Analysis II

Jochen Merker

Sommersemester 2009

Inhaltsverzeichnis

Einleitung

In der Analysis II werden wir uns hauptsächlich mit Konvergenz, Differentiation und Integration im Mehrdimensionalen beschäftigen. Dieser Text ist dazu gedacht, Ihnen einen kurzen Überblick über die Themen zu geben, die dabei behandelt werden. Darüberhinaus sind für interessierte und sehr begabte Studenten auch immer mal wieder weiterführende Literaturtipps angegeben. Da diese Quellen für Studenten im zweiten Semester eigentlich noch zu schwierig sind, lassen Sie sich nicht dadurch entmutigen, wenn Sie dort einen Blick hineinwerfen und nichts verstehen, das verlangt auch keiner von Ihnen.

Falls Sie Fragen, Anregungen oder Wünsche haben, sprechen Sie einfach mich oder einen der Übungsleiter an.

Viel Vergnügen und viel Erfolg beim Studium der Analysis !

Kapitel 1

Konvergenz und Stetigkeit in normierten und metrischen Räumen

Als Vorbereitung für die Differentiations- und Integrationstheorie im Mehrdimensionalen muss man sich (wie auch schon in der Analysis I) zunächst mit der Konvergenz von Folgen vertraut machen – diesmal jedoch im Euklidischen $\mathbb{R}^n$ statt auf der eindimensionalen Zahlengeraden $\mathbb{R}$.

Allgemeiner werden wir die Konvergenz von Folgen gleich in (möglicherweise unendlichdimensionalen) normierten Vektorräumen sowie beliebigen metrischen Räumen diskutieren, da dies keine zusätzliche Schwierigkeit darstellt und insbesondere die spätere Anwendung auf Funktionenräume ermöglicht.

Danach werden wir stetige Funktionen zwischen metrischen Räumen diskutieren, und insbesondere mittels der Begriffe Kompaktheit und Zusammenhang Verallgemeinerungen des Satzes vom Maximum und Minimum bzw. des Zwischenwertsatzes formulieren.

1.1 Normierte Vektorräume

In der linearen Algebra wurden reelle Vektorräume eingeführt und ausführlich diskutiert, d.h. Mengen X mit einer Addition $+ : X \times X \to X$, einem neutralen Element 0 und einer skalaren Multiplikation $\cdot : \mathbb{R} \times X \to X$, so dass $(X, +, 0)$ eine abelsche Gruppe ist und $\cdot$ sowohl die Distributivgesetze als auch $1 \cdot x = x$ erfüllt.

Bekanntestes Beispiel für einen Vektorraum ist sicherlich die Menge $\mathbb{R}^n$ der n-Tupel $(x_1, x_2, \ldots, x_n)$ reeller Zahlen mit der komponentenweisen Addition und skalaren Multiplikation. Tatsächlich ist jeder n-dimensionale Vektorraum linear isomorph zum $\mathbb{R}^n$.

Eine Kurzzusammenfassung der linearen Algebra findet man in [Jänich1].

Wir wollen jetzt auf reellen Vektorräumen einen Längenbegriff einführen, indem wir eine Funktion eine Norm nennen, falls sie die Eigenschaften hat, die wir von einer Länge erwarten.

Definition 1.1 *Eine Funktion* $\|\cdot\| : X \to \mathbb{R}$ *auf einem reellen Vektorraum* X *heißt Norm, falls*

- $\|x\| \geq 0$ *und* $\|x\| = 0 \Leftrightarrow x = 0$ *(Definitheit)*
- $\|\lambda x\| = |\lambda| \|x\|$ *(Homogenität)*
- $\|x + y\| \leq \|x\| + \|y\|$ *(Dreiecksungleichung)*

für alle $x, y \in X$ *und* $\lambda \in \mathbb{R}$ *gilt.*

Einen reellen Vektorraum X zusammen mit einer Norm $\|\cdot\|$ auf X nennen wir einen normierten Vektorraum, oder kurz einen normierten Raum.

Beispiel 1.2 *Normierte Vektorräume sind*

- *der* $\mathbb{R}^n$ *mit der* 1*-Norm* $\|x\|_1 := \sum_{i=1}^{n} |x_i|$ *oder der* ∞*-Norm* $\|x\|_\infty := \sup_{i=1,\ldots,n} |x_i|$,
- *die Menge* $C([0,1], \mathbb{R})$ *der stetigen Funktionen* $u : [0,1] \to \mathbb{R}$ *mit der Supremumsnorm* $\|u\|_\infty := \sup_{x \in [0,1]} |u(x)|$.

Das wichtigste Beispiel für eine Norm ist aber sicherlich die 2-Norm $\|x\|_2 := \sqrt{\sum_{i=1}^{n} |x_i|^2}$ auf dem $\mathbb{R}^n$. Dieses Beispiel wollen wir gleich in einen allgemeineren Kontext stellen.

Definition 1.3 *Eine Funktion* $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$ *auf einem reellen Vektorraum* X *heißt Skalarprodukt, falls*

- $\langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle$ *(Bilinearität)*
- $\langle y, x \rangle = \langle x, y \rangle$ *(Symmetrie)*
- $\langle x, x \rangle \geq 0$ *und* $\langle x, x \rangle = 0 \Leftrightarrow x = 0$ *(Positive Definitheit)*

für alle $x, y, z \in X$ *und* $\lambda, \mu \in \mathbb{R}$ *gilt.*

Einen Vektorraum X zusammen mit einem Skalarprodukt $\langle \cdot, \cdot \rangle$ auf X nennen wir einen Euklidischen Vektorraum.

Beispiel 1.4 *Euklidische Vektorräume sind*

- *der* $\mathbb{R}^n$ *mit dem Skalarprodukt* $\langle x, y \rangle := \sum_{i=1}^{n} x_i y_i$,
- *die Menge* $C_{2\pi}(\mathbb{R}, \mathbb{R})$ *der stetigen* 2π*-periodischen Funktionen* $u, v : \mathbb{R} \to \mathbb{R}$ *mit dem Skalarprodukt* $\langle u, v \rangle := \int_0^{2\pi} u(x) v(x) \, dx$.

Wir wollen nun beweisen, dass ein Euklidischer Vektorraum $(X, \langle\cdot,\cdot\rangle)$ durch die Definition $\|x\| := \sqrt{\langle x, x\rangle}$ ein normierter Vektorraum wird.

Satz 1.5 *In jedem Euklidischen Vektorraum $(X, \langle\cdot,\cdot\rangle)$ gilt für alle $x, y \in X$ die Cauchy-Schwarzsche Ungleichung*

$$|\langle x, y\rangle| \leq \|x\|\|y\| \quad .$$

Beweis: Für $y = 0$ ist die Ungleichung trivial, für $y \neq 0$ setze man $\lambda := \frac{\langle x,y\rangle}{\|y\|^2}$ und berechne

$$0 \leq \|x - \lambda y\|^2 = \|x\|^2 - 2\lambda\langle x, y\rangle + \lambda^2\|y\|^2 = \|x\|^2 - \frac{(\langle x, y\rangle)^2}{\|y\|^2} \quad ,$$

woraus $|\langle x, y\rangle| \leq \|x\|\|y\|$ folgt. □

Korollar 1.6 *Jeder Euklidische Vektorraum $(X, \langle\cdot,\cdot\rangle)$ ist mit $\|x\| := \sqrt{\langle x, x\rangle}$ ein normierter Vektorraum.*

Beweis: Definitheit und Homogenität sind klar, und die Dreiecksungleichung folgt aus

$$\|x + y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2 \quad .$$

□

Außerdem liegt wegen der Cauchy-Schwarzschen Ungleichung die Zahl $\frac{\langle x,y\rangle}{\|x\|\|y\|}$ zwischen -1 und 1, so dass man durch $\cos(\phi) = \frac{\langle x,y\rangle}{\|x\|\|y\|}$ den Winkel ϕ zwischen zwei Vektoren x, y definieren kann. Insbesondere heißen Vektoren x, y mit $\langle x, y\rangle = 0$ orthogonal zueinander.

Bemerkung 1.7

- *Ähnlich kann man Norm und Skalarprodukt auch für komplexe Vektorräume definieren.*
- *Eine Norm $\|\cdot\|$ auf einem Vektorraum X wird genau dann von einem Skalarprodukt induziert, wenn die Parallelogrammgleichung $\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$ für alle $x, y \in X$ gilt (ein Beweis steht z.B. in [Mathieu, 8])*

Aufgabe: *Beweisen Sie, dass in einem Euklidischen Vektorraum die Parallelogrammgleichung gilt.*

1.2 Metrische Räume

Normierte Vektorräume sind noch nicht ausreichend für eine zufriedenstellende Analysis, denn z.B. sind Teilmengen, die keine Unterräume sind, nicht selbst wieder normierte Räume. Deshalb definiert man als Verallgemeinerung metrische Räume, in denen man durch eine Metrik genannte Funktion einen Abstandsbegriff zur Verfügung hat.

Definition 1.8 *Eine Funktion $d : X \times X \to \mathbb{R}$ auf einer Menge X heißt Metrik, falls*

- $d(x, y) \geq 0$ *und* $d(x, y) = 0 \Leftrightarrow x = y$ *(Definitheit)*
- $d(y, x) = d(x, y)$ *(Symmetrie)*
- $d(x, z) \leq d(x, y) + d(y, z)$ *(Dreiecksungleichung)*

für alle $x, y, z \in X$ gilt.

Eine Menge X zusammen mit einer Metrik d nennen wir einen metrischen Raum. Die Elemente von X bezeichnen wir als Punkte, und den Wert $d(x, y)$ als Abstand der Punkte $x, y \in X$.

Beispiel 1.9 *Jeder normierte Vektorraum $(X, \|\cdot\|)$ wird durch die Definition von $d(x, y) := \|x - y\|$ ein metrischer Raum.*

Andere Beispiele für metrische Räume sind

- *jede nichtleere Menge X mit der diskreten Metrik* $d_{diskret}(x, y) := \begin{cases} 0 & \text{falls } x = y \\ 1 & \text{falls } x \neq y \end{cases}$,
- *jede Teilmenge $M \subset X$ eines metrischen Raumes (X, d) mit der auf M eingeschränkten Funktion $d|_{M \times M}$ als Metrik, insbesondere für eine Norm $\|\cdot\|$ auf $\mathbb{R}^n$ jede Teilmenge $M \subset \mathbb{R}^n$ mit der Metrik $d(x, y) := \|x - y\|$ für $x, y \in M$,*
- *jedes Produkt $X \times Y$ metrischer Räume $(X, d_X), (Y, d_Y)$ mit der Produktmetrik* $d_{X \times Y}((x, y), (x', y')) := d_X(x, x') + d_Y(y, y')$.

Wir wollen noch gewisse Teilmengen eines metrischen Raumes benennen.

Definition 1.10

- *Für $r > 0$ nennt man die Teilmenge $B_r(x) := \{y \in X \mid d(x, y) < r\}$ eines metrischen Raumes (X, d) die offene Kugel um den Punkt $x \in X$ mit Radius r.*
- *Eine Teilmenge $U \subset X$ eines metrischen Raumes (X, d) nennt man offen, falls es für jedes $x \in U$ ein $r > 0$ gibt mit $B_r(x) \subset U$, d.h. wenn es um jeden Punkt der Menge U eine offene Kugel gibt, die komplett in U liegt.*

Insbesondere sind offene Kugeln auch offene Teilmengen, denn ist $y \in B_r(x)$, dann gilt ja $d(x, y) < r$, also gibt es ein $\epsilon > 0$ mit $d(x, y) + \epsilon < r$, und mit diesem gilt $B_\epsilon(y) \subset B_r(x)$ wegen $d(x, z) \leq d(x, y) + d(y, z) < d(x, y) + \epsilon < r$ für $z \in B_\epsilon(y)$.

Aufgabe: *Zeichnen Sie im $\mathbb{R}^2$ und $\mathbb{R}^3$ die offenen Einheitskugeln $B_1(0)$ für die Normen $\|\cdot\|_1$, $\|\cdot\|_2$ und $\|\cdot\|_\infty$.*

Das Mengensystem $\{U \subset X \mid U \text{ offen}\}$ aller offenen Teilmengen nennt man die Topologie des metrischen Raumes (X, d).

Satz 1.11 *In einem metrischen Raum* (X, d) *gilt:*

(a) $\emptyset$ *und* X *sind offen.*

(b) *Sind* U *und* V *offen, so auch* $U \cap V$.

(c) *Sind die Mengen* U_i $(i \in I)$ *offen, so ist auch* $\bigcup_{i \in I} U_i$ *offen.*

Beweis:

(a) trivial

(b) Sei $x \in U \cap V$, dann gibt es aufgrund der Offenheit von U, V Radien $r_1, r_2 > 0$ mit $B_{r_1}(x) \subset U$, $B_{r_2}(x) \subset V$. Also gilt mit $r := \min(r_1, r_2)$ auch $B_r(x) \subset U \cap V$. Somit ist $U \cap V$ offen.

(c) Sei $x \in \bigcup_{i \in I} U_i$, dann gibt es ein $i \in I$ mit $x \in U_i$, und da U_i offen ist auch ein $r > 0$ mit $B_r(x) \subset U_i$. Wegen $U_i \subset \bigcup_{i \in I} U_i$ folgt dann aber auch $B_r(x) \subset \bigcup_{i \in I} U_i$. Somit ist die Menge $\bigcup_{i \in I} U_i$ offen.

$\square$

Bemerkung 1.12 *Noch allgemeiner als metrische Räume sind topologische Räume, bei denen man eine Menge* X *nur mit einem System von Teilmengen versieht, das die in Satz 1.11 angegebenen Eigenschaften hat. Wir wollen uns aber mit diesen nicht genauer beschäftigen, ein gut zu lesender Literaturtipp dazu ist [Jänich2].*

Beispiel 1.13

- *Bezüglich der diskreten Metrik* $d_{diskret}$ *auf einer Menge* X *ist jede Teilmenge offen, denn wegen* $B_{1/2}(x) = \{x\}$ *sind Punkte offen, und wegen der dritten Eigenschaft aus Satz 1.11 sind dann auch beliebige Teilmengen offen.*

- *Bezüglich der auf eine Teilmenge* M *eines metrischen Raumes* (X, d) *eingeschränkten Metrik* $d|_{M \times M}$ *sind genau die Teilmengen* U *offen, für die es eine in* X *offene Menge* V *mit* $U = M \cap V$ *gibt. Denn die offene Kugel um* x *mit Radius* r *in* M *ist nichts anderes als der Schnitt* $M \cap B_r(x)$ *mit der offenen Kugel* $B_r(x)$ *in* (X, d). *Die Topologie von* $(M, d|_{M \times M})$ *nennt man die Relativtopologie der Teilmenge* $M \subset X$.

  Beispielsweise ist im Intervall $[0, 1] \subset \mathbb{R}$ *die Menge* $[0, 1/2)$ *offen, da sie der Schnitt von* $[0, 1]$ *mit der in* $\mathbb{R}$ *offenen Menge* $(-1/2, 1/2)$ *ist.*

- *Im Produkt* $X \times Y$ *metrischer Räume* $(X, d_X), (Y, d_Y)$ *mit der Produktmetrik sind genau die Mengen* W *offen, für die es zu jedem Punkt* $(x, y) \in W$ *offene Mengen* $U \subset X$ *und* $V \subset Y$ *mit* $(x, y) \in U \times V \subset W$ *gibt. Denn in der offenen Kugel mit Radius* r *um* (x, y) *bzgl. der Produktmetrik liegen die Mengen* $B_{r'}(x) \times B_{r''}(y)$ *für* $r' + r'' \leq r$, *wobei* $B_{r'}(x)$ *eine Kugel in* X *und* $B_{r''}(y)$ *eine Kugel in* Y *ist.*

  Insbesondere sind in Produkten die Mengen der Form $U \times V$, $U \subset X$ *und* $V \subset Y$ *offen, nicht die einzigen offene Mengen, sondern es gibt auch offenen Mengen, die nicht diese simple Form haben.*

Wir werden später sehen, dass Konvergenz und Stetigkeit nur von der Topologie eines metrischen Raumes abhängen. Deswegen nennt man zwei Metriken auf derselben Menge X äquivalent, falls sie dieselbe Topologie besitzen.

Aufgabe: *Zeigen Sie, dass für einen metrischen Raum (X, d) durch $d'(x, y) := \frac{d(x,y)}{1+d(x,y)}$ eine zu d äquivalente Metrik definiert wird.*

Auch wenn d nicht beschränkt ist, ist die Metrik d' beschränkt, der Begriff der Beschränktheit ist also kein rein topologischer Begriff, sondern hängt von der Wahl der Metrik ab.

Lemma 1.14 *Zwei Normen $\|\cdot\|$ und $\|\cdot\|'$ auf einem Vektorraum X sind genau dann äquivalent, wenn es Konstanten $c, C > 0$ gibt mit*

$$\forall x \in X \;:\; c\|x\| \leq \|x\|' \leq C\|x\| \quad .$$

Beweis:

- Angenommen, die Topologien von $(X, \|\cdot\|)$ und $(X, \|\cdot\|')$ sind dieselben. Dann ist die offene Kugel $B_1(0)$ bzgl. $\|\cdot\|$ auch offen bzgl. $\|\cdot\|'$, und daher gibt es eine bzgl. $\|\cdot\|'$ offene Kugel $B'_r(0)$ mit $B'_r(0) \subset B_1(0)$. Da für beliebiges $x \neq 0$ die Beziehung $\frac{r}{2\|x\|'}x \in B'_r(0) \subset B_1(0)$ gilt, erhält man $\frac{r}{2}\|x\| \leq \|x\|'$ für alle $x \in X$, hat also eine Konstante c mit $c\|x\| \leq \|x\|'$ gefunden.

  Analog erhält man durch Vertauschung der Normen in der Argumentation auch ein $r' > 0$ mit $\frac{r'}{2}\|x\|' \leq \|x\|$, so dass mit $C := \frac{2}{r'}$ auch die andere Ungleichung $\|x\|' \leq C\|x\|$ gilt.

- Andererseits gilt mit Konstanten c, C wie oben trivialerweise

  $$B'_{cr}(x) \subset B_r(x) \subset B'_{Cr}(x) \quad ,$$

  d.h. jede bzgl. $\|\cdot\|$ offene Kugel um x enthält auch eine bzgl. $\|\cdot\|'$ offene Kugel um x und umgekehrt. Somit sind (nach Definition der Offenheit einer Menge) die offenen Mengen bzgl. $\|\cdot\|$ dieselben wie die offenen Mengen bzgl. $\|\cdot\|'$.

$\square$

Aufgabe: *Beweisen Sie, dass Äquivalenz von Normen eine Äquivalenzrelation auf der Menge aller Normen auf X ist.*

1.3 Konvergenz in endlich-dimensionalen normierten Vektorräumen

Bevor wir allgemein über Konvergenz in metrischen Räumen sprechen, wollen wir als Spezialfall zunächst einmal die Konvergenz in (endlich-dimensionalen) normierten Vektorräumen betrachten.

Definition 1.15 *Eine Folge* x_k *von Punkten in einem normierten Vektorraum* $(X, \|\cdot\|)$ *nennt man konvergent gegen den Punkt* $x \in X$*, falls* $\lim\limits_{k\to\infty} \|x_k - x\| = 0$ *in* $\mathbb{R}$ *gilt oder äquivalenterweise*

$$\forall \epsilon > 0 \exists K \in \mathbb{N} : \forall k \geq K : \|x_k - x\| \leq \epsilon \quad .$$

Man beachte, dass bei Konvergenz einer Folge x_k der Grenzwert eindeutig ist, denn konvergiert x_k sowohl gegen x als auch gegen x', dann gilt wegen

$$\|x' - x\| \leq \|x_k - x\| + \|x_k - x'\| \to 0$$

für $k \to \infty$ also $\|x' - x\| = 0$ und somit $x = x'$. Den Grenzwert einer Folge x_k bezeichnet man mit $\lim\limits_{k\to\infty} x_k$.

Im $\mathbb{R}^n$ mit den drei von uns zuvor diskutierten Normen läßt sich Konvergenz ganz einfach charakterisieren.

Lemma 1.16 *Eine Folge von Punkten* $x_k = (x_{1k}, x_{2k}, \ldots, x_{nk})$ *im durch* $\|\cdot\|_1$, $\|\cdot\|_2$ *oder* $\|\cdot\|_\infty$ *normierten* $\mathbb{R}^n$ *konvergiert genau dann gegen* $x = (x_1, x_2, \ldots, x_n)$*, wenn* $\lim\limits_{k\to\infty} x_{jk} = x_j$ *für jedes* $j = 1, 2, \ldots, n$ *gilt, d.h. wenn die Komponenten der Folge in* $\mathbb{R}$ *konvergieren.*

Beweis: Für jede der drei Normen gilt die Ungleichung

$$|x_{jk} - x_j| \leq \|x_k - x\| \leq |x_{1k} - x_1| + \cdots + |x_{nk} - x_n|$$

für beliebiges $j = 1, 2, \ldots, n$. Also folgt einerseits aus $\lim\limits_{k\to\infty} \|x_k - x\| = 0$ auch die Konvergenz $\lim\limits_{k\to\infty} |x_{jk} - x_j| = 0$ für jedes $j = 1, 2, \ldots, n$, und andererseits aus der Konvergenz $\lim\limits_{k\to\infty} |x_{jk} - x_j| = 0$ für jedes $j = 1, 2, \ldots, n$ auch $\lim\limits_{k\to\infty} \|x_k - x\| = 0$. $\square$

Im folgenden fassen wir den $\mathbb{R}^n$ als mit einer der drei Normen ausgestatteten normierten Vektorraum auf. Da die Konvergenz im so normierten $\mathbb{R}^n$ nach dem vorigen Lemma auf die in $\mathbb{R}$ zurückgeführt werden kann, übertragen sich auch die zwei wichtigsten Eigenschaften. Um diese zu formulieren, nennen wir analog zum Eindimensionalen eine Folge x_k von Punkten im $\mathbb{R}^n$

- eine Cauchy-Folge, wenn $\forall \epsilon > 0 \exists K \in \mathbb{N} : \forall k, l \geq K : \|x_k - x_l\| \leq \epsilon$.
- beschränkt, wenn es eine Kugel $B_r(x)$ um einen Punkt x gibt, in der alle x_k liegen, oder äquivalenterweise, wenn es eine Konstante C mit $\|x_k\| \leq C$ für alle $k \in \mathbb{N}$ gibt.

Satz 1.17

(a) *Jede Cauchy-Folge im* $\mathbb{R}^n$ *konvergiert, d.h.* $\mathbb{R}^n$ *ist vollständig.*

(b) *Jede beschränkte Folge im* $\mathbb{R}^n$ *besitzt eine konvergente Teilfolge, d.h. der Satz von Bolzano-Weierstraß gilt im* $\mathbb{R}^n$*.*

Beweis:

(a) Ist x_k eine Cauchy-Folge, so sind wegen $|x_{jk} - x_{jl}| \leq \|x_k - x_l\|$ auch alle Komponenten x_{jk} für $j = 1, 2, \ldots, n$ Cauchy-Folgen in $\mathbb{R}$ und konvergieren somit. Daher konvergiert nach Lemma 1.16 auch die Folge x_k in $\mathbb{R}^n$.

(b) Nehmen wir für einen Beweis per Induktion an, dass der Satz von Bolzano-Weierstraß für Folgen im $\mathbb{R}^{n-1}$ gilt. Sei nun x_k eine beschränkte Folge im $\mathbb{R}^n$. Dann sind wegen $\|(x_{1k}, x_{2k}, \ldots, x_{(n-1)k})\|_{\mathbb{R}^{n-1}} \leq \|x_k\|_{\mathbb{R}^n}$ und $|x_{nk}| \leq \|x_k\|_{\mathbb{R}^n}$ sowohl die Folge $(x_{1k}, x_{2k}, \ldots, x_{(n-1)k})$ im $\mathbb{R}^{n-1}$ als auch die Folge x_{nk} in $\mathbb{R}$ beschränkt. Demzufolge kann man zunächst eine Teilfolge von x_k auswählen, für die x_{nk} in $\mathbb{R}$ konvergiert, und dann davon wiederum eine Teilfolge, für die $(x_{1k}, x_{2k}, \ldots, x_{(n-1)k})$ im $\mathbb{R}^{n-1}$ konvergiert. Bei dieser Teilfolge von x_k konvergieren dann alle Komponenten und daher auch die Teilfolge selbst im $\mathbb{R}^n$.

□

Nun gibt es aber natürlich noch völlig andere Normen auf dem $\mathbb{R}^n$ als nur die drei bisher von uns betrachteten, und die Frage stellt sich, ob die gerade bewiesenen Eigenschaften auch bei Verwendung dieser Normen gültig bleiben. Die positive Antwort darauf ergibt sich aus den folgenden beiden Sätzen.

Satz 1.18 *Sei x_k eine Folge im Vektorraum X und seien $\|\cdot\|$, $\|\cdot\|'$ zwei äquivalente Normen auf X. Dann gilt:*

Die Folge x_k konvergiert bzgl. $\|\cdot\|$ genau dann gegen x, wenn sie bzgl. $\|\cdot\|'$ gegen x konvergiert.

Beweis: Nach Lemma 1.14 gibt es Konstanten $c, C > 0$ mit

$$c\|x_k - x\| \leq \|x_k - x\|' \leq C\|x_k - x\| \quad .$$

Also folgt aus $\lim\limits_{k\to\infty} \|x_k - x\| = 0$ auch $\lim\limits_{k\to\infty} \|x_k - x\|' = 0$ und umgekehrt. □

Satz 1.19 *Je zwei Normen auf dem $\mathbb{R}^n$ sind äquivalent.*

Beweis: Es reicht zu zeigen, dass jede Norm $\|\cdot\|$ auf dem $\mathbb{R}^n$ zur Euklidischen Norm $\|\cdot\|_2$ äquivalent ist.

Nun gilt einerseits mit den Einheitsvektoren $e_1, e_2, \ldots, e_n$ die Gleichung $x = \sum\limits_{j=1}^{n} x_j e_j$ und daher nach der Dreiecksungleichung und der Cauchy-Schwarzschen Ungleichung

$$\|x\| = \|\sum_{j=1}^{n} x_j e_j\| \leq \sum_{j=1}^{n} |x_j|\|e_j\| \leq \left(\sqrt{\sum_{j=1}^{n} \|e_j\|^2}\right) \|x\|_2 \quad ,$$

also mit $C := \sqrt{\sum\limits_{j=1}^{n} \|e_j\|^2}$ die Ungleichung $\|x\| \leq C\|x\|_2$.

Andererseits ist die größte Konstante $c \geq 0$, für die $c\|x\|_2 \leq \|x\|$ gilt, durch die Gleichung $c = \inf\{\|x\| \,|\, x \in S^{n-1}\}$ gegeben, wobei $S^{n-1} := \{x \in \mathbb{R}^n \,|\, \|x\|_2 = 1\}$ die Euklidische Sphäre vom Radius Eins bezeichnet. Nun kann man $c > 0$ zeigen, indem man die Annahme $c = 0$ zum Widerspruch führt:

Angenommen, es wäre $c = 0$, d.h. es gäbe eine Folge x_k mit $\lim\limits_{k\to\infty} \|x_k\| = 0$ und $\|x_k\|_2 = 1$. Da wegen $\|x_k\|_2 = 1$ die Folge x_k in der Euklidischen Norm beschränkt ist, enthält sie nach dem Satz von Bolzano-Weierstraß eine bzgl. $\|\cdot\|_2$ konvergente Teilfolge. Der Grenzwert x dieser Teilfolge $x_{k'}$ hat dann die Euklidische Norm 1 wegen

$$\|x\|_2^2 = \sum_{j=1}^{n} |x_j|^2 = \lim_{k'\to\infty} \sum_{j=1}^{n} |x_{jk'}|^2 = 1 \quad .$$

Aufgrund von $\lim\limits_{k\to\infty} \|x_{k'}\| = 0$ folgt aber mit der Ungleichung $\|\cdot\| \leq C\|\cdot\|_2$ auch

$$\|x\| \leq \|x - x_{k'}\| + \|x_{k'}\| \leq C\|x - x_{k'}\|_2 + \|x_{k'}\| \to 0$$

für $k' \to \infty$, d.h. $\|x\| = 0$ und somit $x = 0$. Dies steht aber im Widerspruch zu $\|x\|_2 = 1$. $\square$

Insbesodere folgt aus diesen beiden Sätzen, dass unabhängig von der Wahl der Norm im $(\mathbb{R}^n, \|\cdot\|)$ Konvergenz mit komponentenweiser Konvergenz identisch ist, jede Cauchy-Folge konvergiert und der Satz von Bolzano-Weierstraß gilt.

Satz 1.19 kann man auch auf beliebige endlich-dimensionale Vektorräume übertragen.

Aufgabe: *Zeigen Sie, dass für eine injektive lineare Abbildung $A : X \to Y$ und eine Norm $\|\cdot\|_Y$ auf Y durch $\|x\|_X := \|Ax\|_Y$ eine Norm auf X definiert wird.*

Korollar 1.20 *Je zwei Normen auf einem endlich-dimensionalen reellen Vektorraum sind äquivalent.*

Beweis: Aus der linearen Algebra ist bekannt, dass jeder reelle Vektorraum Y von endlicher Dimension n zum $\mathbb{R}^n$ linear isomorph ist, d.h. es gibt eine bijektive lineare Abbildung $A : \mathbb{R}^n \to Y$.

Ist nun $\|\cdot\|_Y$ eine Norm auf Y, so induziert A durch $\|x\| := \|Ax\|_Y$ eine entsprechende Norm $\|\cdot\|$ auf $\mathbb{R}^n$, und wegen $\|Ax\|_Y = \|x\|$ ist A nicht nur ein linearer Isomorphismus, sondern erhält auch noch die Norm (gerade so wurde eben die Norm $\|\cdot\|$ auf $\mathbb{R}^n$ konstruiert).

Macht man dasselbe mit einer anderen Norm $\|\cdot\|'_Y$ auf Y, so erhält man mit der zugehörigen Norm $\|\cdot\|'$ auf dem $\mathbb{R}^n$ nach Satz 1.19 Konstanten $c, C > 0$ mit

$$\forall x \in \mathbb{R}^n \,:\, c\|x\| \leq \|x\|' \leq C\|x\| \quad .$$

Wegen $\|y\|_Y = \|A(A^{-1}y)\|_Y = \|A^{-1}x\|$ und $\|y\|'_Y = \|A(A^{-1}y)\|'_Y = \|A^{-1}x\|'$ hat man dann aber auch

$$\forall y \in Y \,:\, c\|y\|_Y \leq \|y\|'_Y \leq C\|y\|_Y \quad .$$

□

Die Konvergenz einer Folge hängt in einem endlich-dimensionalen normierten Vektorraum also überhaupt nicht von der Wahl der Norm ab. **Deshalb gibt man – solange es nur um Konvergenz geht – bei einem endlich-dimensionalen Vektorraum meistens die gewählte Norm gar nicht mehr explizit an**.

1.4 Konvergenz in metrischen Räumen

Konvergenz in metrischen Räumen kann man ganz analog zu Definition 1.15 erklären.

Definition 1.21 *Eine Folge x_k von Punkten in einem metrischen Raum (X, d) nennt man konvergent gegen den Punkt $x \in X$, falls $\lim\limits_{k\to\infty} d(x_k, x) = 0$ in $\mathbb{R}$ gilt oder äquivalenterweise*

$$\forall \epsilon > 0 \exists K \in \mathbb{N} : \forall k \geq K : d(x_k, x) \leq \epsilon \quad .$$

Wie zuvor im Spezialfall normierter Vektorräume zeigt man auch hier mittels der Dreiecksungleichung die Eindeutigkeit des Grenzwertes einer Folge x_k und bezeichnet den Grenzwert mit $\lim\limits_{k\to\infty} x_k$.

Aufgabe: *Zeigen Sie, dass im mit der diskreten Metrik $d_{diskret}$ versehenen $\mathbb{R}^n$ eine Folge x_k genau dann gegen x konvergiert, wenn sie ab einem gewissen Index K konstant gleich x ist, d.h. wenn $\exists K \in \mathbb{N} : \forall k \geq K : x_k = x$.*

Konvergenz bezüglich der diskreten Metrik ist also völlig verschieden von der üblichen Konvergenz bezüglich einer Norm auf dem $\mathbb{R}^n$. Insbesondere hängt die Konvergenz bei einem nur mit einer Metrik versehenen endlich-dimensionalen Vektorraum sehr wohl von der Wahl der Metrik ab.

Wir wollen Konvergenz noch mittels rein topologischer Begriffe charakterisieren und führen dazu den Begriff der Umgebung eines Punktes ein.

Definition 1.22 *Man nennt eine Teilmenge V eines metrischen Raumes (X, d) Umgebung des Punktes $x \in X$, wenn es eine Kugel $B_\epsilon(x)$ $(\epsilon > 0)$ mit $B_\epsilon(x) \subset V$ gibt, oder äquivalenterweise, wenn es eine offene Menge $U \subset V$ mit $x \in U$ gibt.*

Insbesondere ist eine Teilmenge eines metrischen Raumes genau dann offen, wenn sie Umgebung jedes ihrer Punkte ist.

Desweiteren besitzen zwei verschiedene Punkte $x, x' \in X$ in einem metrischen Raum (X, d) punktfremde Umgebungen, d.h. zu $x \neq x'$ gibt es Umgebungen U von x und U' von x' mit $U \cap U' = \emptyset$. Diese auch als Hausdorffsches Trennungsaxiom bezeichnete Eigenschaft gilt in einem metrischen Raum einfach deswegen, weil man ja bei $x \neq x'$ auch $r := d(x, x') > 0$ hat und damit einfach $U := B_{r/2}(x)$ und $U' := B_{r/2}(x')$ wählen kann. Gäbe es nämlich ein $y \in U \cap U'$, dann wäre $d(x, y) < r/2$, $d(y, x') < r/2$ und damit auch $d(x, x') \leq d(x, y) + d(y, x') < r$ im Widerspruch zu $d(x, x') = r$.

Nun kommen wir zur angekündigten Charakterisierung der Konvergenz mittels rein topologischer Begriffe.

Lemma 1.23 *Eine Folge x_k von Punkten in einem metrischen Raum (X, d) konvergiert genau dann gegen $x \in X$, wenn in jeder Umgebung von x alle bis auf endlich viele Folgenglieder x_k liegen, d.h. wenn*

$$\forall \text{ Umgebungen } V \text{ von } x \, \exists K \in \mathbb{N} : \forall k \geq K : x_k \in V$$

Beweis: Konvergiere x_k gegen x. Ist V eine Umgebung von x, so gibt es eine Kugel $B_\epsilon(x) \subset V$. Zu diesem $\epsilon > 0$ gibt es wegen der Konvergenz von x_k gegen x ein $K \in \mathbb{N}$ mit $d(x_k, x) < \epsilon$ für alle $k \geq K$. Also gibt es zu V ein K mit $x_k \in V$ für alle $k \geq K$.

Umgekehrt ist natürlich $B_\epsilon(x)$ eine Umgebung von x, es gibt also zu $\epsilon > 0$ ein K mit $x_x \in B_\epsilon(x) \Leftrightarrow d(x_k, x) \leq \epsilon$ für alle $k \geq K$. □

Die wichtigste Konsequenz des vorigen Lemmas ist, dass Konvergenz in metrischen Räumen nicht direkt von der Wahl der Metrik abhängt, sondern nur von der Topologie.

In Verallgemeinerung von 1.16 charakterisiert das folgende Lemma Konvergenz in Produkten.

Lemma 1.24 *Eine Folge (x_k, y_k) konvergiert im Produkt $X \times Y$ metrischer Räume X, Y genau dann gegen (x^*, y^*), wenn x_k in X gegen x^* und y_k in Y gegen y^* konvergiert.*

Beweis: Dies folgt direkt aus der Definition der Produktmetrik $d_{X \times Y}((x_k, y_k), (x^*, y^*)) = d_X(x_k, x^*) + d_Y(y_k, y^*)$ und der Definition von Konvergenz.

Alternativ kann man die Aussage auch rein topologisch durch die Beobachtung beweisen, dass eine Teilmenge W des Produktes $X \times Y$ genau dann eine Umgebung von (x^*, y^*) ist, wenn es Umgebungen U von x^* und V von y^* mit $U \times V \subset W$ gibt. □

Wir wollen noch einige weitere topologische Begriffe einführen, mittels derer wir Teilmengen eines metrischen Raumes genauer beschreiben können.

Definition 1.25 *Ein Punkt x heißt Berührpunkt einer Teilmenge M in einem metrischen Raum (X, d), wenn es eine Folge $x_k \in M$ mit $\lim\limits_{k \to \infty} x_k = x$ gibt, oder wenn äquivalenterweise in jeder Umgebung von x ein Punkt aus M liegt.*

Die Menge aller Berührpunkte von M bezeichnet man mit $\overline{M}$ und nennt sie den Abschluss von M. Es gilt $M \subset \overline{M}$, da für $x \in M$ die konstante Folge $x, x, \ldots$ gegen x konvergiert.

Definition 1.26 *Eine Teilmenge $A \subset X$ eines metrischen Raumes (X, d) nennt man abgeschlossen, wenn sie jeden ihrer Berührpunkte enthält, d.h. wenn $A = \overline{A}$ gilt, oder – mit anderen Worten – wenn für jede Folge $x_k \in A$, die in X konvergiert, auch der Grenzwert $\lim\limits_{k \to \infty} x_k$ in A liegt.*

Äquivalenterweise heißt eine Teilmenge $A \subset X$ abgeschlossen, wenn sie das Komplement $A = X \setminus U$ einer offenen Menge $U \subset X$ ist.

Die Äquivalenz der beiden angegebenen Bedingungen kann man folgendermaßen einsehen: Enthält A all seine Berührpunkte, dann ist jeder Punkt $x \in X \setminus A$ kein Berührpunkt von A, also gibt es zu jedem Punkt $x \in X \setminus A$ eine Umgebung V von x, in der kein Punkt aus A liegt, d.h. eine Umgebung V von x mit $V \subset X \setminus A$. Somit ist $U := X \setminus A$ Umgebung all seiner Punkte und daher offen, d.h. $A = X \setminus U$ ist das Komplement einer offenen Menge.

Ist umgekehrt $A = X \setminus U$ Komplement einer offene Menge U und x ein Berührpunkt von A, dann kann x nicht in U liegen. Denn läge x in U, dann gäbe es aufgrund der Offenheit von U eine Umgebung um x, die komplett in U läge und daher keinen Punkt aus A enthielte, im Widerspruch dazu, dass x Berührpunkt von A ist. Also liegt x (und somit jeder Berührpunkt von A) automatisch in A.

Festhalten wollen wir noch, dass die Bezeichnung von $\overline{M}$ als Abschluss von M Sinn macht, da $\overline{M}$ die minimale M enthaltende abgeschlossene Menge ist. Analog dazu nennt man die maximale in M enthaltene offene Menge das Innere von M und bezeichnet Sie mit M^o. Die Menge M^o besteht gerade aus den inneren Punkten von M, d.h. aus denjenigen Punkten $x \in M$, für die es eine Kugel $B_r(x)$ $(r > 0)$ um x mit $B_r(x) \subset M$ gibt.

Zu guter letzt definiert man noch den Rand ∂M einer Teilmenge M durch $\partial M := \overline{M} \setminus M^o$. Der Rand von M besteht genau aus denjenigen Punkten x, für die in jeder Umgebung von x sowohl Punkte aus M als auch aus $X \setminus M$ liegen.

1.5 Vollständigkeit

Ganz analog zum Fall (endlich-dimensionaler) normierter Vektorräume nennt man eine Folge x_k in einem beliebigen metrischen Raum (X, d) eine Cauchy-Folge, wenn die Aussage $\forall \epsilon > 0 \, \exists K \in \mathbb{N} : \forall k, l \geq K : d(x_k, x_l) \leq \epsilon$ gilt. Jede konvergente Folge ist automatisch eine Cauchy-Folge, aber im Gegensatz zu endlich-dimensionalen Vektorräumen konvergieren Cauchy-Folgen in unendlich-dimensionalen normierten Räumen oder allgemeinen metrischen Räumen nicht unbedingt.

Definition 1.27 *Ein metrischer Raum (X, d) heißt vollständig, wenn jede Cauchy-Folge x_k in X konvergiert.*

Einen vollständigen normierten Vektorraum $(X, \|\cdot\|)$ nennt man Banach-Raum, und einen vollständigen Euklidischen Vektorraum $(X, \langle\cdot,\cdot\rangle)$ nennt man Hilbert-Raum.

Die Vollständigkeit eines metrischen Raumes ist sehr wichtig, da man sich Lösungen von Gleichungen häufig nur unter der Annahme verschaffen kann, dass der zugrundeliegende Raum vollständig ist.

Beispiel 1.28 *Die Menge $\mathbb{Q}$ der rationalen Zahlen ist mit der vom Betrag $|\cdot|$ induzierten Metrik nicht vollständig. Deswegen hatte man sich in der Analysis I ja gerade entschieden, nicht in $\mathbb{Q}$ zu arbeiten, sondern im vollständigen metrischen Raum $(\mathbb{R}, |\cdot|)$.*

In $\mathbb{R}$ konnte man dann im Gegensatz zu $\mathbb{Q}$ z.B. die Existenz von Wurzeln aus beliebigen positiven Zahlen zeigen.

Hier noch zwei unendlich-dimensionale Beispiele:

- *Die Menge $C([0,1],\mathbb{R})$ der stetigen Funktionen $u : [0,1] \to \mathbb{R}$ ist mit der Supremumsnorm $\|u\|_\infty := \sup\limits_{x\in[0,1]} |u(x)|$ ein Banach-Raum.*

  Denn ist u_k eine Cauchy-Folge in $C([0,1],\mathbb{R})$, dann ist wegen $|u_k(x) - u_l(x)| \leq \|u_k - u_l\|_\infty$ für jedes $x \in [0,1]$ auch die Folge $u_k(x)$ in $\mathbb{R}$ eine Cauchy-Folge und konvergiert somit. Bezeichnet man den Grenzwert mit $u(x)$, so wird dadurch eine Funktion $u : [0,1] \to \mathbb{R}$ definiert. Diese Funktion ist nicht nur der punktweise Grenzwert der Folge von Funktionen u_k, sondern u_k konvergiert auch gleichmäßig gegen u. Tatsächlich, zu $\epsilon > 0$ gibt es ein $K \in \mathbb{N}$ mit $|u_k(x) - u_l(x)| \leq \epsilon$ für alle $k, l \geq K$ und alle $x \in [0,1]$, und der Grenzübergang $l \to \infty$ zeigt, dass es zu $\epsilon > 0$ es ein $K \in \mathbb{N}$ mit $|u_k(x) - u(x)| \leq \epsilon$ für alle $k \geq K$ und alle $x \in [0,1]$ gibt, also u_k gleichmäßig gegen u konvergiert. Nun sind die Funktionen u_k auch noch stetig, also ist auch u stetig, und somit gilt wirklich $\lim\limits_{k\to\infty} u_k = u$ in $C([0,1],\mathbb{R})$ bzgl. der Supremumsnorm. Dieses Resultat war natürlich schon als Cauchy-Kriterium für gleichmäßige Konvergenz aus der Analysis I bekannt.

- *Die Menge $C_{2\pi}(\mathbb{R},\mathbb{R})$ der stetigen 2π-periodischen Funktionen u ist mit der vom Skalarprodukt $\langle u, v\rangle := \int_0^{2\pi} u(x)v(x)\,dx$ induzierten Norm $\|u\| := \sqrt{\int_0^{2\pi} |u(x)|^2\,dx}$ nicht vollständig und somit kein Hilbert-Raum.*

  Denn ist $u_k(x) := \frac{a_0}{2} + \sum_{n=1}^k a_n \sin(nx) + b_n \cos(nx)$ beispielsweise die Fourier-Reihe zu einer (nichtkonstanten) 2π-periodischen Treppenfunktion u, dann ist u_k bekanntlich eine Cauchy-Folge bzgl. $\|\cdot\|$ und konvergiert im quadratischen Mittel gegen u, d.h. es gilt $\lim\limits_{k\to\infty} \|u_k - u\| = 0$. Aber die Treppenfunktion u ist nicht stetig, und daher gibt es auch keine stetige 2π-periodische Funktion, gegen die u_k konvergiert.

Die Vollständigkeit eines metrischen Raumes ist äquivalent zum Schachtelungsprinzip. Um dieses zu formulieren, definieren wir den Durchmesser einer nichtleeren Teilmenge M eines metrischen Raumes (X, d) durch

$$\mathrm{diam}(M) := \sup\{d(x,y) \mid x, y \in M\} \quad .$$

Man beachte, dass $\mathrm{diam}(M) = \mathrm{diam}(\overline{M})$ gilt.

Lemma 1.29 *Ein metrischer Raum (X, d) ist genau dann vollständig, wenn das Schachtelungsprinzip gilt, d.h. es zu jeder Folge $A_1 \supset A_2 \supset A_3 \supset \ldots$ ineinander geschachtelter nichtleerer abgeschlossener Teilmengen $A_k \subset X$ mit $\lim\limits_{k\to\infty} \mathrm{diam}(A_k) = 0$ genau einen Punkt $x \in X$ mit $x \in A_k$ für alle $k \in \mathbb{N}$ gibt.*

Beweis: Gilt das Schachtelungsprinzip und ist x_k eine Cauchy-Folge, dann betrachte man den Abschluß $A_K := \overline{\{x_k \mid k \geq K\}}$ der Endstücke der Folge. Offenbar sind die

Teilmengen A_K ineinander geschachtelt, nichtleer und abgeschlossen, und da es zu jedem $\epsilon > 0$ ein K mit $d(x_k, x_l) \leq \epsilon$ für alle $k, l \geq K$ gibt, folgt $\lim\limits_{K\to\infty} \operatorname{diam}(A_K) = 0$. Also gibt es einen Punkt x, der in allen A_K liegt. Wegen $d(x_k, x) \leq \operatorname{diam}(A_K)$ für $k \geq K$ ist dieser Punkt x auch der Grenzwert der Folge x_k.

Konvergiert andererseits jede Cauchy-Folge in X, dann wähle man einfach jeweils einen Punkt $x_k \in A_k$. Die so definierte Folge ist wegen $d(x_k, x_l) \leq \operatorname{diam}(A_K)$ für $k, l \geq K$ und $\lim\limits_{K\to\infty} \operatorname{diam}(A_K) = 0$ eine Cauchy-Folge, sie konvergiert also gegen einen Punkt $x \in X$. Wegen $x_k \in A_K$ für alle $k \geq K$ und der Abgeschlossenheit von A_K gilt $x \in A_K$, und zwar für jedes $K \in \mathbb{N}$. Also haben wir einen gemeinsamen Punkt aller A_k gefunden, und wegen $\lim\limits_{k\to\infty} \operatorname{diam}(A_k) = 0$ kann es auch keinen weiteren solchen Punkt geben. $\square$

Nicht jede Teilmenge M eines vollständigen metrischen Raumes (X, d) ist mit der Metrik $d|_{M\times M}$ selbst wieder vollständig. So ist z.B. jedes offene Intervall $(a, b) \subset \mathbb{R}$ nicht vollständig, denn die Folge $x_k := a + \frac{1}{k}$ ist wegen $d(x_k, x_l) = |\frac{1}{k} - \frac{1}{l}| < \epsilon$ für $k, l \geq \lceil \frac{1}{\epsilon} \rceil$ natürlich eine Cauchy-Folge, konvergiert aber gegen kein Element von (a, b) (sondern in $\mathbb{R}$ gegen a). Das folgende Lemma besagt jedoch, dass sich Vollständigkeit auf abgeschlossene Teilmengen überträgt.

Lemma 1.30 *Jede abgeschlossene Teilmenge M eines vollständigen metrischen Raumes (X, d) ist mit der eingeschränkten Metrik $d|_{M\times M}$ vollständig.*

Beweis: Ist $x_k \in M$ eine Cauchy-Folge, dann konvergiert x_k aufgrund der Vollständigkeit von (X, d) gegen einen Punkt $x \in X$. Aber aufgrund der Abgeschlossenheit von M liegt x schon in M (siehe Definition 1.26), d.h. x_k konvergiert in M. Somit ist M vollständig. $\square$

Es stellt sich die Frage, ob man zu einem metrischen Raum (X, d) auch immer eine vollständigen metrischen Raum $(\tilde{X}, \tilde{d})$ finden kann, der X in einem gewissen Sinne enthält. Solch eine Vervollständigung von X kann man tatsächlich immer konstruieren. Dies klärt insbesondere eine der letzten offenen Fragen aus der Analysis I, ob man nämlich $\mathbb{R}$ nicht nur axiomatisch definieren kann, sondern auch konstruktiv aus $\mathbb{Q}$ gewinnen kann.

Um die Vervollständigung von (X, d) zu konstruieren, betrachten wir die Menge aller Cauchy-Folgen in X und nennen zwei Cauchy-Folgen x_k und y_k äquivalent, wenn $\lim\limits_{k\to\infty} d(x_k, y_k) = 0$ gilt. Dies definiert tatsächlich eine Äquivalenzrelation $\sim$ auf der Menge aller Cauchy-Folgen wegen

- $x_k \sim x_k$ aufgrund von $\lim\limits_{k\to\infty} d(x_k, x_k) = 0$
- $(x_k \sim y_k) \Rightarrow (y_k \sim x_k)$ aufgrund von $\lim\limits_{k\to\infty} d(x_k, y_k) = \lim\limits_{k\to\infty} d(y_k, x_k)$
- $((x_k \sim y_k) \wedge (y_k \sim z_k)) \Rightarrow (x_k \sim z_k)$ aufgrund der Ungleichung $0 \leq d(x_k, z_k) \leq d(x_k, y_k) + d(y_k, z_k) \to 0$ bei $k \to \infty$ und somit $\lim\limits_{k\to\infty} d(x_k, z_k) = 0$.

Nun sei $\tilde{X}$ die Menge aller Äquivalenzklassen $[x_k]$ von Cauchy-Folgen x_k bezüglich der soeben definierten Äquivalenzrelation $\sim$. Auf $\tilde{X}$ kann man dann eine Metrik durch $\tilde{d}([x_k],[y_k]) := \lim_{k\to\infty} d(x_k, y_k)$ definieren.

Tatsächlich existiert dieser Limes in $\mathbb{R}$ für beliebige Cauchy-Folgen x_k und y_k in X, denn wegen $|d(x_k, y_k) - d(x_l, y_l)| \leq d(x_k, x_l) + d(y_k, y_l)$ ist die Folge $d(x_k, y_k)$ eine Cauchy-Folge in $\mathbb{R}$. Darüberhinaus hängt der Grenzwert nicht von der Wahl der Repräsentanten ab, denn gilt $x_k \sim \tilde{x}_k$ und $y_k \sim \tilde{y}_k$, dann folgt wegen

$$|d(\tilde{x}_k, \tilde{y}_k) - d(x_k, y_k)| \leq d(x_k, \tilde{x}_k) + d(y_k, \tilde{y}_k) \to 0$$

für $k \to \infty$ die Gleichheit $\lim_{k\to\infty} d(\tilde{x}_k, \tilde{y}_k) = \lim_{k\to\infty} d(x_k, y_k)$, also ist $\tilde{d}$ wohldefiniert. Desweiteren ist $\tilde{d}$ eine Metrik auf $\tilde{X}$, denn aus $\tilde{d}([x_k],[y_k]) = 0$ folgt $x_k \sim y_k$ nach Definition von $\sim$ und somit $[x_k] = [y_k]$, und die Symmetrie sowie die Dreiecksungleichung für d übertragen sich direkt auf $\tilde{d}$.

Um zu zeigen, dass $(\tilde{X}, \tilde{d})$ vollständig ist, betrachte man eine Cauchy-Folge $[x_k]_n$ in $\tilde{X}$, d.h. eine Folge von Äquivalenzklassen von Cauchy-Folgen in X, für die zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ existiert mit $\tilde{d}([x_k]_m, [x_k]_n) \leq \epsilon/3$ für alle $m, n \geq N$. Wählt man sich zu jedem n einen Repräsentanten x_{kn} von $[x_k]_n$, so ist für festes n die Folge $k \mapsto x_{kn}$ ja eine Cauchy-Folge, also gibt es ein $K(n) \in \mathbb{N}$ mit $d(x_{kn}, x_{ln}) \leq \epsilon/3$ für $k, l \geq K(n)$.

Um sich einen Grenzwert der Cauchy-Folge $[x_k]_n$ in $\tilde{X}$ zu verschaffen, betrachte man einfach die Folge $y_n := x_{K(n)n}$. Diese Folge y_n ist eine Cauchy-Folge in X, denn seien $m, n \geq N$, dann gibt es nach Definition von $\tilde{d}$ ein $l \geq K(n), K(m)$ mit $d([x_{lm}, x_{ln}) \leq \epsilon/3$, und mit diesem gilt

$$d(y_m, y_n) = d(x_{K(m)m}, x_{K(n)n}) \leq d(x_{K(m)m}, x_{lm}) + d(x_{lm}, x_{ln}) + d(x_{ln}, x_{K(n)n}) \leq \epsilon \quad .$$

Und tatsächlich konvergiert die Cauchy-Folge $[x_k]_n$ in $\tilde{X}$ auch gegen die Äquivalenzklasse $[y_k]$. Denn da (wie gerade nachgewiesen) y_k eine Cauchy-Folge in X ist, gibt es ein M mit $d(y_n, y_k) \leq \frac{2}{3}\epsilon$ für $n, k \geq M$. Sei nun $n \geq M$, dann gilt für jedes $k \geq M, K(n)$ die Ungleichung

$$d(x_{kn}, y_k) \leq d(x_{kn}, y_n) + d(y_n, y_k) = d(x_{kn}, x_{K(n)n}) + d(y_n, y_k) \leq \epsilon \quad ,$$

weil $d(x_{kn}, x_{K(n)n}) \leq \epsilon/3$ aufgrund der Wahl von $k \geq K(n)$ gilt. Also gilt $\tilde{d}([x_k]_n, [y_k]) = \lim_{k\to\infty} d(x_{kn}, y_k) \leq \epsilon$ für jedes $n \geq M$, d.h. die Cauchy-Folge $[x_k]_n$ konvergiert in $\tilde{X}$ gegen $[y_k]$. Somit ist $(\tilde{X}, \tilde{d})$ vollständig.

Zu guter letzt kann man noch (X, d) in den so gewonnenen metrischen Raum $(\tilde{X}, \tilde{d})$ isometrisch einbetten, indem man einen Punkt x mit der Äquivalenzklasse der konstanten Folge $\iota(x) := x, x, \ldots$ identifiziert. Die so definierte Abbildung $\iota : X \to \tilde{X}$ erfüllt offensichtlich $\tilde{d}(\iota(x), \iota(y)) = d(x, y)$, d.h. ι ist eine Isometrie und insbesondere injektiv. Somit hat X aufgefasst als Teilmenge $\iota(X)$ von $\tilde{X}$ dieselben metrischen Eigenschaften wie vorher. Darüberhinaus liegt X dicht in $\tilde{X}$, d.h. es gibt zu jedem $[x_k] \in \tilde{X}$ und $\epsilon > 0$ ein $\iota(x)$ mit $\tilde{d}([x_k], \iota(x)) \leq \epsilon$, denn sei K so groß, dass $d(x_k, x_l) \leq \epsilon$ für alle $k, l \geq K$ gilt, dann erfüllt $x := x_K$ diese Bedingung. Wir fassen zusammen:

Satz 1.31 *Zu jedem metrischen Raum (X, d) gibt es einen vollständigen metrischen Raum $(\tilde{X}, \tilde{d})$ und eine Isometrie $\iota : X \to \tilde{X}$, deren Bild $\iota(X)$ in $\tilde{X}$ dicht liegt.*

Man kann desweiteren noch zeigen, dass die Vervollständigung eines metrischen Raumes bis auf Isometrie eindeutig ist.

Aufgabe: *Die Vervollständigung eines normierten Vektorraumes ist ein Banach-Raum, die Vervollständigung eines Euklidischen Vektorraumes ist ein Hilbert-Raum.*

1.6 Der Banachsche Fixpunktsatz

Wie schon erwähnt kann man sich bei Vollständigkeit des zugrundeliegenden Raumes häufig Lösungen von Gleichungen verschaffen. Wir wollen dies im folgenden exemplarisch am Beispiel einer Fixpunktgleichung darstellen.

Sei $f : X \to X$ eine Abbildung von einem metrischen Raum (X, d) in sich selbst. Wir fragen uns, ob f einen Fixpunkt besitzt, d.h. ob es einen Punkt $x^* \in X$ mit $f(x^*) = x^*$ gibt. Der folgende Satz von Banach beantwortet diese Frage positiv, falls f eine Kontraktion und (X, d) vollständig ist.

Satz 1.32 *Sei (X, d) ein vollständiger metrischer Raum und $f : X \to X$ eine Kontraktion, d.h. es gibt ein $L < 1$ mit $d(f(x), f(y)) \leq Ld(x, y)$. Dann gibt es einen eindeutigen Punkt $x^* \in X$ mit $f(x^*) = x^*$.*

Genauer konvergiert die durch $x_{k+1} := f(x_k)$ rekursiv definierte Folge für jeden Startwert $x_0 \in X$ gegen den Fixpunkt x^* von f.

Beweis: Zunächst einmal kann es höchstens einen Fixpunkt x^* von f geben, denn ist $\tilde{x}^*$ ein weiterer Fixpunkt, dann folgt aus $x^* = f(x^*)$ und $\tilde{x}^* = f(\tilde{x}^*)$ wegen

$$d(x^*, \tilde{x}^*) = d(f(x^*), f(\tilde{x}^*)) \leq Ld(x^*, \tilde{x}^*)$$

und $L < 1$ schon $d(x^*, \tilde{x}^*) = 0$ und somit $x^* = \tilde{x}^*$.

Nun wollen wir uns den gesuchten Fixpunkt konstruktiv verschaffen, indem wir beginnend mit einem beliebigen Startwert $x_0 \in X$ durch $x_{k+1} := f(x_k)$ rekursiv eine Folge definieren und zeigen, dass diese eine Cauchy-Folge ist. Dann konvergiert x_k aufgrund der Vollständigkeit von (X, d) nämlich gegen einen Punkt x^*, und daher auch $f(x_k)$ gegen $f(x^*)$ wegen $d(f(x_k), f(x^*)) \leq Ld(x_k, x^*) \to 0$. Aufgrund von

$$x^* = \lim_{k\to\infty} x_{k+1} = \lim_{k\to\infty} f(x_k) = f(x^*)$$

ist somit x^* ein Fixpunkt von f.

Zu zeigen bleibt also nur noch, dass die durch $x_{k+1} := f(x_k)$ rekursiv definierte Folge x_k eine Cauchy-Folge ist. Dazu bemerke man, dass wegen $d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leq Ld(x_n, x_{n-1})$ bei $l > k \geq 0$ die Ungleichung

$$d(x_l, x_k) \leq \sum_{n=k}^{l-1} d(x_{n+1}, x_n) \leq \sum_{n=k}^{l-1} L^{n-k} d(x_{k+1}, x_k) \leq \left(\sum_{n=0}^{l-k-1} L^n \right) L^k d(x_1, x_0)$$

gilt. Nutzt man nun noch aus, dass $\sum_{n=0}^{l-k-1} L^n \leq \sum_{n=0}^{\infty} L^n = \frac{1}{1-L}$ wegen $0 \leq L < 1$ gilt, so erhält man schließlich

$$d(x_l, x_k) \leq \frac{L^k}{1-L} d(x_1, x_0)$$

für alle $l > k \geq 0$. Also ist x_k wirklich eine Cauchy-Folge, denn zu $\epsilon > 0$ findet man wegen $L^k \to 0$ ein $K \in \mathbb{N}$, ab dem $\frac{L^k}{1-L} d(x_1, x_0) \leq \epsilon$ für alle $k \geq K$ gilt, und mit diesem K gilt dann auch $d(x_l, x_k) \leq \epsilon$ für alle $k \geq K$. $\square$

Explizit wollen wir noch den folgenden Spezialfall festhalten, der sich direkt aus dem Banachschen Fixpunktsatz und Lemma 1.30 ergibt.

Korollar 1.33 *Ist $D \subset \mathbb{R}^n$ eine abgeschlossene Teilmenge und $f : D \to \mathbb{R}^n$ eine kontrahierende Selbstabbildung von D, d.h. es gilt $f(D) \subset D$ und es gibt ein $L < 1$ mit $\|f(x) - f(y)\| \leq L\|x - y\|$, dann gibt es einen eindeutigen Fixpunkt $x^* \in D$ von f.*

1.7 Stetigkeit

Stetigkeit von Abbildungen $f : X \to Y$ zwischen metrischen Räumen kann man ganz analog zum Eindimensionalen durch das $\epsilon - \delta$-Kriterium oder das Folgenkriterium definieren.

Definition 1.34 *Eine Abbildung $f : X \to Y$ von einem metrischen Raum (X, d_X) in einen metrischen Raum (Y, d_Y) nennt man stetig im Punkt $a \in X$, wenn*

$$\forall \epsilon > 0 \, \exists \delta > 0 \, : \, \forall x \in X \, : \, (d_X(x, a) \leq \delta \Rightarrow d_Y(f(x), f(a)) \leq \epsilon)$$

gilt, oder wenn äquivalenterweise für jede Folge $x_k \in X$ mit $\lim_{k\to\infty} x_k = a$ auch schon $\lim_{k\to\infty} f(x_k) = f(a)$ gilt.

Die Äquivalenz der beiden Kriterien kann man wortwörtlich wie in der Analysis I zeigen: Ist das $\epsilon - \delta$-Kriterium erfüllt und konvergiert x_k gegen a, dann gibt es zu $\epsilon > 0$ ein $\delta > 0$ mit

$$\forall x \in X \, : \, (d_X(x, a) \leq \delta \Rightarrow d_Y(f(x), f(a)) \leq \epsilon) \quad ,$$

und da x_k gegen a konvergiert, gibt es zum gerade gefundenen $\delta > 0$ ein $K \in \mathbb{N}$ mit $d_X(x_k, a) \leq \delta$ für alle $k \geq K$. Demnach hat man zu $\epsilon > 0$ ein $K \in \mathbb{N}$ mit $d_Y(f(x_k), f(a)) \leq \epsilon$ für alle $k \geq K$ gefunden, d.h. die Bildfolge $f(x_k)$ konvergiert gegen $f(a)$.

Ist umgekehrt das Folgenkriterium erfüllt und nehmen wir an, das $\epsilon - \delta$-Kriterium sei nicht erfüllt, dann gäbe es ein $\epsilon > 0$, für das zu jedem $\delta := \frac{1}{k}$ ein Punkt x_k mit $d_X(x_k, a) \leq \frac{1}{k}$ aber $d_Y(f(x_k), f(a)) > \epsilon$ existierte. Dies würde aber $\lim_{k\to\infty} x_k = a$ bedeuten, während $f(x_k)$ nicht gegen $f(a)$ konvergiert, im Widerspruch zur Gültigkeit des Folgenkriteriums.

Wie schon in der Analysis I nennt man eine Abbildung $f : X \to Y$ stetig, wenn sie in jedem Punkt $a \in X$ stetig ist.

Das folgende Lemma zeigt, dass man Stetigkeit alternativ auch mittels des rein topologischen Begriffs der Umgebung definieren kann. Insbesondere hängt die Stetigkeit einer Abbildung nicht direkt von der Wahl der Metrik auf den Räumen X und Y ab, sondern nur von deren Topologie.

Lemma 1.35 *Eine Abbildung $f : X \to Y$ ist genau dann stetig im Punkt $a \in X$, wenn es zu jeder Umgebung $V \subset Y$ von $f(a)$ eine Umgebung $U \subset X$ von a mit $f(U) \subset V$ gibt.*

Beweis: Wir zeigen die Äquivalenz zum $\epsilon - \delta$-Kriterium.

Gilt das $\epsilon - \delta$-Kriterium und ist $V \subset Y$ eine Umgebung von $f(a)$, dann gibt es eine Kugel mit Radius $\epsilon > 0$ und $B_\epsilon(f(a)) \subset V$. Zu diesem $\epsilon > 0$ gibt es nun ein $\delta > 0$ mit

$$\forall x \in X \, : \, (d_X(x,a) \leq \delta \Rightarrow d_Y(f(x), f(a)) \leq \epsilon) \quad .$$

Dies bedeutet aber nichts anderes als $f(B_\delta(a)) \subset B_\epsilon(f(a))$, und somit hat man zu V die Umgebung $U := B_\delta(a)$ mit $f(U) \subset V$ gefunden.

Gibt es umgekehrt zu jeder Umgebung $V \subset Y$ von $f(a)$ eine Umgebung $U \subset X$ von a mit $f(U) \subset V$, dann auch zur Umgebung $V := B_\epsilon(f(a))$ von $f(a)$. In der so gefundenen Umgebung U von a liegt aber wiederum eine Kugel $B_\delta(a) \subset U$, und somit gilt $f(B_\delta(a)) \subset V = B_\epsilon(f(a))$, was gleichbedeutend ist mit

$$\forall x \in X \, : \, (d_X(x,a) \leq \delta \Rightarrow d_Y(f(x), f(a)) \leq \epsilon) \quad .$$

$\square$

Mit anderen Worten ist eine Abbildung $f : X \to Y$ also genau dann in $a \in X$ stetig, wenn für jede Umgebung V von $f(a)$ das Urbild $f^{-1}(V) := \{x \in X \,|\, f(x) \in V\}$ eine Umgebung von a ist.

Noch deutlicher als in Lemma 1.35 ist der rein topologische Charakter der Stetigkeit an folgendem Lemma zu sehen.

Satz 1.36 *Eine Abbildung $f : X \to Y$ ist genau dann stetig, wenn für jede in Y offene Menge V das Urbild $f^{-1}(V)$ offen in X ist.*

Beweis: Wir weisen die Äquivalenz der angegebenen Bedingung zu der in Lemma 1.35 angegebenen Bedingung nach.

Sei $f : X \to Y$ stetig, $V \subset Y$ offen und $x \in f^{-1}(V)$ ein beliebiger Punkt. Dann gibt es aufgrund der Stetigkeit von f in x eine Umgebung U von x mit $f(U) \subset V$, d.h. $U \subset f^{-1}(V)$. Also ist $f^{-1}(V)$ Umgebung jedes Punktes $x \in f^{-1}(V)$ und somit offen.

Sei umgekehrt $f^{-1}(V) \subset X$ offen für jedes offene $V \subset Y$, sei $x \in X$ und $V' \subset Y$ eine Umgebung von $f(x)$. Als Umgebung von $f(x)$ enthält V' auch eine offene Menge V mit $f(x) \in V \subset V'$. Deren Urbild $U := f^{-1}(V)$ ist offen und enthält x, ist also wegen $f(U) \subset V \subset V'$ eine Umgebung von x, deren Bild in V' liegt. Somit ist f stetig. $\square$

Korollar 1.37 *Für eine stetige Funktion $f : X \to \mathbb{R}$ und einen Wert $c \in \mathbb{R}$ ist die Menge $U := \{x \in X \,|\, f(x) < c\}$ offen und die Menge $A := \{x \in X \,|\, f(x) \leq c\}$ abgeschlossen.*

Beweis: U ist das Urbild der offenen Menge $(-\infty, c)$, A ist das Urbild der abgeschlossenen Menge $(-\infty, c]$. □

Als nächstes wollen wir einige Beispiele stetiger Abbildungen kennenlernen.

Beispiel 1.38

- *Jede konstante Abbildung $f : X \to Y$, $f(x) := b$, mit einem fest gewählten Punkt $b \in Y$ ist stetig.*

- *Die Norm $\|\cdot\| : X \to \mathbb{R}$ auf einem normierten Vektorraum X ist stetig in jedem Punkt $a \in X$, denn wegen $|\,\|x\| - \|a\|\,| \leq \|x - a\|$ folgt bei $x_k \to a$ auch schon $\|x_k\| \to \|a\|$ für $k \to \infty$.*

- *Die Addition $+ : X \times X \to X$ eines normierten Vektorraumes X ist stetig in jedem Punkt $(x^*, y^*) \in X \times X$, denn wegen*

  $$\|(x+y) - (x^* + y^*)\|_X \leq \|x - x^*\|_X + \|y - y^*\|_X = \|(x,y) - (x^*, y^*)\|_{X \times X}$$

  folgt bei $(x_k, y_k) \to (x^*, y^*)$ auch schon $x_k + y_k \to x^* + y^*$ für $k \to \infty$.

  Analog dazu ist auch die skalare Multiplikation $\cdot : \mathbb{R} \times X \to X$ stetig wegen

  $$\|\lambda x - \lambda^* x^*\|_X \leq \|\lambda x - \lambda^* x\|_X + \|\lambda^* x - \lambda^* x^*\|_X = |\lambda - \lambda^*| \|x\|_X + |\lambda^*| \|x - x^*\|_X \,,$$

  Lemma 1.24 und der zuvor bewiesenen Stetigkeit der Norm. Insbesondere ist auch die Multiplikation $\cdot : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ reeller Zahlen stetig.

- *Die Division $\frac{\cdot}{\cdot} : \mathbb{R} \times \mathbb{R} \setminus \{0\} \to \mathbb{R}$ reeller Zahlen ist stetig in jedem Punkt (x^*, y^*) mit $y^* \neq 0$, denn aus*

  $$\left| \frac{x_k}{y_k} - \frac{x^*}{y^*} \right| = \left| \frac{x_k y^* - x^* y_k}{y_k y^*} \right| \leq \frac{1}{|y_k y^*|} \left(|x_k - x^*||y^*| + |y_k - y^*||x^*| \right)$$

  folgt bei $x_k \to x^*$ und $y_k \to y^*$ auch die Konvergenz von $\frac{x_k}{y_k}$ gegen $\frac{x^*}{y^*}$ (wobei man wieder Lemma 1.24 benutzt).

- *Die Projektion $\pi_X : X \times Y \to X$, $(x, y) \mapsto x$, eines Produkt $X \times Y$ auf den Faktor X ist stetig in (x^*, y^*), denn zu einer Umgebung U von x^* ist $U \times Y$ eine Umgebung (x^*, y^*) mit $\pi_X(U \times Y) \subset U$ (analog für die Projektion $\pi_Y(x, y) := y$ auf Y)*

  Insbesondere sind die Koordinatenfunktionen $\mathbb{R}^n \to \mathbb{R}$, $x \mapsto x_i$, stetig.

Die folgenden Sätze dienen dazu, die Stetigkeit von Abbildungen zu beweisen, die aus elementaren stetigen Abbildungen zusammengesetzt sind.

Lemma 1.39 *Eine Abbildung $f = (f_1, f_2) : X \to Y \times Z$ ist genau dann stetig im Punkt $a \in X$, wenn die Komponentenabbildungen $f_1 : X \to Y$ und $f_2 : X \to Z$ stetig in a sind.*

Beweis: Dies folgt einerseits direkt aus dem Folgenkriterium und Lemma 1.24.

Andererseits kann man die Aussage auch rein topologisch wie folgt beweisen: Sei f in a stetig, V eine Umgebung von $f_1(a)$ und W eine Umgebung von $f_2(a)$. Dann ist $V \times W$ eine Umgebung von $f(a) = (f_1(a), f_2(a))$, und somit gibt es aufgrund der Stetigkeit von f eine Umgebung U von a mit $f(U) \subset V \times W$. Mit diesem U gilt also $f_1(U) \subset V$ und $f_2(U) \subset W$, d.h. f_1 und f_2 sind stetig in a.

Umgekehrt seien f_1 und f_2 stetig. In jeder Umgebung von $f(a) = (f_1(a), f_2(a))$ liegt dann eine Umgebung der Form $V \times W$, und zu diesen gibt es Umgebungen U_1, U_2 von a mit $f_1(U_1) \subset V$ und $f_2(U_2) \subset W$. Also ist $U_1 \cap U_2$ eine Umgebung von a, die unter f in $V \times W$ hinein und somit auch in die ursprüngliche Umgebung hinein abgebildet wird. Daher ist f stetig. $\square$

Beispielsweise ist nach diesem Lemma für eine stetige Abbildung $f : X \to Y$ und einen festen Punkt $z \in Z$ auch die Abbildung $X \to Y \times Z$, $x \mapsto (f(x), z)$, stetig, denn $f : X \to Y$ und die konstante Abbildung $x \mapsto z$ sind stetig.

Satz 1.40 *Ist die Abbildung $f : X \to Y$ im Punkt $a \in X$ stetig und die Abbildung $g : Y \to Z$ im Punkt $f(a) \in Y$ stetig, dann ist auch die Hintereinanderausführung $g \circ f : X \to Z$ im Punkt $a \in X$ stetig.*

Beweis: Sei W eine Umgebung von $g(f(a))$, dann gibt es aufgrund der Stetigkeit von g eine Umgebung V von $f(a)$ mit $g(V) \subset W$, und zu V aufgrund der Stetigkeit von f eine Umgebung U von a mit $f(U) \subset V$. Diese Umgebung U von a erfüllt dann also

$$(g \circ f)(U) = g(f(U)) \subset g(V) \subset W \quad ,$$

und somit ist $g \circ f$ stetig in a. $\square$

Korollar 1.41 *Sind $f, g : X \to \mathbb{R}$ stetige Funktionen, dann sind auch $f + g$ und $f \cdot g$ stetig auf X. Gilt darüberhinaus $\forall x \in X \ : \ g(x) \neq 0$, so ist auch $\frac{f}{g}$ stetig auf X.*

Beweis: Die Abbildung $(f, g) : X \to \mathbb{R} \times \mathbb{R}$ ist nach Lemma 1.39 stetig, und die angegebenen Funktionen sind nichts anderes als die Verknüpfungen $+ \circ (f, g)$, $\cdot \circ (f, g)$ und $\div \circ (f, g)$ mit den in Beispiel 1.38 explizit diskutierten Abbildungen $+$, $\cdot$ und $\div$. $\square$

Beispielsweise folgt durch wiederholte Anwendung dieses Korollars, dass jede polynomiale Funktion $p : \mathbb{R}^n \to \mathbb{R}$ vom Grad N in n Variablen,

$$p(x_1, \ldots, x_n) := \sum_{k_1 + \cdots + k_n \leq N} a_{k_1, \ldots, k_n} x_1^{k_1} \cdot \cdots \cdot x_n^{k_n} \quad ,$$

stetig ist, denn p entsteht gerade aus Verknüpfungen von Koordinatenfunktionen $x \mapsto x_i$, Addition $+$, Multiplikation $\cdot$ und konstanten Funktionen. Ebenso sind rationale Funktionen $f := \frac{p}{q}$ in n Variablen (wobei p, q polynomiale Funktionen sind) stetig in denjenigen Punkten $x \in \mathbb{R}^n$ mit $q(x) \neq 0$. Andere Beispiele für stetige Funktionen in mehreren Variablen kann man z.B. durch Verknüpfung der stetigen Funktionen exp, sin, cos auf $\mathbb{R}$ mit Koordinatenfunktionen erhalten.

Beispiel 1.42 *Funktionen $f : X \times Y \to Z$ sind* ***nicht*** *bereits dann stetig in einem Punkt (x^*, y^*), wenn $x \mapsto f(x, y^*)$ und $y \mapsto f(x^*, y)$ stetig sind, letzteres ist eine wesentlich schwächere Eigenschaft! Dies liegt daran, dass nach Lemma 1.24 auch Folgen gegen (x^*, y^*) konvergieren, die nicht auf den Achsen verlaufen.*

Beispielsweise gilt für die durch $f(x, y) := \frac{2xy}{x^2+y^2}$ bei $(x, y) \neq (0, 0)$ und $f(0, 0) := 0$ definierte Funktion zwar $f(x, 0) = 0 = f(0, y)$ für alle $x, y \in \mathbb{R}$, aber f ist nicht stetig im Punkt $(0, 0)$, denn es gilt $f(x, x) = 1$, d.h. entlang der Diagonalen konvergiert $f(x, y)$ für $(x, y) \to (0, 0)$ nicht gegen Null. Also ist f nicht stetig im Punkt $(0, 0)$.

Homöomorphismen Ist die Abbildung $f : X \to Y$ zwischen metrischen Räumen X, Y bijektiv und stetig, so muss ihre Umkehrabbildung $f^{-1} : Y \to X$ nicht unbedingt stetig sein. Beispielsweise hat die stetige und bijektive Abbildung $f : [0, 2\pi) \to S^1$, $t \mapsto (\cos(t), \sin(t))$, auf den Kreis $S^1 := \{(x, y) \in \mathbb{R}^2 \,|\, \|(x, y)\|_2 = 1\}$ keine stetige Umkehrabbildung.

Eine stetige Bijektion $f : X \to Y$, deren Umkehrabbildung $f^{-1} : Y \to X$ auch stetig ist, heißt ein Homöomorphismus zwischen X und Y, und die Räume X und Y werden in diesem Fall homöomorph genannt. Für solch eine Abbildung ist $x_k \to x$ äquivalent zu $f(x_k) \to f(x)$, d.h. homöomorphe Räume haben dieselben topologischen Eigenschaften, auch wenn sie unterschiedliche metrische Eigenschaften besitzen können.

Beispiel 1.43 *Die offene Kugel $B_1(0)$ in einem normierten Vektorraum X ist homöomorph zu X, denn die Abbildung $f : B_1(0) \to X$, $x \mapsto \frac{x}{1-\|x\|}$ ist*

- *stetig, da die Identität und die Norm stetig sind und $\|x\| < 1$ für $x \in B_1(0)$ gilt,*
- *bijektiv, da $f(x) = y \Leftrightarrow x = \frac{y}{1+\|y\|}$ für $x \in B_1(0)$ und $y \in X$ gilt,*
- *und sogar ein Homöomorphismus, da auch die Umkehrabbildung $f^{-1}(y) := \frac{y}{1+\|y\|}$ stetig ist (aufgrund der Stetigkeit der Identität und der Norm).*

Gleichmäßige Stetigkeit und Lipschitz-Stetigkeit Während Stetigkeit ein rein topologischer Begriff ist, hängen die beiden folgenden Begriffe direkt von der Wahl der Metrik ab.

Definition 1.44 *Eine Abbildung $f : X \to Y$ zwischen metrischen Räumen (X, d_X) und (Y, d_Y) heißt*

- *gleichmäßig stetig, wenn*

$$\forall \epsilon > 0 \, \exists \delta > 0 \; : \; \forall x, x' \in X \; : \; (d_X(x, x') \leq \delta \Rightarrow d_Y(f(x), f(x')) \leq \epsilon) \quad ,$$

- *Lipschitz-stetig, wenn* $\exists L \geq 0 : \forall x, x' \in X : d(f(x), f(x')) \leq L d_X(x, x')$.

Bei gleichmäßiger Stetigkeit ist also im Gegensatz zur Stetigkeit die Zahl $\delta = \delta(\epsilon)$ unabhängig von dem Punkt, in dem Stetigkeit getestet werden soll. Bei Lipschitz-Stetigkeit ist der Anstieg von f durch die Lipschitz-Konstante L beschränkt. Insbesondere sind Kontraktionen gerade Abbildungen mit Lipschitz-Konstante $L < 1$.

Lemma 1.45 *Jede Lipschitz-stetige Abbildung ist gleichmäßig stetig, jede gleichmäßig stetige Abbildung ist stetig.*

Beweis: Hat f die Lipschitz-Konstante L, so wähle man zu $\epsilon > 0$ einfach $\delta := \epsilon/L$, dann folgt aus $d_X(x, x') \leq \delta$ schon $d_Y(f(x), f(x')) \leq L d_X(x, x') \leq L\delta \leq \epsilon$.

Ist f gleichmäßig stetig, so gilt insbesondere

$$\forall a \in X\, \forall \epsilon > 0\, \exists \delta > 0 : \forall x \in X : (d_X(x, a) \leq \delta \Rightarrow d_Y(f(x), f(a)) \leq \epsilon) ,$$

also ist f stetig in jedem Punkt $a \in X$. $\square$

Beispiel 1.46

- *Die Norm* $\|\cdot\| : X \to \mathbb{R}$ *auf einem normierten Vektorraum* X *hat wegen der Gültigkeit von* $|\,\|x\| - \|y\|\,| \leq \|x - y\|$ *die Lipschitz-Konstante Eins.*

- *Die Metrik* $d : X \times X \to \mathbb{R}$ *auf einem metrischen Raum* X *hat als Abbildung vom mit der Produktmetrik versehenen Raum* $X \times X$ *nach* $\mathbb{R}$ *wegen* $|d(x, y) - d(x', y')| \leq d(x, x') + d(y, y')$ *die Lipschitz-Konstante Eins und ist insbesondere stetig.*

Grenzwerte von Abbildungen Analog zum Eindimensionalen kann man nicht nur über die Stetigkeit einer Abbildung $f : X \to Y$ zwischen metrischen Räumen X, Y in einem Punkt $a \in X$ sprechen, sondern auch über den Grenzwert $\lim_{a \neq x \to a} f(x)$ einer Abbildung $f : D \to Y$ $(D \subset X)$ in einem Berührpunkt a von $D \setminus \{a\}$. Solch einen Punkt a nennt man auch einen Häufungspunkt von D, und a ist genau dann Häufungspunkt von D, wenn jede Umgebung von a unendlich viele Punkte aus D enthält.

Definition 1.47 *Der Punkt* $y \in Y$ *heißt Grenzwert der Abbildung* $f : D \to Y$ *im Häufungspunkt* a *von* D*, falls es zu jeder Umgebung* V *von* y *eine Umgebung* $U \subset D$ *von* a *mit* $f(U \setminus \{a\}) \subset V$ *gibt, und in diesem Fall schreibt man* $\lim\limits_{a \neq x \to a} f(x) = y$.

Die Eindeutigkeit von y folgt direkt aus dem Hausdorffschen Trennungsaxiom, das im metrischen Raum Y automatisch gilt.

Im Falle der Existenz des Grenzwertes y von f in a erhält man durch die Definition $f(a) := y$ eine in a stetige Funktion $f : D \cup \{a\} \to \mathbb{R}$, die stetige Fortsetzung von f (bei $a \notin D$) oder stetige Abänderung von f (bei $a \in D$) genannt wird.

Gleichmäßige Konvergenz von Folgen von Abbildungen Wie im Eindimensionalen kann man auch Folgen von stetigen Abbildungen $f_k : X \to Y$ zwischen metrischen Räumen X, Y betrachten und sich fragen, unter welchen Voraussetzungen diese eine stetige Grenzfunktion besitzen.

Definition 1.48 *Eine Folge $f_k : X \to Y$ von Abbildungen zwischen metrischen Räumen (X, d_X), (Y, d_Y) heißt gleichmäßig konvergent gegen $f : X \to Y$, falls die Grenzwertbeziehung $\lim\limits_{k\to\infty} \left(\sup\limits_{x\in X} d_Y(f_k(x), f(x)) \right) = 0$ gilt.*

Ist Y vollständig, so konvergiert f_k genau dann gleichmäßig, wenn das Cauchy-Kriterium

$$\forall \epsilon \, \exists K \in \mathbb{N} \, : \, \forall k, l \geq K \, : \sup_{x\in X} d_Y(f_k(x), f_l(x)) \leq \epsilon$$

erfüllt ist, und wie in der Analysis I gilt auch hier:

Lemma 1.49 *Die Grenzabbildung $f(x) = \lim\limits_{k\to\infty} f_k(x)$ einer gleichmäßig konvergenten Folge $f_k : X \to Y$ stetiger Abbildungen zwischen metrischen Räumen X, Y ist stetig.*

Beweis: Sei $\epsilon > 0$ und $a \in X$. Wegen der gleichmäßigen Konvergenz von f_k gegen f gibt es ein $K \in \mathbb{N}$ mit $d(f_k(x), f(x)) \leq \epsilon/3$ für alle $x \in X$ und alle $k \geq K$. Sei nun $k \geq K$ fest gewählt, dann gibt es aufgrund der Stetigkeit von f_k in a eine Umgebung U von a, so dass aus $x \in U$ die Ungleichung $d_Y(f_k(x), f_k(a)) \leq \epsilon/3$ folgt. Also folgt aus $x \in U$ auch

$$d_Y(f(x), f(a)) \leq d_Y(f(x), f_k(x)) + d_Y(f_k(x), f_k(a)) + d_Y(f_k(a), f(a)) \leq \epsilon \quad ,$$

was die Stetigkeit von f in a beweist. □

1.8 Stetige lineare Abbildungen

Seien X, Y normierte (möglicherweise unendlichdimensionale) Vektorräume. Wir betrachten im folgenden lineare Abbildungen $A : X \to Y$, die also die Gleichungen $A(x + x') = A(x) + A(x')$ und $A(\lambda x) = \lambda A(x)$ erfüllen.

Jede lineare Abbildung $A : X \to Y$ zwischen endlich-dimensionalen normierten Vektorräumen X, Y ist stetig, denn nach Korollar 1.20 sind alle Normen auf ihnen äquivalent, also kann man sie einfach durch $X \cong \mathbb{R}^m$ bzw. $Y \cong \mathbb{R}^n$ mit dem $\mathbb{R}^n$ identifizieren und die Stetigkeit der induzierten linearen Abbildung $A : \mathbb{R}^m \to \mathbb{R}^n$ testen. Die Komponenten dieser Abbildung sind aber einfach polynomiale Funktionen $x \mapsto \sum\limits_{j=1}^{m} a_{ij} x_j$ ersten Grades und somit stetig.

Im allgemeinen müssen aber lineare Abbildungen nicht stetig sein.

Lemma 1.50 *Eine lineare Abbildung $A : X \to Y$ zwischen normierten Vektorräumen X, Y ist genau dann stetig, wenn es eine Konstante $L < \infty$ gibt mit*

$$\forall x \in X \, : \, \|Ax\|_Y \leq L \|x\|_X \quad .$$

Beweis: Sei A stetig, dann ist A auch stetig im Nullpunkt. Also gibt es zu $\epsilon := 1$ ein $\delta > 0$ mit $\|Ax\|_Y \leq 1$ für alle $x \in X$ mit $\|x\|_X \leq \delta$. Daher gilt $\|A\left(\frac{\delta}{\|x\|_X}x\right)\|_Y \leq 1$ für alle $x \neq 0$ (denn $\frac{\delta}{\|x\|_X}x$ hat die Norm δ), und somit folgt $\|Ax\|_Y \leq \frac{1}{\delta}\|x\|_X$ für alle $x \in X$.

Umgekehrt folgt aus $\|Ax\|_Y \leq L\|x\|_X$ für alle $x \in X$ durch die Substitution $x \mapsto x - x'$ auch die Ungleichung $\|Ax - Ax'\|_Y \leq L\|x - x'\|_X$ für alle $x, x' \in X$, d.h. A ist Lipschitz-stetig mit Lipschitz-Konstante L, und insbesondere stetig. □

Aus dem Beweis des Lemmas ergibt sich insbesondere, dass eine lineare Abbildung $A : X \to Y$ zwischen beliebigen normierten Vektorräumen genau dann stetig ist, wenn sie Lipschitz-stetig ist. Die kleinstmögliche Lipschitz-Konstante L von A nennt man die Operatornorm von A und bezeichnet sie mit $\|A\|_{L(X,Y)}$. Nach dem Lemma ist die Operatornorm alternativ durch

$$\|A\|_{L(X,Y)} = \sup_{0 \neq x \in X} \frac{\|Ax\|_Y}{\|x\|_X}$$

gegeben. Tatsächlich ist die so definierte Funktion $\|\cdot\|_{L(X,Y)}$ eine Norm auf dem Vektorraum $L(X,Y)$ aller stetigen linearen Abbildungen von X nach Y.

Lemma 1.51 *Die Funktion $\|\cdot\|_{L(X,Y)}$ ist eine Norm auf dem Vektorraum $L(X,Y)$ aller stetigen linearen Abbildungen von X nach Y und erfüllt $\|Ax\|_Y \leq \|A\|_{L(X,Y)}\|x\|_X$ für alle $x \in X$.*

Sind $A : X \to Y$ und $B : Y \to Z$ lineare stetige Abbildungen zwischen normierten Vektorräumen X, Y, Z, so gilt zusätzlich

$$\|B \circ A\|_{L(X,Z)} \leq \|B\|_{L(Y,Z)}\|A\|_{L(X,Y)} \quad .$$

Beweis: Die Funktion $\|\cdot\|_{L(X,Y)}$ hat alle Eigenschaften einer Norm, denn

- gilt $\|A\| = 0$, dann folgt $\|Ax\| = 0$ für alle $x \in X$ und somit $A = 0$,
- $\|Ax\|_Y \leq L\|x\|_X$ ist bei $\lambda \neq 0$ äquivalent zu $\|\lambda Ax\|_Y \leq |\lambda|L\|x\|_X$, also gilt $\|\lambda A\|_{L(X,Y)} = |\lambda|\|A\|_{L(X,Y)}$,
- aus $\|(A + A')(x)\|_Y \leq \|Ax\|_Y + \|A'x\|_Y \leq (\|A\|_{L(X,Y)} + \|A'\|_{L(X,Y)})\|x\|_X$ für alle $x \in X$ folgt $\|A + A'\|_{L(X,Y)} \leq \|A\|_{L(X,Y)} + \|A'\|_{L(X,Y)}$,

und die Ungleichung $\|Ax\|_Y \leq \|A\|_{L(X,Y)}\|x\|_X$ besagt gerade, dass $\|A\|_{L(X,Y)}$ eine Lipschitz-Konstante von A ist.

Darüberhinaus folgt aus $\|B(Ax)\|_Z \leq \|B\|_{L(Y,Z)}\|Ax\|_Y \leq \|B\|_{L(Y,Z)}\|A\|_{L(X,Y)}\|x\|_X$ sofort die Abschätzung $\|B \circ A\|_{L(X,Z)} \leq \|B\|_{L(Y,Z)}\|A\|_{L(X,Y)}$. □

Meistens läßt man die Indizes an den Normen weg, da intuitiv klar ist, welche Norm man zu betrachten hat.

Da die Operatornorm $\|A\|$ anschaulich den Dehnungskoeffizienten der linearen Abbildung A angibt, ist sie auch im Fall endlich-dimensionaler Vektorräume eine interessante

Größe. Später werden wir insbesondere im endlich-dimensionalen Fall Stetigkeit von Abbildungen in den Raum $L(X, Y)$ testen müssen. Dabei hilft das folgende Lemma.

Lemma 1.52 *Für endlich-dimensionale Vektorräume Y, Z ist $f : X \to L(Y, Z)$ genau dann stetig, wenn für jedes $y \in Y$ (oder auch nur für y aus einer Basis von Y) die Abbildungen $X \to Z$, $x \mapsto f(x)(y)$, stetig sind.*

Beweis: Wir zeigen, dass die Konvergenz $A_k \to A$ einer Folge in $L(Y, Z)$ äquivalent zu $\forall y \in Y \ : \ A_k(y) \to A(y)$ ist. Dann folgt die Aussage des Lemmas nach dem Folgenkriterium.

Die eine Richtung ist trivial, da aus der Konvergenz $\|A_k - A\| \to 0$ natürlich die Konvergenz $\|A_k(y) - A(y)\| \leq \|A_k - A\|\|y\| \to 0$ folgt.

Zum Beweis der anderen Richtung konstruieren wir eine neue Norm auf $L(Y, Z)$. Sei $e_1, \ldots, e_m$ eine Basis von Y. Wir definieren für $y = \sum_{i=1}^{m} y_i e_i$ die neue Norm $\|y\|' := \sum_{i=1}^{n} |y_i|$ auf Y. Wegen

$$\|Ay\| \leq \sum_{i=1}^{n} |y_i|\|Ae_i\| \leq \|y\|' \max_{i=1,\ldots,n} \|Ae_i\|$$

erfüllt die Operatonorm $\|A\|'$ von A bzgl. der neuen Norm $\|\cdot\|'$ auf Y die Ungleichung $\|A\|' \leq \max_{i=1,\ldots,n} \|Ae_i\|$. Insbesondere folgt aus $A_k e_i \to Ae_i$ in der neuen Operatornorm die Konvergenz $\|A_k - A\|' \to 0$. Aufgrund der Endlich-Dimensionalität von $L(Y, Z)$ sind die neue und die alte Operatornorm aber äquivalent, und somit folgt auch $\|A_k - A\| \to 0$ in der alten Operatornorm. $\square$

Abschließend wollen wir uns noch speziell mit dem Raum $L(X, X)$ der stetigen linearen Endomorphismen $A : X \to X$ eines Banach-Raumes X beschäftigen.

Aufgrund der Vollständigkeit von X ist auch $L(X, X)$ vollständig, denn gilt die Ungleichung $\|A_k - A_l\| \leq \epsilon$ für $k, l \geq K$, dann auch $\|A_k x - A_l x\| \leq \epsilon\|x\|$ für jedes $x \in X$, und somit ist $A_k x$ für jedes $x \in X$ eine Cauchy-Folge, konvergiert im vollständigen Raum X also gegen einen Punkt Ax. Die so definierte Abbildung $A : X \to X$ ist offensichtlich linear. Außerdem ist sie wegen $\|A_k - A\| = \lim_{l \to \infty} \|A_k - A_l\| \leq \epsilon$ für alle $k \geq K$ auch Grenzwert der A_k bzgl. der Operatornorm, und wegen $\|A\| \leq \|A_k - A\| + \|A_k\| < \infty$ auch stetig. Dies beweist die Vollständigkeit von $L(X, X)$.

Nun ist aber $L(X, X)$ nicht nur ein Banach-Raum, sondern man kann stetige lineare Abbildungen $A, B : X \to X$ auch hintereinanderschalten. Statt $A \circ B$ schreibt man einfach AB, und da diese Operation bilinear und assoziativ ist, hat sie alle Eigenschaften, die man von einer Multiplikation verlangt. Zusätzlich erfüllt die Norm auf $L(X, X)$ noch die Ungleichung $\|AB\| \leq \|A\|\|B\|$, und dies bringt $L(X, X)$ den Namen Banach-Algebra ein. Wir fassen zusammen:

Lemma 1.53 *Für einen Banach-Raum X ist $L(X, X)$ zusammen mit der Hintereinanderausführung von stetigen linearen Abbildungen eine Banach-Algebra.*

Während man schon in beliebigen normierten Vektorräumen X über die Konvergenz von Reihen $\sum_{k=0}^{\infty} x_k$ zu einer Folge x_k in X sprechen kann, indem man analog zum Eindimensionalen die Folge der Partialsummen $S_N := \sum_{k=0}^{N} x_k$ betrachtet, kann man in Banach-Räumen zusätzlich noch die Konvergenz absolut konvergenter Reihen nachweisen. Dabei nennt man eine Reihe $\sum_{k=0}^{\infty} x_k$ in X absolut konvergent, wenn die Reihe $\sum_{k=0}^{\infty} \|x_k\|$ in $\mathbb{R}$ konvergiert. Dann gibt es aber zu jedem ϵ ein K mit $\sum_{k=M+1}^{N} \|x_k\| \leq \epsilon$ für $N > M \geq K$, und daher ist die Folge S_N eine Cauchy-Folge in X wegen

$$\|S_N - S_M\| \leq \sum_{k=M+1}^{N} \|x_k\| \leq \epsilon$$

für $N > M \geq K$. Aufgrund der Vollständigkeit von X konvergiert also S_N und somit auch die Reihe. Daher haben wir bewiesen:

Lemma 1.54 *Jede absolut konvergente Reihe in einem Banach-Raum konvergiert.*

In einer Banach-Algebra wie $L(X, X)$ kann man nun sogar auch Potenzreihen $\sum_{k=0}^{\infty} c_k A^k$ betrachten, denn man hat ja eine Multiplikation zur Verfügung, um A^k zu definieren. Diese Reihen konvergieren absolut für alle A, deren Norm $\|A\|$ kleiner als der Konvergenzradius R der eindimensionalen Potenzreihe $x \mapsto \sum_{k=0}^{\infty} c_k x^k$ ist, denn es gilt $\|c_k A^k\| \leq |c_k| \|A\|^k$ und somit $\sum_{k=0}^{\infty} \|c_k A^k\| \leq \sum_{k=0}^{\infty} |c_k| \|A\|^k < \infty$ bei $\|A\| < R$. Darüberhinaus ist die so definierte Abbildung $f : B_R(0) \subset L(X, X) \to L(X, X)$, $A \mapsto \sum_{k=0}^{\infty} c_k A^k$, auf jeder abgeschlossenen Kugel $\overline{B_r(0)}$ $(r < R)$ Lipschitz-stetig, denn aus $A^k - B^k = \sum_{i=0}^{k-1} A^{k-1-i}(A - B)B^i$ folgt $\|A^k - B^k\| \leq \|A - B\| k r^{k-1}$ bei $\|A\|, \|B\| \leq r$ und somit $\|f(A) - f(B)\| \leq \left(\sum_{k=1}^{\infty} k |c_k| r^{k-1} \right) \|A - B\|$ mit Lipschitz-Konstanter $\sum_{k=1}^{\infty} k |c_k| r^{k-1} < \infty$. Beim Beweis des folgenden Satzes werden diese Resultate speziell auf die Neummansche Reihe $f(A) := \sum_{k=0}^{\infty} A^k$ angewendet.

Satz 1.55 *Die Gruppe $GL(X)$ der invertierbaren stetigen linearen Abbildungen auf einem Banach-Raum X ist eine offene Teilmenge von $L(X, X)$, und die Inversion $A \mapsto A^{-1}$ auf $GL(X)$ ist stetig.*

Beweis: Zunächst einmal hat für jedes $A \in L(X, X)$ mit $\|A\| < 1$ das Element $\mathrm{Id} - A \in L(X, X)$ ein Inverses $(\mathrm{Id} - A)^{-1}$. Genauer ist das Inverse mittels der geometrischen Reihe durch $(\mathrm{Id} - A)^{-1} = \sum_{k=0}^{\infty} A^k =: f(A)$ gegeben, denn

$$f(A)(\mathrm{Id} - A) = \sum_{k=0}^{\infty} A^k - \sum_{k=1}^{\infty} A^k = \mathrm{Id}$$

und analog $(\mathrm{Id} - A)f(A) = \mathrm{Id}$ gelten.

Ist nun $A \in GL(X) \subset L(X, X)$ invertierbar, dann liegt auch die Kugel um A mit Radius $r < 1/\|A\|^{-1}$ in $GL(X)$, denn wegen $\|\mathrm{Id} - A^{-1}B\| \leq \|A\|^{-1}\|A - B\| < 1$ für jedes $B \in B_r(A)$ ist $\mathrm{Id} - (\mathrm{Id} - A^{-1}B) = A^{-1}B$ invertierbar und daher auch B. Also ist $GL(X)$ offen, und die Inversion ist auf $B_r(A)$ durch $B \mapsto f(\mathrm{Id} - A^{-1}B)A^{-1}$ gegeben. Da die Abbildungen $B \mapsto \mathrm{Id} - A^{-1}B$, f und $C \mapsto CA^{-1}$ Lipschitz-stetig sind, ist also auch die Inversion Lipschitz-stetig. □

1.9 Kompaktheit

Neben der Vollständigkeit ist die Kompaktheit eines metrischen Raumes eine weitere wichtige Eigenschaft, um Existenzaussagen beweisen zu können. Um Kompaktheit eines metrischen Raumes X definieren zu können, führen wir zunächst offene Überdeckungen ein: Eine Familie U_i ($i \in I$ mit einer beliebig großen Indexmenge I) von offenen Mengen heißt offene Überdeckung von X, wenn $X \subset \bigcup_{i \in I} U_i$ gilt, d.h. wenn jeder Punkt von X zumindest in einer der offenen Mengen U_i liegt.

Definition 1.56 *Ein metrischer Raum X heißt kompakt, falls es in* ***jeder*** *offenen Überdeckung U_i $(i \in I)$ von X eine endliche Teilüberdeckung gibt, d.h. endlich viele Indizes $i_1, \ldots, i_n$ mit $X \subset \bigcup_{j=1}^{n} U_{i_j}$ existieren.*

Natürlich hätte man in der Definition auch einfach $=$ statt $\subset$ schreiben können, aber so verallgemeinert sie sich sofort auf den Fall, in dem X Teilmenge eines größeren Raumes ist und die U_i offene Teilmengen des größeren Raumes sind (beachte dabei, dass die offenen Mengen der Relativtopologie von X genau die Schnitte von X mit offenen Mengen im größeren Raum sind).

Man bemerke außerdem den rein topologischen Charakter der Definition, Kompaktheit eines metrischen Raumes X hängt also nicht von der Wahl der Metrik ab, sondern nur von der Topologie von X. Nun wollen wir kompakte metrische Räume charakterisieren.

Lemma 1.57 *Ein metrischer Raum X ist genau dann kompakt, wenn jede Folge in X eine konvergente Teilfolge besitzt.* [1].

[1]Die Äquivalenz von Kompaktheit und Folgenkompaktheit (so wird allgemein die im Satz von Bolzano-Weierstraß bewiesene Eigenschaft genannt, dass jede Folge in X eine konvergente Teilfolge besitzt) gilt nicht für beliebige topologische Räume, sondern nur für metrische Räume.

Beweis: Sei X kompakt und x_k eine Folge in X. Wir betrachten die der Folge x_k zugrundeliegende Menge $A := \{x_k \,|\, k \in \mathbb{N}\}$. Ist die Menge A endlich, so hat x_k sogar eine konstante (und somit konvergente) Teilfolge.

Habe die Menge A nun unendlich viele Elemente, dann besitzt A auch einen Häufungspunkt in X. Ansonsten würde es nämlich um jeden Punkt $x \in X$ eine offene Umgebung $U(x)$ geben, die nur endlich viele Punkte aus A enthält. Die offenen Menge $U(x)$ würden dann X überdecken, und aufgrund der Kompaktheit von X würde X schon von endlich vielen Mengen $U(x_1), \ldots, U(x_n)$ überdeckt werden. Somit würde A aber auch nur endlich viele Punkte enthalten, Widerspruch.

Sei also $a \in X$ ein Häufungspunkt von A. Dann gibt es in jeder Kugel um a unendlich viele Elemente aus A, und insbesondere gibt es eine streng monoton wachsende Folge k_n von Indizes mit $d(x_{k_n}, a) \leq \frac{1}{n}$, so dass die Teilfolge x_{k_n} wie gewünscht einen Grenzwert hat (genauer ist a ihr Grenzwert). Also folgt aus der Kompaktheit von X, dass jede Folge in X eine konvergente Teilfolge besitzt.

Um die Umkehrung zu beweisen, nehmen wir an, dass jede Folge in X eine konvergente Teilfolge besitzt. Dann gibt es zu jedem $\epsilon > 0$ insbesondere endlich viele Punkte $x_1, \ldots, x_N$ mit $X \subset \bigcup\limits_{i=1}^{N} B_\epsilon(x_i)$ (*). Denn ansonsten würde es für jedes $k \in \mathbb{N}$ eine Punkt x_k geben mit $x_k \notin B_\epsilon(x_i)$ für alle $i < k$, und die so zu beliebigem x_1 rekursiv definierte Folge x_k hätte wegen $d(x_k, x_i) \geq \epsilon$ keine Cauchy-Folge als Teilfolge, also auch keine konvergente Folge, Widerspruch.

Mittels der soeben bewiesenen Eigenschaft (*) wollen wir nun die Annahme zum Widerspruch führen, es gäbe eine offene Überdeckung U_i von X, die keine endliche Teilüberdeckung von X enthält. Betrachten wir zunächst (*) bei $\epsilon := 1$, dann kann mindestens eine der Kugeln $B_1(x_i)$ nicht durch endlich viele U_i überdeckt werden, d.h. es gibt einen Punkt x_0, für den $B_1(x_0)$ nicht durch endlich viele U_i überdeckt werden kann. Nun halbieren wir ϵ und wenden wieder (*) an. Dann kann der Schnitt von $B_1(x_0)$ mit einer der so gewonnenen Kugeln vom Radius $1/2$ wiederum nicht durch endlich viele U_i überdeckt werden. Deren Mittelpunkt bezeichnen wir mit x_1. Dies führen wir nun fort und gewinnen dadurch eine Folge x_k, für die $B_1(x_0) \cap \cdots \cap B_{2^{-k}}(x_k)$ nicht durch endlich viele U_i überdeckt werden kann. Die Folge x_k ist eine Cauchy-Folge, denn nach Konstruktion gilt $d(x_k, x_{k+1}) \leq 2^{-k} + 2^{-(k+1)} = 3 \cdot 2^{-(k+1)}$. Andererseits enthält x_k nach Voraussetzung eine konvergente Teilfolge, und daher muß schon die gesamte Folge x_k konvergieren. Sei x ihr Grenzwert und U_{i^*} eine offene Menge der Überdeckung, die x enthält. Dann gibt es aufgrund der Offenheit von U_{i^*} einen Radius $r > 0$ mit $B_r(x) \subset U_{i^*}$. Ist k nun so groß, dass $d(x_k, x) < r/2$ und $2^{-k} < r/2$ gilt, dann folgt

$$B_1(x_0) \cap \cdots \cap B_{2^{-k}}(x_k) \subset B_{2^{-k}}(x_k) \subset B_r(x) \subset U_{i^*} \quad ,$$

also kann man $B_1(x_0) \cap \cdots \cap B_{2^{-k}}(x_k)$ doch durch endlich viele der U_i überdecken, nämlich durch die eine Menge U_{i^*}, Widerspruch. □

In allgemeinen metrischen Räumen X ist zwar jede kompakte Teilmenge K auch abgeschlossen und beschränkt, denn

- wäre K nicht abgeschlossen, so gäbe es eine in X konvergente Folge $x_k \in K$, deren

Grenzwert nicht in K liegt. Aufgrund der Kompaktheit müßte x_k aber eine in K konvergente Teilfolge besitzen, Widerspruch.

- wäre K nicht beschränkt und $x \in K$, so gäbe es zu jedem $k \in \mathbb{N}$ einen Punkt $x_k \in K$ mit $d(x_k, x) \geq k$. Die Folge x_k kann aber im Widerspruch zur Kompaktheit keine konvergente Teilfolge besitzen, da $d(x_k, x_l) \geq |d(x_k, x) - d(x_l, x)|$ bei festem l für $k \to \infty$ beliebig groß wird.

Aber nicht jede abgeschlossene und beschränkte Teilmenge von X ist umgekehrt auch kompakt. Beispielsweise ist in $\mathbb{R}$ mit der zur Euklidischen Metrik äquivalenten Metrik $\tilde{d}(x, y) := \frac{\|x-y\|_2}{1+\|x-y\|_2}$ die Menge $\mathbb{N}$ abgeschlossen und beschränkt, aber nicht kompakt. In normierten endlich-dimensionalen Vektorräumen gilt jedoch auch die Umkehrung.

Satz 1.58 *Eine Teilmenge $K \subset X$ eines endlich-dimensionalen normierten Vektorraums X ist genau dann kompakt, wenn sie abgeschlossen und beschränkt ist.*

Beweis: Zunächst kann man nach Satz 1.19 und seinem Korollar ohne Einschränkung $X = \mathbb{R}^n$ annehmen.

Es ist nur noch zu zeigen, für jede Folge $x_k \in K$ in einer abgeschlossenen und beschränkten Teilmenge $K \subset \mathbb{R}^n$ eine Teilfolge existiert, die einen Grenzwert in K besitzt. Nun ist eine solche Folge x_k aber auch beschränkt, und daher hat sie nach Satz 1.17 eine im $\mathbb{R}^n$ konvergente Teilfolge. Der Grenzwert dieser Teilfolge liegt aber aufgrund der Abgeschlossenheit (siehe Definition 1.26) schon in K, was zu zeigen war. $\square$

Insbesondere sind alle mehrdimensionalen abgeschlossenen und beschränkten Intervalle $\prod_{i=1}^{n} [a_i, b_i] \subset \mathbb{R}^n$ $(-\infty < a_i < b_i < \infty)$ und die bzgl. einer Norm $\|\cdot\|$ abgeschlossenen Kugeln $\overline{B_r(x)} := \{y \in \mathbb{R}^n \mid \|x - y\| \leq r\}$ kompakt.

Stetige Abbildungen auf kompakten Räumen Wie in der Analysis I für stetige Funktionen auf Intervallen $[a, b]$ kann man auch für stetige Abbildungen auf kompakten metrischen Räumen viele nützliche Aussagen beweisen.

Satz 1.59 *Das Bild $f(X)$ eines kompakten metrischen Raumes X unter einer stetigen Abbildung $f : X \to Y$ ist eine kompakte Teilmenge des metrischen Raumes Y.*

Beweis: Sei V_i eine offene Überdeckung von $f(X)$, dann überdecken die offenen Mengen $U_i := f^{-1}(V_i)$ ganz X, und somit überdecken aufgrund der Kompaktheit von X schon endlich viele $U_{i_1}, \ldots, U_{i_n}$ ganz X. Insbesondere überdecken also $V_{i_1}, \ldots, V_{i_n}$ ganz $f(X)$. $\square$

Das folgende Korollar verallgemeinert den Satz vom Maximum und Minimum aus der Analysis I und garantiert die Exstenz von Maximal- bzw. Minimalstellen einer stetigen Funktion auf einem Kompaktum.

Korollar 1.60 *Jede stetige Funktion $f : X \to \mathbb{R}$ auf einem kompakten metrischen Raum X nimmt Maximum und Minimum an.*

Beweis: $f(X) \subset \mathbb{R}$ ist beschränkt, also hat $f(X)$ ein endliches Supremum M und Infimum m. Außerdem ist $f(X)$ abgeschlossen, also sind M und m Elemente von $f(X)$ und somit existieren Maximum und Minimum. □

Als Anwendung wollen wir beweisen, dass der Abstand

$$\operatorname{dist}(K, A) := \inf_{x \in K,\, y \in A} d(x, y)$$

einer kompakten Teilmenge K und einer abgeschlossenen Teilmenge A eines metrischen Raumes mit leerem Schnitt $K \cap A = \emptyset$ positiv ist. Tatsächlich, die Funktion $x \mapsto d(x, A) := \inf\limits_{y \in A} d(x, y)$ ist stetig, denn es gilt $d(x, A) \leq d(x, y) + d(y, A)$ nach der Dreiecksungleichung und aufgrund der Symmetrie in x, y auch

$$|d(x, A) - d(y, A)| \leq d(x, y) \quad ,$$

also hat die Funktion $d(\cdot, A)$ die Lipschitz-Konstante Eins. Also nimmt $d(\cdot, A)$ auf dem Kompaktum K sein Minimum an, d.h. es gibt ein $x^* \in K$ mit $d(x^*, A) = \operatorname{dist}(K, A)$. Da x^* nicht in A liegt und A abgeschlossen ist, gibt es eine Kugel $B_r(x^*)$ $(r > 0)$ um x^*, die A nicht schneidet. Somit ist $\operatorname{dist}(K, A) = d(x^*, A) > 0$ positiv.

Auch zum folgenden Satz ist das eindimensionale Analogon schon bekannt.

Satz 1.61 *Jede stetige Abbildung $f : X \to Y$ eines kompakten metrischen Raumes X in einen metrischen Raum Y ist gleichmäßig stetig.*

Beweis: Sei $\epsilon > 0$. Da f stetig ist, gibt es zu jedem Punkt $a \in X$ eine Kugel $B_{\delta(a)}(a)$ mit $d(f(x), f(a)) \leq \epsilon/2$ für alle $x \in B_{\delta(a)}(a)$. Nun überdecken aufgrund der Kompaktheit von X schon endlich viele der Kugeln $B_{\delta(a)/2}(a)$ ganz X. Sind ihre Radien $\delta(a_1)/2, \ldots, \delta(a_n)/2$, dann erfüllt das von a unabhängige $\delta := \frac{1}{2} \min(\delta(a_1), \ldots, \delta(a_n))$ die bei gleichmäßiger Stetigkeit verlangte Bedingung.

Tatsächlich, sind $a, x \in X$ zwei Punkte mit $d(x, a) \leq \delta$, dann gibt es eine Kugel $B_{\delta(a_i)/2}(a_i)$, in der a liegt, dann liegt wegen $d(x, a) \leq \delta \leq \delta(a_i)/2$ aber x schon in $B_{\delta(a_i)}(a_i)$, und daher gilt

$$d(f(x), f(a)) \leq d(f(x), f(a_i)) + d(f(a_i), f(a)) \leq \epsilon/2 + \epsilon/2 = \epsilon \quad .$$

□

Im Gegensatz zum allgemeinen Fall ist jede stetige Bijektion auf einem Kompaktum automatisch ein Homöomorphismus. Um dies zu zeigen, benötigen wir das folgende, auch schon für sich allein interessante Lemma.

Lemma 1.62 *Jede abgeschlossene Teilmenge eines kompakten metrischen Raumes ist kompakt.*

Beweis: Ist $A \subset X$ abgeschlossene Teilmenge des Kompaktums X und U_i eine offene Überdeckung von A, dann ist U_i zusammen mit $X \setminus A$ eine offene Überdeckung von X. Also gibt es endlich viele $U_{i_1}, \ldots, U_{i_n}$, die zusammen mit $X \setminus A$ ganz X überdecken, und somit überdecken die endlich vielen offenen Mengen $U_{i_1}, \ldots, U_{i_n}$ ganz A. □

Satz 1.63 *Ist eine stetige Abbildung $f : X \to Y$ von einem kompakten metrischen Raum X in einen metrischen Raum Y bijektiv und stetig, dann ist ihre Umkehrabbildung f^{-1} automatisch stetig.*

Beweis: Da die Urbilder von f^{-1} gerade die Bilder von f sind, müssen wir nur zeigen, dass das Bild einer offenen Menge unter f offen ist, oder äquivalenterweise, dass das Bild einer abgeschlossenen Menge $A \subset X$ unter f abgeschlossen ist. Nun ist A nach dem vorigen Lemma kompakt, also auch $f(A)$ als Bild eines Kompaktums unter einer stetigen Abbildung kompakt, und somit ist insbesondere $f(A)$ in Y abgeschlossen, was zu zeigen war. □

Diesen Satz kann man beispielsweise auf raumfüllende Kurven anwenden. Eine stetige Abbildung $c : [0,1] \to [0,1]^2$ heißt parametrisierte Kurve in $[0,1]^2$, und ist c surjektiv, so wird die Kurve raumfüllend genannt. Tatsächlich gibt es solche merkwürdigen parametrisierten Kurven, wie ein Beispiel von Peano zeigt. In Satz 1.74 werden wir aber beweisen, dass $\mathbb{R}$ und $\mathbb{R}^n$ für $n > 1$ nicht homöomorph sind, und ebenso sind auch $[0,1]$ und $[0,1]^2$ nicht homöomorph. Also kann die Parametrisierung c einer raumfüllenden Kurve nach Satz 1.63 nicht injektiv sein, muss sich also irgendwo selbst schneiden.

Produkte mit kompaktem Faktor Abschließend sollen noch Produkte mit kompakten metrischen Räumen diskutiert werden. Das folgende etwas technische Tubenlemma ist dabei der Ausgangspunkt.

Lemma 1.64 *Ist X ein metrischer Raum, K ein kompakter metrischer Raum, $x^* \in X$ ein Punkt und $W \subset X \times K$ eine offene Menge mit $\{x^*\} \times K \subset W$, dann gibt es eine offene Umgebung $U \subset X$ von $x^* \in X$ mit $U \times K \subset W$.*

Beweis: Aufgrund der Offenheit von W gibt es zu jedem Punkt $(x^*, y) \in \{x^*\} \times K$ offene Umgebungen $U_y \subset X$ von x^*, $V_y \subset K$ von y mit $U_y \times V_y \subset W$. Da K kompakt ist, reichen endlich viele $V_{y_1}, \dots, V_{y_n}$ aus, um K zu überdecken. Der Durchschnitt $U := U_{y_1} \cap \cdots \cap U_{y_n}$ der zugehörigen offenen Umgebungen von x^* ist dann aber wieder eine offene Umgebung von x^* und erfüllt außerdem $U \times K \subset W$. □

Anschaulich besagt das Tubenlemma, dass in jeder Umgebung der Faser $\{x^*\} \times K$ eine Tube $U \times K$ liegt.

Korollar 1.65 *Das Produkt $K \times K'$ kompakter metrischer Räume K, K' ist kompakt.*

Beweis: Sei W_i eine offene Überdeckung von $K \times K'$. Jede Faser $\{x\} \times K'$ wird aufgrund der Kompaktheit von K' bereits von endlich vielen $W_{i_1,x}, \dots, W_{i_{n(x)},x}$ überdeckt. Nach dem Tubenlemma überdecken diese endlich vielen offenen Mengen aber bereits eine ganze Tube $U_x \times K'$, wobei U_x eine offene Umgebung von x ist. Von den so gewonnenen Umgebungen U_x überdecken ihrerseits aber wieder endlich viele den kompakten Raum K. Somit reichen endlich viele W_i aus, um $K \times K'$ zu überdecken. □

Dieses Korollar erfährt im Satz von Tychonoff übrigens noch eine weitgehende Verallgemeinerung (siehe z.B. [Jänich2]).

Eine weitere praktische Anwendung des Tubenlemmas ist der folgende Nachweis der Stetigkeit parameterabhängiger Integrale.

Satz 1.66 *Sei X ein metrischer Raum und $f : X \times [a,b] \to \mathbb{R}$ eine stetige Funktion. Dann ist die Funktion $F : X \to \mathbb{R}$, $F(x) := \int\limits_a^b f(x,t)\,dt$, wohldefiniert und stetig.*

Beweis: Für jedes feste $x \in X$ ist die Funktion $t \mapsto f(x,t)$ als stetige Funktion über das kompakte Intervall $[a,b]$ Riemann-integrierbar, also ist F wohldefiniert. Um die Stetigkeit in $x^* \in X$ zu beweisen, sei $\epsilon > 0$ vorgegeben. Da die Funktion f stetig ist, ist die $\{x^*\} \times [a,b]$ enthaltende Menge $W := \{(x,t) \in X \times [a,b] \,|\, |f(x,t) - f(x^*,t)| < \epsilon\}$ offen. Nach dem Tubenlemma findet man eine offene Umgebung U von x^*, mit der $U \times [a,b] \subset W$ gilt. Für $x \in U$ erhält man

$$|F(x) - F(x^*)| \leq \int\limits_a^b |f(x,t) - f(x^*,t)|\,dt \leq |b-a|\epsilon \quad ,$$

also ist die Funktion F stetig in x^*. □

1.10 Zusammenhang

Ein weiterer wichtiger Satz der Analysis I über stetige Funktionen auf einem Intervall ist der Zwischenwertsatz. Dieser läßt sich verallgemeinern, indem man zusammenhängende Räume einführt. Man bemerke, dass die folgende Definition wiederum nur von der Topologie und nicht von der Wahl der Metrik abhängt.

Definition 1.67 *Ein metrischer Raum X heißt zusammenhängend, falls es keine Zerlegung $X = U \cup V$ von X in disjunkte, offene und nichtleere Teilmengen $U, V \subset X$ gibt, oder äquivalenterweise $\emptyset$ und X die einzigen Teilmengen von X sind, die sowohl offen als auch abgeschlossen sind.*

Insbesondere nennt man eine Teilmenge $M \subset X$ eines metrischen Raumes zusammenhängend, wenn M bzgl. der induzierten Metrik und somit bzgl. der Relativtopologie zusammenhängend ist.

Um die Äquivalenz der beiden in der Definition angegebenen Bedingungen einzusehen, bemerke man dass bei einer Zerlegung $X = U \cup V$ von X in disjunkte, offene und nichtleere Teilmengen $U, V \subset X$ die Mengen U, V wegen $U = X \setminus V$ und $V = X \setminus U$ auch abgeschlossen wären. Umgekehrt ist das Komplement $X \setminus U$ einer offenen und abgeschlossenen Teilmenge $U \subset X$ wieder offen und abgeschlossen, und es gilt dann $X = U \cup (X \setminus U)$.

Analog zu Satz 1.59 gilt:

Satz 1.68 *Das Bild $f(X)$ eines zusammenhängenden metrischen Raumes X unter einer stetigen Abbildung $f : X \to Y$ ist eine zusammenhängende Teilmenge des metrischen Raumes Y.*

Beweis: Wäre $f(X)$ nicht zusammenhängend, dann gäbe es disjunkte, offene und nichtleere Teilmengen $U, V \subset f(X)$ mit $f(X) = U \cup V$. Dann aber wären auch $f^{-1}(U), f^{-1}(V)$ disjunkte, offene und nichtleere Teilmengen von X, mit denen die Zerlegung $X = f^{-1}(U) \cup f^{-1}(V)$ im Widerspruch zum Zusammenhang von X gilt. $\square$

Um aus diesem Satz eine direkte Verallgemeinerung des Zwischenwertsatz folgern zu können, müssen wir uns zunächst darüber klar werden, welche Teilmengen von $\mathbb{R}$ zusammenhängend sind.

Beispiel 1.69 *Eine nichtleere Teilmenge $M \subset \mathbb{R}$ ist genau dann zusammenhängend, wenn sie ein Intervall ist.*

Tatsächlich, sei M ein Intervall. Ohne Einschränkung habe M zwei oder mehr Punkte, andernfalls ist nichts zu zeigen. Angenommen, es gäbe eine Zerlegung $M = U \cup V$ von M in disjunkte, offene und nichtleere Teilmengen $U, V \subset M$. Seien $u \in U$, $v \in V$, und die Benennung sei so, dass $u < v$ gilt. Da M ein Intervall ist, gilt $[u, v] \subset M$. Da U in M sowohl offen als auch abgeschlossen ist, liegt das Supremum s der Teilmenge $[u, v] \cap U$ in U, und wegen der Disjunktheit von U und V gilt $s \neq v$, so dass insbesondere kein Punkt aus $(s, v]$ zu U gehört. Aber als offene Teilmenge enthält U eine ganze Kugel um s, d.h. es existiert ein $r > 0$ mit $(s - r, s + r) \subset U$, im Widerspruch dazu, dass keiner der Punkte aus $(s, v]$ in U liegt.

Ist andererseits $M \subset \mathbb{R}$ nichtleer und kein Intervall, dann gibt es Punkte $u, v \in M$ und einen Punkt $s \notin M$ zwischen diesen. Daher sind $U := M \cap (-\infty, s)$ und $V := M \cap (s, \infty)$ disjunkte, in M offene und nichtleere Teilmengen mit $M = U \cup V$, also ist M nicht zusammenhängend.

Sei nun $f : X \to \mathbb{R}$ eine stetige Funktion auf einem zusammenhängenden metrischen Raum X. Da das Bild $f(X) \subset \mathbb{R}$ nach Satz 1.68 zusammenhängend und daher nach dem vorigen Beispiel ein Intervall ist, ergibt sich sofort die folgende Verallgemeinerung des Zwischenwertsatzes.

Korollar 1.70 *Sei X ein zusammenhängender metrischer Raum, seien $a, b \in X$ Punkte und sei $f : X \to \mathbb{R}$ eine stetige Funktion. Dann nimmt f jeden Wert zwischen $f(a)$ und $f(b)$ an.*

Diesen Satz kann man auch benutzen, um nachzuweisen, dass ein metrischer Raum nicht zusammenhängend ist. Die Teilmenge $O(n) := \{A \in L(\mathbb{R}^n, \mathbb{R}^n) \mid A^T A = \mathrm{Id}\}$ der orthogonalen Matrizen ist beispielsweise nicht zusammenhängend, denn die Determinante $\det : L(\mathbb{R}^n, \mathbb{R}^n) \to \mathbb{R}$ ist als Polynom n-ten Grades in den Einträgen der Matrix natürlich stetig, nimmt aber auf $O(n)$ nur die beiden Werte -1 und 1 an (Drehungen haben Determinante 1, Spiegelungen Determinante -1).

Wegzusammenhang Einen alternativen Zusammenhangsbegriff liefert die folgende Definition.

Definition 1.71 *Einen metrischen Raum X nennt man wegzusammenhängend, falls es zu je zwei Punkten $x, x' \in X$ eine stetige Abbildung $c : [0, 1] \to X$ mit $c(0) = x$ und*

$c(1) = x'$ gibt.

Mit anderen Worten heißt ein Raum X also wegzusammenhängend, wenn je zwei Punkte durch eine Kurve verbunden werden können.

Satz 1.72 *Jeder wegzusammenhängende metrische Raum ist auch zusammenhängend.*

Beweis: Sei X wegzusammenhängend. Gäbe es disjunkte, offene und nichtleere Teilmengen $U, V \subset X$ mit $X = U \cup V$, dann könnte man zwei Punkte $u \in U$, $v \in V$ wählen und aufgrund des Wegzusammenhangs eine parametrisierte Kurve $c : [0,1] \to X$ mit $c(0) = u$, $c(1) = v$ finden. Mit dieser wäre aber $[0,1] = c^{-1}(U) \cup c^{-1}(V)$ eine Zerlegung von $[0,1]$ in disjunkte, offene und nichtleere Teilmengen, im Widerspruch dazu, dass $[0,1]$ zusammenhängend ist. □

Angemerkt sei, dass zwar eine offene Teilmenge $M \subset \mathbb{R}^n$ genau dann zusammenhängend ist, wenn sie wegzusammenhängend ist (siehe [Königsberger II]), dass aber bereits für beliebige Teilmengen von $\mathbb{R}^n$ Zusammenhang eine schwächere Eigenschaft als Wegzusammenhang ist.

Abschließend soll gezeigt werden, dass die eindimensionale Gerade $\mathbb{R}$ nicht homöomorph zu einem höherdimensionalen Raum $\mathbb{R}^n$, $n > 1$, ist. Tatsächlich ist dies keineswegs eine triviale Frage, denn zwischen $\mathbb{R}$ und $\mathbb{R}^n$ gibt es schließlich eine Bijektion, und das Beispiel einer raumfüllenden parametrisierten Kurve mag auch Verunsicherung aufkommen lassen. Mittels des Zusammenhangbegriffs findet man aber recht schnell einen Beweis.

Beispiel 1.73 *Für $n > 1$ ist $\mathbb{R}^n \setminus \{0\}$ wegzusammenhängend und insbesondere zusammenhängend.*

Denn sind $x, z \in \mathbb{R}^n \setminus \{0\}$ zwei Punkte, dann ist mit einem dritten Punkt $y \in \mathbb{R}^n \setminus \{0\}$, für den der Nullpunkt weder auf der Strecke von x nach y noch auf der Strecke von y nach z liegt, die durch $c(t) := x + 2t(y - x)$ für $t \in [0, 1/2]$ und $c(t) := y + (2t - 1)(z - y)$ für $t \in [1/2, 1]$ definierte parametrisierte Kurve $c : [0,1] \to \mathbb{R}^n \setminus \{0\}$ ein Polygonzug, der x und z verbindet.

Satz 1.74 *Für $n > 1$ ist $\mathbb{R}$ nicht zu $\mathbb{R}^n$ homöomorph.*

Beweis: Wäre $f : \mathbb{R}^n \to \mathbb{R}$ ein Homöomorphismus, dann wäre auch die Einschränkung $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R} \setminus \{f(0)\}$ ein Homöomorphismus. Aber $\mathbb{R}^n \setminus \{0\}$ ist zusammenhängend, während $\mathbb{R} \setminus \{f(0)\}$ nicht zusammenhängend ist, so dass f wegen Satz 1.68 nicht sowohl stetig als auch surjektiv sein kann, Widerspruch. □

Für beliebige $m \neq n$ sind auch $\mathbb{R}^n$ und $\mathbb{R}^m$ nicht homöomorph, doch um dies zu zeigen, muss man stärkere Hilfsmittel aus der algebraischen Topologie einsetzen (siehe beispielsweise [Hatcher, Theorem 2.26]).

Kapitel 2

Differentialrechnung

In diesem Kapitel wird die Differentialrechnung auf dem $\mathbb{R}^n$ und allgemeiner auf Banach-Räumen formuliert. Die konsequente Nutzung von Banach-Räumen an den Stellen, wo es keinerlei zusätzlichen Aufwand bedeutet, ermöglicht insbesondere die direkte Anwendung auf unendlichdimensionale Funktionenräume, wie sie beispielsweise in der Physik als Räume von Feldern auftreten. Auch wenn uns hier in erster Linie der endlichdimensionale Fall interessiert, ist es doch recht hilfreich zu wissen, dass sich ein Großteil der Resultate der Differentialrechnung direkt auf solche unendlichdimensionalen Situationen überträgt.

Beginnen werden wir mit der Differentiation parametrisierter Kurven, die noch weitgehend parallel zum eindimensionalen Fall verläuft. Dann wird ausführlich auf die Differentiation von Abbildungen eingegangen, die auf mehrdimensionalen Räumen definiert sind, und insbesondere der Unterschied zwischen partieller Differenzierbarkeit und (totaler) Differenzierbarkeit geklärt. Für (total) differenzierbare Abbildungen kann man die Stetigkeit und die Gültigkeit der Kettenregel zeigen, und dies erlaubt eine weitreichende Analysis. Insbesondere kann man stetig differenzierbare Abbildungen durch Taylorpolynome approximieren und dadurch für zweimal stetig differenzierbare Funktionen hinreichende Kriterien für das Vorliegen eines Extremums gewinnen.

Abschließend werden mit dem Satz über lokale Umkehrbarkeit und dem Satz über implizite Funktionen sehr mächtige Werkzeuge der nichtlinearen Analysis bereitgestellt, mit deren Hilfe man beispielsweise Untermannigfaltigkeiten und Extrema unter Nebenbedingungen diskutieren kann.

2.1 Differenzierbare Kurven

Wie zuvor schon angedeutet nennt man eine stetige Abbildung $c : I \to X$ von einem Intervall $I \subset \mathbb{R}$ in einen metrischen Raum (X, d) eine parametrisierte Kurve in X. Ihr Bild $C := c(I) \subset X$ ist die eigentliche Kurve und wird häufig die Spur von c genannt.

Hier wollen wir uns speziell mit parametrisierten Kurven in Banach-Räumen $(X, \|\cdot\|)$ beschäftigen, die differenzierbar sind. Da der Definitionsbereich einer parametrisierten

Kurve ein Intervall ist, entspricht die Theorie der Differentiation von parametrisierten Kurven weitgehend dem aus der Analysis I bekannten eindimensionalen Fall, ganz im Gegensatz zu der in den folgenden Abschnitten diskutierten Theorie der Differentiation von Abbildungen, die auf einem mehrdimensionalen Raum definiert sind.

Definition 2.1 *Eine parametrisierte Kurve $c : I \to X$ in einem Banach-Raum X heißt differenzierbar in $t^* \in I$, wenn der Grenzwert $\dot{c}(t^*) := \lim\limits_{t\to t^*} \frac{c(t)-c(t^*)}{t-t^*}$ existiert, und $\dot{c}(t^*) \in X$ nennt man dann die Ableitung von c in t^*. Desweiteren heißt c differenzierbar, falls c in jedem Punkt $a \in I$ differenzierbar ist.*

Will man explizit die Variable t angeben, nach der man $c(t)$ ableitet, so schreibt man statt $\dot{c}$ auch $\frac{dc}{dt}$ für die Ableitung.

Im endlichdimensionalen Fall $X = \mathbb{R}^n$ ist $c = (c_1, \ldots, c_n)$ wegen Lemma 1.16 genau dann differenzierbar, wenn für jedes $i = 1, \ldots, n$ die Komponentenfunktionen $c_i : I \to \mathbb{R}$ differenzierbar sind.

Anschaulich macht man sich die Bedeutung der Ableitung am besten im Fall $X = \mathbb{R}^2$ oder $X = \mathbb{R}^3$ klar. Modelliert der Parameter $t \in I$ von c die Zeit und gibt $c(t) \in X$ die Position eines Punktteilchens zum Zeitpunkt t an, das die Kurve $C = c(I) \subset X$ durchläuft, dann ist die Ableitung $\dot{c}(t^*)$ gerade der Geschwindigkeitsvektor des Punktteilchens zum Zeitpunkt t^*. Die Norm $\|\dot{c}(t^*)\|$ entspricht der Geschwindigkeit, mit der sich das Punktteilchen bewegt, und die normierte Richtung des Punktteilchens ist durch $\frac{\dot{c}(t^*)}{\|\dot{c}(t^*)\|}$ gegeben (falls $\|\dot{c}(t^*)\| \neq 0$ ist).

Wie in der Analysis I folgt aus Differenzierbarkeit schon die Stetigkeit.

Lemma 2.2 *Existiert für eine Abbildung $c : I \to X$ der Grenzwert $\lim\limits_{t\to t^*} \frac{c(t)-c(t^*)}{t-t^*}$ in t^*, dann ist c stetig in t^*.*

Beweis: Da der Nenner von $\lim\limits_{t\to t^*} \frac{c(t)-c(t^*)}{t-t^*}$ gegen Null konvergiert, kann der Grenzwert nur dann existieren, wenn auch der Zähler gegen $0 \in X$ konvergiert, d.h. wenn $\lim\limits_{t\to t^*} c(t) = c(t^*)$ gilt und daher c stetig in t^* ist. □

Insbesondere sind differenzierbare parametrisierte Kurven automatisch stetig. Die Ableitung ist als Abbildung $\dot{c} : t \mapsto \dot{c}(t)$ aber nicht automatisch stetig. Ist sie es doch, so nennt man die Kurve c stetig differenzierbar oder eine C^1-Kurve. In diesem Fall kann man $\dot{c}$ wiederum als Kurve auffassen und bei Differenzierbarkeit von $\dot{c}$ von der zweiten Ableitung $\ddot{c}$ sprechen, sowie analog Ableitungen höherer Ordnung und C^k-Differenzierbarkeit definieren.

Die Ableitung von parametrisierten Kurven ist linear, d.h. $\frac{d}{dt}(\lambda c + \mu \tilde{c}) = \lambda \dot{c} + \mu \dot{\tilde{c}}$ gilt für Konstanten $\lambda, \mu \in \mathbb{R}$. Die Kettenregel nimmt die folgende Form an:

Lemma 2.3 *Ist $\phi : I \to \tilde{I}$ differenzierbar in t^* und $c : \tilde{I} \to X$ differenzierbar in $\phi(t^*)$, dann ist $c \circ \phi$ differenzierbar in t^* mit Ableitung $\frac{d}{dt}(c \circ \phi)(t^*) = \dot{c}(\phi(t^*)) \cdot \phi'(t^*)$.*

Zumindest im endlichdimensionalen Fall $X = \mathbb{R}^n$ folgen all diese Regeln aus denen der Analysis I, da Differentiation nichts anderes als komponentenweise Differentiation ist und für die Komponentenfunktionen die Ableitungsregeln schon bekannt sind. Für unendlichdimensionale Banach-Räume X vertrösten wir darauf, dass diese Regeln nur Spezialfälle von später allgemein bewiesenen Sätzen der Differentiationstheorie sind.

Während die Ableitungsregeln also noch wie im Eindimensionalen gelten, ist dies für den Mittelwertsatz bereits nicht mehr der Fall.

Beispiel 2.4 *Für die parametrisierte Kurve $c : [0, 2\pi] \to \mathbb{R}^2$, $c(t) := (\cos(t), \sin(t))$, in der Euklidischen Ebene $\mathbb{R}^2$, deren Bild der Kreis $S^1 := \{(x, y) \in \mathbb{R}^2 \,|\, x^2+y^2 = 1\}$ ist, gibt es kein $\xi \in (0, 2\pi)$ mit $\dot{c}(\xi) = \frac{c(2\pi)-c(0)}{2\pi-0} = 0$. Denn der Vektor $\dot{c}(\xi) = (-\sin(\xi), \cos(\xi))$ hat für jedes $\xi \in (0, 2\pi)$ die Länge $\|\dot{c}(\xi)\|_2 = 1$ und ist insbesondere nie der Nullvektor.*

Im folgenden wollen wir einige Themen vorstellen, bei denen die Differentiation von parametrisierten Kurven eine Rolle spielt.

Lineare Differentialgleichungen und die Exponentialabbildung Viele naturwissenschaftliche Probleme lassen sich auf das mathematische Problem zurückführen, dass man eine parametrisierte Kurve $c : I \to X$ in einen Banach-Raum X sucht, deren Ableitung $\dot{c}(t)$ für jedes $t \in I$ mit einem vorgegebenen Vektor $f(c(t))$ am gerade durchlaufenen Punkt $c(t)$ identisch ist.

Genauer sucht man zu einer vorgegebenen Abbildung $f : X \to X$, die Vektorfeld genannt wird, und einem vorgebenen Anfangswert $x \in X$ eine differenzierbare parametrisierte Kurve $c : I \to X$ auf einem Intervall I mit $0 \in I$, die für jedes $t \in I$ die Gleichung $\dot{c}(t) = f(c(t))$ erfüllt und in Null den Wert $c(0) = x$ hat. Diese Aufgabe nennt man auch ein Anfangswertproblem für ein System von autonomen gewöhnlichen Differentialgleichungen.

Der einfachste Fall ist der eines linearen Vektorfeldes $f(x) := Ax$, wobei $A : X \to X$ eine stetige lineare Abbildung ist und daher das Differentialgleichungssystem $\dot{c}(t) = Ac(t)$ linear (und homogen) genannt wird.

Beispiel 2.5 *Betrachtet man die Position $x(t) \in \mathbb{R}^n$ eines Punktteilchen ($n = 1, 2, 3$), so ändert diese sich in Abhängigkeit von der Geschwindigkeit $v(t) \in \mathbb{R}^n$ nach dem Gesetz $\dot{x}(t) = v(t)$, und die Geschwindigkeit ändert sich ihrerseits wieder in Abhängigkeit von der Kraft F pro Masse, die auf das Punktteilchen einwirkt. Hängt die Kraft $F = \alpha \cdot x(t)$ linear von der Position $x(t)$ über einen Faktor $\alpha \in \mathbb{R}$ ab, so erhält man insgesamt das Differentialgleichungssystem*

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix}(t) = \begin{pmatrix} 0_n & E_n \\ \alpha E_n & 0_n \end{pmatrix} \begin{pmatrix} x \\ v \end{pmatrix}(t) \quad ,$$

und gesucht ist eine Kurve $c(t) = (x(t), v(t)) \in \mathbb{R}^{2n}$, die dieses System zu vorgegebener Anfangsposition und -geschwindigkeit löst.

Beispiel 2.6 *Betrachtet man eine Population von Raubtieren und Beutetieren, bei der $x(t) \in \mathbb{R}$ die (als kontinuierlich angenommene) Anzahl von Raubtieren und $y(t) \in \mathbb{R}$ die (als kontinuierlich angenommene) Anzahl von Beutetieren zum Zeitpunkt t angibt,* *so macht es Sinn, die zeitliche Änderung des Tierbestandes durch das Differentialgleichungssystem*

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix}(t) = \begin{pmatrix} -\alpha & \beta \\ -\gamma & \delta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}(t)$$

auf dem $\mathbb{R}^2$ *zu modellieren, wobei* $\alpha > 0$ *die Sterberate der Raubtiere angibt (Raubtiere ohne Essen töten sich gegenseitig),* $\beta > \alpha$ *die Rate angibt, mit der sich die Raubtiere bei genügend Nahrung vermehren,* $\gamma > 0$ *die Rate angibt, mit der die Beutetiere von den Raubtieren getötet werden, und* $\delta > 0$ *die Vermehrungsrate der Beutetiere angibt.*

Das Anfangswertproblem für ein lineares Differentialgleichungssystem $\dot{c}(t) = Ac(t)$ bei $c(0) = x$ kann man für $A \in L(X,X)$ mit Hilfe der Exponentialabbildung $\exp : L(X,X) \to L(X,X)$ ganz einfach lösen, denn die Lösung ist $c(t) := \exp(tA)x$.

Um dies zu zeigen, müssen wir zunächst einmal die Exponentialabbildung exp auf der Banach-Algebra $L(X,X)$ der stetigen linearen Abbildungen definieren. Dies kann ganz analog zum Eindimensionalen durch die Potenzreihe $\exp(A) := \sum\limits_{k=0}^{\infty} \frac{A^k}{k!}$ geschehen, die einen unendlichen Konvergenzradius besitzt und deshalb nach Abschnitt 1.8 auch für jedes $A \in L(X,X)$ konvergiert. Wie in der Analysis I kann man desweiteren für $A, B \in L(X,X)$ mit $AB = BA$ die Funktionalgleichung $\exp(A+B) = \exp(A)\exp(B)$ mittels des Cauchy-Produktes von absolut konvergenten Reihen unter Ausnutzung der Kommutativität $AB = BA$ beweisen.

Beispiel 2.7

- *Für eine Diagonalmatrix* $A := \mathrm{diag}(\alpha_1, \ldots, \alpha_n) \in L(\mathbb{R}^n, \mathbb{R}^n)$ *gilt die Formel* $\exp(A) = \mathrm{diag}(e^{\alpha_1}, \ldots, e^{\alpha_n})$.
- *Für den Jordanblock* $A := \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$ *gilt wegen* $A^k = \begin{pmatrix} \lambda^k & k\lambda^{k-1} \\ 0 & \lambda^k \end{pmatrix}$ *die Gleichung* $\exp(A) = e^{\lambda} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.
- *Es gilt* $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ *und auch wirklich*

  $$\exp\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} e & e-1 \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} e & e \\ 0 & 1 \end{pmatrix} = \exp\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \exp\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

  wegen $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}^k = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$.

Insbesondere erfüllt die Kurve $t \mapsto \exp(tA)$ in $L(X,X)$ die Gleichung $\exp((s+t)A) = \exp(sA)\exp(tA)$ sowie $\exp(0A) = \mathrm{Id}$. Da für $|t| \leq 1$ die Abschätzung

$$\|\exp(tA) - \mathrm{Id} - tA\| \leq |t|^2 \sum_{k=2}^{\infty} \frac{\|A\|^k}{k!}$$

gilt, folgt desweiteren für die Kurve $c(t) := \exp(tA)x$ nicht nur $c(0) = x$, sondern auch

$$\dot{c}(t) = \lim_{h\to 0} \frac{c(t+h) - c(t)}{h} = \lim_{h\to 0} \frac{\exp((t+h)A)x - \exp(tA)x}{h} =$$
$$\exp(tA)\left(\lim_{h\to 0} \frac{\exp(hA) - \mathrm{Id}}{h} x\right) = \exp(tA)Ax = A\exp(tA)x = Ac(t) \quad .$$

Dies zeigt, dass $c(t) = \exp(tA)x$ wirklich die Differentialgleichung $\dot{c}(t) = Ac(t)$ zum Anfangswert $c(0) = x$ löst.

Darüberhinaus ist die Lösung des Anfangswertproblems auch eindeutig, denn löst c die Differentialgleichung $\dot{c}(t) = Ac(t)$, dann gilt analog zur obigen Rechnung

$$\frac{d}{dt}(\exp(-tA)c(t)) = \exp(-tA)(-A)c(t) + \exp(-tA)\dot{c}(t) = \exp(-tA)(-Ac(t) + Ac(t)) = 0 \quad ,$$

also ist $\exp(-tA)c(t)$ konstant. Wegen $\exp(0) = \mathrm{Id}$ und des Anfangswertes $c(0) = x$ gilt somit $\exp(-tA)c(t) = x$ und also $c(t) = \exp(tA)x$.

Injektiv immersierte Kurven In diesem Abschnitt wollen wir uns mit Kurven $C \subset X$, die stetig differenzierbar parametrisiert werden können, als Teilmengen des Banach-Raumes X beschäftigen und keinen Wert mehr darauf legen, wie die Kurve parametrisiert wird.

Die Ableitung $\dot{c}(t^*)$ einer Parametrisierung c von C nennt man in diesem Zusammenhang auch einen Tangentialvektor im Punkt $c(t^*) \in C$ der Kurve.

Allerdings kann man aus einer Parametrisierung $c : I \to X$ von C durch eine Parametertransformation $\phi : I \to I$ auch eine neue Parametrisierung $c \circ \phi$ derselben Kurve gewinnen. Nach Lemma 2.3 hat diese die Ableitung $\frac{d}{dt}(c \circ \phi)(t^*) = \dot{c}(\phi(t^*)) \cdot \phi'(t^*)$, und da $\phi'(t^*)$ alle möglichen Werte aus $\mathbb{R}$ annehmen kann[1], ist nur die Menge T_xC aller Tangentialvektoren an einem Punkt $x \in C$ eindeutig durch die Kurve C bestimmt.

Die Menge T_xC aller Tangentialvektoren in einem Punkt $x \in C$ kann in dieser allgemeinen Situation noch recht kompliziert sein. Beispielsweise kann es Punkte von C geben, an denen die Ableitung **jeder** stetig differenzierbaren Parametrisierung von C verschwindet und somit $T_xC = \{0\}$ gilt. Solche Punkte nennt man singuläre Punkte von C.

Beispiel 2.8 *Die durch $t \mapsto (t^2, t^3)$ parametrisierte Neilsche Parabel C hat $(0,0)$ als singulären Punkt. Denn ist $c(t) = (x(t), y(t))$ eine beliebige Parametrisierung mit $c(0) = (0,0)$, dann muß $x'(0) = 0$ gelten, da $x(t)$ bei $t = 0$ ein globales Minimum hat, und wegen $y(t)^2 = x(t)^3$ (ansonsten würde c gar nicht die Kurve C parametrisieren) gilt dann auch $y'(0) = \pm\frac{3}{2}\sqrt{|x(0)|}x'(0) = 0$.*

Selbst wenn es keine singulären Punkte gibt (in diesem Fall nennt man C eine immersierte Kurve), könnte sich die Kurve noch selbst schneiden, d.h. c könnte nicht injektiv

[1] Hier müßte man eigentlich noch beweisen, dass es zu jedem $r \in \mathbb{R}$ und $t^* \in I$ wirklich eine (beliebig oft) differenzierbare Funktion $\phi : I \to I$ mit $\phi(t^*) = t^*$ und $\phi'(t^*) = r$ gibt.

sein und es könnte somit $t_1 \neq t_2$ mit $c(t_1) = c(t_2)$ geben, bei denen $\dot{c}(t_1)$ nicht ein Vielfaches von $\dot{c}(t_2)$ ist.

Beispiel 2.9 *Die beliebig oft differenzierbare Kurve* $c(t) := (t^2 - 1, t^3 - t)$ *hat den Doppelpunkt* $c(1) = (0,0) = c(-1)$*, bei dem* $\dot{c}(1) = (2,2)$ *kein Vielfaches von* $\dot{c}(-1) = (-2,2)$ *ist.*

Gibt es aber eine stetig differenzierbare Parametrisierung $c : I \to X$ von C, die injektiv ist und $\dot{c}(t) \neq 0$ für alle $t \in I$ erfüllt, so heißt C eine injektiv immersierte Kurve. In diesem Fall ist die Menge T_xC aller möglichen Tangentialvektoren am Punkt $x \in C$ ein eindimensionaler Unterraum von X. Aber selbst injektiv immersierte Kurven können noch Punkte besitzen, in deren Nähe die Kurve nicht homöomorph zu einem Intervall ist.

Beispiel 2.10 *Die durch* $c(t) := (\sin(2\arctan(t)), 2\sin(4\arctan(t)))$ *parametrisierte Kurve* C *sieht aus wie eine liegende Acht, obwohl* c *injektiv ist und* $\dot{c}(t) \neq 0$ *für alle* $t \in \mathbb{R}$ *erfüllt.*

Insbesondere gibt es keinen Homöomorphismus von $C \cap U$*,* U *eine offene Umgebung von* $(0,0)$*, auf ein offenes Intervall* I*. Denn* $(C \cap U) \setminus \{(0,0)\}$ *hat vier Zusammenhangskomponenten, während* $I \setminus \{r\}$ *(*$r \in I$ *beliebig) zwei Zusammenhangskomponenten hat, was der Eigenschaft widerspricht, dass ein Homöomorphismus zusammenhängende Mengen auf zusammenhängende Mengen abbildet.*

Ist die Parametrisierung $c : I \to C$ jedoch nicht nur eine injektive Immersion, sondern sogar ein Homöomorphismus auf ihr mit der Relativtopologie versehenes Bild $C \subset X$, dann kann solch ein Fall nicht auftreten und man nennt C eine eindimensionale Untermannigfaltigkeit von X. Allgemeine mehrdimensionale Untermannigfaltigkeiten werden wir später genauer besprechen.

Bogenlänge Obwohl die Diskussion der Bogenlänge einer Kurve eher zur Integrationstheorie gehört, wollen wir sie hier schon ansprechen.

Eine parametrisierte Kurve $c : I \to X$ in einem Banach-Raum $(X, \|\cdot\|)$ heißt rektifizierbar, wenn die Längen aller Sehnenpolygone von c beschränkt sind. Dabei ist das Sehnenpolygon zu einer Zerlegung $Z = \{t_0 < t_1 < \cdots < t_k\}$ von I $(t_i \in I)$ das durch die Strecken von $c(t_{i-1})$ nach $c(t_i)$ gegebene Polygon, und dieses hat die Länge $\sum_{i=1}^{k} \|c(t_i) - c(t_{i-1})\|$.

Definition 2.11 *Die Bogenlänge* $L(c)$ *einer rektifizierbaren parametrisierten Kurve* c *ist das Supremum* $L(c) := \sup_Z \sum_{i=1}^{k} \|c(t_i) - c(t_{i-1})\|$ *aller Längen der Sehnenpolygone von* c*, wobei das Supremum über alle Zerlegungen* $Z = \{t_0 < t_1 < \cdots < t_k\}$ *von* I *(*$t_i \in I$*) gebildet wird.*

Beispiel 2.12 *Jede Lipschitz-stetige Parametrisierung $c : [a,b] \to X$ auf einem kompakten Intervall $[a,b]$ ist rektifizierbar, denn mit einer Lipschitz-Konstanten L von c gilt*

$$\sum_{i=1}^{k} \|c(t_i) - c(t_{i-1})\| \leq \sum_{i=1}^{k} L|t_i - t_{i-1}| = L(b-a) \quad ,$$

und insbesondere ist $L(c) \leq L \cdot (b-a)$.

Die Bogenlänge einer stetig differenzierbaren Parametrisierung $c : [a,b] \to X$ kann man ganz einfach ausrechnen.

Satz 2.13 *Jede stetig differenzierbare Parametrisierung $c : [a,b] \to X$ einer Kurve in einem Banach-Raum X ist rektifizierbar und hat die Bogenlänge $L(c) := \int_a^b \|\dot{c}(t)\|\,dt$.*

Beweis: Für den Beweis dieses Satzes definieren wir zunächst einmal das Riemann-Integral einer Kurve mit Werten in einem Banach-Raum, welches auch für sich genommen schon nützlich ist:

Eine Abbildung $c : [a,b] \to X$ heißt Riemann-integrierbar, wenn es einen Vektor $x \in X$ gibt, für den zu jedem $\epsilon > 0$ ein $\delta > 0$ existiert, so dass für jede Zerlegung $Z = \{a = t_0 < t_1 < \cdots < t_k = b\}$ von $[a,b]$ der Feinheit $\max\limits_{i=1,\ldots,k} |t_i - t_{i-1}| \leq \delta$ und jede Wahl von Stützstellen $\tau_i \in [t_{i-1}, t_i]$ die Ungleichung $\|x - \sum\limits_{i=1}^{k} c(\tau_i)(t_i - t_{i-1})\| \leq \epsilon$ gilt. Den Vektor x nennt man dann das Riemann-Integral von f und bezeichnet ihn mit $\int_a^b c(t)\,dt$. Man bemerke, dass diese Definition völlig analog zur Definition des Riemann-Integrals mittels Riemann-Summen im Eindimensionalen ist.

Nun ist jede stetige Abbildung $c : [a,b] \to X$ Riemann-integrierbar, denn als stetige Abbildung auf dem kompakten Intervall $[a,b]$ ist c gleichmäßig stetig, und daher gibt es zu jedem $\epsilon > 0$ ein $\delta > 0$, mit dem aus $|s-t| \leq \delta$ schon $\|c(s) - c(t)\| \leq \frac{\epsilon}{b-a}$ folgt. Somit bilden die Riemannschen Summen $\sum\limits_{i=1}^{k} c(\tau_i^{(n)})(t_i^{(n)} - t_{i-1}^{(n)})$ für jede Folge von Zerlegungen $Z^{(n)} = \{a = t_0^{(n)} < t_1^{(n)} < \cdots < t_k^{(n)} = b\}$ mit gegen Null konvergierender Feinheit und jede Wahl von Stützstellen $\tau_i^{(n)} \in [t_{i-1}^{(n)}, t_i^{(n)}]$ eine Cauchy-Folge, und aufgrund der Vollständigkeit von X gibt es einen Vektor $x \in X$, gegen den diese konvergiert, d.h. das Integral $\int_a^b c(t)\,dt$ existiert.

Da das Integral mittels Riemannscher Summen definiert ist und für diese Summen die Dreiecksungleichung anwendbar ist, gilt auch $\| \int_a^b c(t)\,dt\| \leq \int_a^b \|c(t)\|\,dt$.

Als letzten Schritt benötigen wird noch die Gleichung $c(t_i) - c(t_{i-1}) = \int_{t_{i-1}}^{t_i} \dot{c}(t)\,dt$ für stetig differenzierbare Kurven c. Ist $\phi : X \to \mathbb{R}$ ein stetiges lineares Funktional auf X, dann gilt aufgrund der Vertauschbarkeit von ϕ mit Riemannschen Summen auch $\phi\left(\int_{t_{i-1}}^{t_i} \dot{c}(t)\,dt\right) = \int_{t_{i-1}}^{t_i} \phi(\dot{c}(t))\,dt$, sowie aufgrund der Vertauschbarkeit von ϕ mit Differenzenquotienten auch $\frac{d}{dt}(\phi \circ c) = \phi \circ \dot{c}$. Also gilt nach dem Hauptsatz der Differential-und Integralrechnung für jedes stetige lineare Funktional ϕ auf X die Gleichung

$$\phi\left(\int_{t_{i-1}}^{t_i} \dot{c}(t)\,dt\right) = \int_{t_{i-1}}^{t_i} \phi(\dot{c}(t))\,dt = \int_{t_{i-1}}^{t_i} \frac{d}{dt}(\phi\circ c)(t)\,dt = \phi(c(t_i)) - \phi(c(t_{i-1})) = \phi(c(t_i) - c(t_{i-1}))$$

Im Endlich-dimensionalen folgt daraus trivialerweise die Gleichung $c(t_i) - c(t_{i-1}) = \int_{t_{i-1}}^{t_i} \dot{c}(t)\,dt$, denn ϕ könnte die Projektion eines Vektors auf eine seiner Komponenten sein. Für unendlich-dimensionale Banach-Räume folgt die Gleichung aus dem Satz von Hahn-Banach, aufgrund dessen aus $\phi(x) = \phi(\tilde{x})$ für alle stetigen linearen Funktionale ϕ auf X auch schon $x = \tilde{x}$ folgt. Dieser Satz ist ganz und gar nicht elementar, einen Beweis kann man beispielsweise in [Heuser III] finden.

Mit diesen Hilfsmitteln können wir nun die Formel $L(c) := \int_a^b \|\dot{c}(t)\|\,dt$ relativ leicht zeigen. Für die Länge eines Sehnenpolygons gilt einerseits die Ungleichung

$$\sum_{i=1}^{k} \|c(t_i) - c(t_{i-1})\| = \sum_{i=1}^{k} \left\| \int_{t_{i-1}}^{t_i} \dot{c}(t)\,dt \right\| \leq \sum_{i=1}^{k} \int_{t_{i-1}}^{t_i} \|\dot{c}(t)\|\,dt = \int_a^b \|\dot{c}(t)\|\,dt$$

und daher $L(c) \leq \int_a^b \|\dot{c}(t)\|\,dt$. Um auch die umgekehrte Ungleichung zu zeigen, beweisen wir, dass es zu jedem $\epsilon > 0$ ein Sehnenpolygon mit $\sum\limits_{i=1}^{k} \|c(t_i) - c(t_{i-1})\| \geq \int\limits_a^b \|\dot{c}(t)\|\,dt - \epsilon$ gibt:

Zum Integral $\int\limits_a^b \dot{c}(t)\,dt$ gibt es aufgrund der gleichmäßigen Stetigkeit von $\dot{c}$ auf $[a, b]$ eine Riemannsche Summe $\sum\limits_{i=1}^{k} \dot{c}(\tau_i)(t_i - t_{i-1})$, für die $\|\dot{c}(t) - \dot{c}(\tau_i)\| \leq \frac{\epsilon}{2(b-a)}$ bei $t \in [t_{i-1}, t_i]$ gilt. Daraus folgt

$$\sum_{i=1}^{k} \|c(t_i) - c(t_{i-1})\| = \sum_{i=1}^{k} \| \int_{t_{i-1}}^{t_i} \dot{c}(t)\,dt\| \geq$$

$$\sum_{i=1}^{k} \left(\|\dot{c}(\tau_i)\| - \frac{\epsilon}{2(b-a)}\right)(t_i - t_{i-1}) \geq \left(\sum_{i=1}^{k} \|\dot{c}(\tau_i)\|(t_i - t_{i-1})\right) - \frac{\epsilon}{2} \geq \int_a^b \|\dot{c}(t)\|\,dt - \epsilon \quad ,$$

was zu beweisen war. □

Bemerkung 2.14 *Der Satz gilt offenbar auch für nur stückweise stetig differenzierbare $c : [a, b] \to X$. Dabei heißt c stückweise stetig differenzierbar, falls es eine Zerlegung $a = t_0 < t_1 < \cdots < t_k = b$ von $[a, b]$ gibt, so dass c auf den offenen Intervallen (t_{i-1}, t_i), $i = 1, \ldots, k$, stetig differenzierbar ist.*

Angemerkt sei außerdem, dass [Forster II] einen einfacheren Beweis des Satzes angibt, der aber nur im Euklidischen $\mathbb{R}^n$ funktioniert.

Man bemerke, dass die Länge einer stetig differenzierbaren Parametrisierung $c : [a, b] \to X$ einer Kurve nicht von der Parametrisierung abhängt, sondern nur von der Kurve

$C \subset X$ selbst. Denn ist $\phi : [a', b'] \to [a, b]$ eine stetig differenzierbare Parametertransformation mit stetig differenzierbarer Umkehrabbildung ϕ^{-1}, dann gilt

$$\int_{a'}^{b'} \|\frac{d}{dt}(c \circ \phi)\| \, ds = \int_{a'}^{b'} \|\dot{c}(\phi(s))\| \cdot |\phi'(s)| \, ds = \int_a^b \|\dot{c}(t)\| \, dt$$

nach der Kettenregel und der Substitutionsregel.

Eine ausgezeichnete Parametrisierung von C ist die Parametrisierung auf Bogenlänge. Ist $c : [a, b] \to X$ eine beliebige stetig differenzierbare Parametrisierung von C mit $\dot{c}(t) \neq 0$ für alle $t \in [a, b]$, so ist $\phi(t) := \int_a^t \|\dot{c}(\tilde{t})\| \, d\tilde{t}$ wegen $\phi'(t) = \|\dot{c}(t)\|$ eine stetig differenzierbare Parametertransformation mit stetig differenzierbarer Umkehrabbildung. Die Parametrisierung $\gamma := c \circ \phi^{-1}$ von C hat dann die Ableitung $\dot{\gamma}(s) = \frac{\dot{c}(\phi^{-1}(s))}{\|\dot{c}(\phi^{-1}(s))\|}$, also gilt $\|\dot{\gamma}(s)\| = 1$ für alle s. Somit ist $\int_0^s \|\dot{\gamma}(s)\| = s$, d.h. der Parameter s der Parametrisierung $\gamma : [0, L(c)] \to X$ gibt gerade die Länge des Kurvenstücks von $\gamma(0)$ nach $\gamma(s)$ an. Diese Parametrisierung auf Bogenlänge spielt eine große Rolle in der Relativitätstheorie und definiert dort die Eigenzeit eines Beobachters.

2.2 Partiell differenzierbare Abbildungen

Im Gegensatz zur Theorie differenzierbarer Kurven gibt es bei auf mehrdimensionalen Räumen definierten Abbildungen verschiedene Differentiationsbegriffe. Wir beginnen mit dem schwächsten Begriff, der partiellen Differenzierbarkeit.

Definition 2.15 *Eine Abbildung $f : X \to Y$ zwischen Banach-Räumen X, Y heißt im Punkt $a \in X$ partiell differenzierbar in Richtung $h \in X$, falls der Grenzwert $\partial_h f(a) := \lim\limits_{t \to 0} \frac{f(a+th)-f(a)}{t}$ in Y existiert, und $\partial_h f(a)$ bezeichnet man als partielle Ableitung von f im Punkt a in Richtung h.*

Ist f im Punkt a in jede Richtung partiell differenzierbar, so nennt man f partiell differenzierbar in a, und ist die Abbildung $h \mapsto \partial_h f(a)$ von X nach Y darüberhinaus noch linear und stetig, dann heißt f Gâteaux-differenzierbar in a.

Man bemerke, dass nach dieser Definition f genau dann im Punkt a partiell differenzierbar in Richtung h ist, wenn die durch $c(t) := f(a + th)$ definierte parametrisierte Kurve $c : \mathbb{R} \to Y$ in 0 differenzierbar ist, und es gilt $\partial_h f(a) = \dot{c}(0)$.

Insbesondere ist eine Funktion $f : X \to \mathbb{R}$ genau dann im Punkt a partiell differenzierbar in Richtung h, wenn die Funktion $t \mapsto f(a + th)$ in 0 im Sinne der Analysis I differenzierbar ist.

Im Falle einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$ bedeutet dies, dass die partielle Ableitung im Punkt $a \in \mathbb{R}^n$ in die i-te Koordinatenrichtung e_i (wobei $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ den i-ten Einheitsvektor des $\mathbb{R}^n$ bezeichnet) genau dann existiert, wenn die Funktion $\phi : \xi \mapsto f(a_1, \ldots, a_{i-1}, \xi, a_{i+1}, \ldots, a_n)$ auf $\mathbb{R}$, bei der alle Koordinaten bis auf die i-te Komponente festgehalten wurden, in a_i differenzierbar ist. In diesem Fall symbolisiert man die partielle Ableitung $\partial_{e_i} f(a)$ üblicherweise durch $\frac{\partial f}{\partial x_i}(a)$, und mit der zuvor definierten Funktion ϕ auf $\mathbb{R}$ gilt $\frac{\partial f}{\partial x_i}(a) = \phi'(a_i)$. Ist $f : \mathbb{R}^n \to \mathbb{R}$ sogar Gâteaux-differenzierbar, so kann man wegen der Linearität von $h \mapsto \partial_h f(a)$ die partielle Ableitung in Richtung h mittels der partiellen Ableitungen in Koordinatenrichtung durch $\partial_h f(a) = \sum\limits_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) h_i$ berechnen.

Der Begriff der partiellen Ableitung einer Abbildung f im Punkt a macht natürlich auch noch dann Sinn, wenn f nur auf einer Umgebung $U \subset X$ von a definiert ist und nicht auf dem ganzen Banach-Raum X. Man kann also auch bei $f : U \to Y$, $U \subset X$ Umgebung von $a \in U$, über die partielle Differenzierbarkeit von f in a sprechen.

Darüberhinaus nennt man eine Abbildung f auf X (oder einer offenen Teilmenge $U \subset X$) partiell differenzierbar bzw. Gâteaux-differenzierbar, wenn f in jedem Punkt $a \in X$ (oder $a \in U$) partiell differenzierbar bzw. Gâteaux-differenzierbar ist.

Beispiel 2.16 *Im Euklidischen $\mathbb{R}^n$ ist die Euklidische Norm als Funktion $f : X \to \mathbb{R}$, $f(x) := \|x\|_2 = \sqrt{\sum\limits_{i=1}^{n} x_i^2}$, in jedem Punkt $a \neq 0$ partiell differenzierbar in jede Koordinatenrichtung e_i, denn die Funktion $\xi \mapsto \left(a_1^2 + \cdots + a_{i-1}^2 + \xi^2 + a_{i+1}^2 + \cdots + a_n^2\right)^{1/2}$ auf $\mathbb{R}$ ist bei $a \neq 0$ in a_i differenzierbar und hat dort nach den üblichen Differentiationsregeln aus der Analysis I die Ableitung*

$$\frac{\partial f}{\partial x_i}(a) = \frac{2a_i}{2\left(a_1^2 + \cdots + a_{i-1}^2 + a_i^2 + a_{i+1}^2 + \cdots + a_n^2\right)^{1/2}} = \frac{a_i}{\|a\|_2} \quad , \quad i = 1, \ldots, n \quad .$$

Tatsächlich ist die Euklidische Norm in Punkten $a \neq 0$ sogar partiell differenzierbar in jede Richtung und auch Gâteaux-differenzierbar. Dies beweisen wir gleich im allgemeineren Kontext eines beliebigen Hilbert-Raumes $(X, \langle\cdot,\cdot\rangle)$ für $f(x) := \|x\| = \sqrt{\langle x, x\rangle}$: Die Funktion

$$t \mapsto f(a + th) = \sqrt{\langle a + th, a + th\rangle} = \left(\|a\|^2 + 2t\langle a, h\rangle + t^2\|h\|^2\right)^{1/2}$$

ist bei $a \neq 0$ nach den üblichen Differentiationsregeln aus der Analysis I in $t = 0$ differenzierbar mit Ableitung

$$\partial_h f(a) = \frac{2\langle a, h\rangle}{2\left(\|a\|^2\right)^{1/2}} = \frac{\langle a, h\rangle}{\|a\|} \quad .$$

Offensichtlich ist $h \mapsto \partial_h f(a) = \frac{\langle a,h\rangle}{\|a\|}$ linear, und die Stetigkeit dieser linearen Abbildung folgt aus der Cauchy-Schwarzschen Ungleichung $|\langle a, h\rangle| \leq \|a\|\|h\|$, nach der die lineare Abbildung $h \mapsto \partial_h f(a)$ die Operatornorm $\frac{\|a\|}{\|a\|} = 1 < \infty$ besitzt. Also ist f Gâteaux-differenzierbar auf $X \setminus \{0\}$, da $\partial_h f(a)$ bei $a \neq 0$ für jedes $h \in X$ existiert sowie linear und stetig in h ist.

Die folgende Überlegung ist nützlich, da sie insbesondere besagt, dass eine Abbildung f mit Werten im $\mathbb{R}^n$ genau dann partiell differenzierbar ist, wenn ihre Komponentenfunktionen partiell differenzierbar sind.

Lemma 2.17 *Seien X, Y, Z Banach-Räume. Eine Abbildung $f = (f_1, f_2) : X \to Y \times Z$ ist genau dann im Punkt $a \in X$ partiell differenzierbar in Richtung h, wenn die Abbildungen f_1 und f_2 im Punkt $a \in X$ partiell differenzierbar in Richtung h sind, und es gilt $\partial_h f(a) = (\partial_h f_1(a), \partial_h f_2(a))$.*

Beweis: Nach Lemma 1.24 existiert der Grenzwert $\lim\limits_{t\to 0} \frac{f(a+th)-f(a)}{t}$ genau dann, wenn die Grenzwerte $\lim\limits_{t\to 0} \frac{f_1(a+th)-f_1(a)}{t}$ und $\lim\limits_{t\to 0} \frac{f_2(a+th)-f_2(a)}{t}$ existieren, und mit den entsprechenden Bezeichnungen für die Grenzwerte gilt $\partial_h f(a) = (\partial_h f_1(a), \partial_h f_2(a))$. □

Abgesehen davon, dass man partielle Ableitungen oft leicht ausrechnen kann, ist die alleinige Existenz von partiellen Ableitungen ein zu schwacher Begriff für eine weitreichende Analysis. Denn aus der partiellen Differenzierbarkeit und selbst aus der Gâteaux-Differenzierbarkeit einer Abbildung folgt weder die Stetigkeit, noch ist die Kettenregel gültig.

Beispiel 2.18 *Die Funktion $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) := \frac{\prod_{i=1}^n x_i}{\|x\|_2^{2n}}$ bei $x \neq 0$ und $f(0) := 0$, ist nicht nur in allen Punkten $a \neq 0$ partiell differenzierbar, sondern auch im Nullpunkt in die Koordinatenrichtungen e_i partiell differenzierbar, denn wegen $f(te_i) = 0$ für alle $t \in \mathbb{R}$ gilt*

$$\frac{\partial f}{\partial x_i}(0) = \lim_{t\to 0} \frac{f(te_i) - 0}{t} = 0 \quad .$$

Aber f ist nicht stetig im Nullpunkt, denn beispielsweise existiert der Grenzwert von $f(t, t, \ldots, t) = \frac{t^n}{|t|^{2n} n^n}$ für $t \to 0$ nicht und ist somit insbesondere nicht gleich $f(0) = 0$.

Aufgabe:

- *Zeigen Sie, dass die durch $f(x, y) := \frac{xy^2}{x^2+y^4}$ für $(x, y) \neq (0, 0)$ und $f(0, 0) := 0$ definierte Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ im Nullpunkt partielle Ableitungen in jede Richtung besitzt, aber dass $h \mapsto \partial_h f(0)$ nicht linear ist und somit f in 0 nicht Gâteaux-differenzierbar ist, sowie f im Nullpunkt nicht stetig ist.*

- *Zeigen Sie, dass die durch $g(x, y) := \frac{2xe^{-1/y^2}}{x^2+e^{-2/x^2}}$ für $(x, y) \neq (0, 0)$ und $g(0, 0) := 0$ definierte Funktion $g : \mathbb{R}^2 \to \mathbb{R}$ zwar im Nullpunkt Gâteaux-differenzierbar ist, aber dort trotzdem nicht stetig ist.*

Beispiel 2.19 *Die parametrisierte Kurve $f : \mathbb{R} \to \mathbb{R}^2$, $f(t) := (t, t^3)$, ist offensichtlich (Gâteaux-)differenzierbar, und auch die Funktion $g : \mathbb{R}^2 \to \mathbb{R}$, $g(x, y) := \frac{x^3 y}{x^6+y^2}$ für $(x, y) \neq (0, 0)$ sowie $g(0, 0) := 0$, ist Gâteaux-differenzierbar. Denn die Gâteaux-Differenzierbarkeit in allen Punkten ungleich $(0, 0)$ ist trivial, und im Nullpunkt ist g Gâteaux-differenzierbar mit partieller Ableitung $\partial_h g(0, 0) = 0$ für alle $h \in \mathbb{R}^2$ wegen*

$$\partial_h g(0,0) = \lim_{t\to 0} \frac{\frac{t^4 h_1^3 h_2}{t^6 h_1^6 + t^2 h_2^2} - 0}{t} = \lim_{t\to 0} t \frac{h_1^2 h_2^2}{t^4 h_1^6 + h_2^2} = 0 \quad ,$$

wobei man beachte, dass bei $h_2 = 0$ der Zähler verschwindet.

Aber die Verkettung $g \circ f : \mathbb{R} \to \mathbb{R}$ ist in 0 nicht differenzierbar, denn

$$(g \circ f)(t) = g(f(t)) = \frac{t^3 t^3}{t^6 + (t^3)^2} = \frac{1}{2}$$

gilt für jedes $t \neq 0$, jedoch $(g \circ f)(0) = g(f(0)) = g(0,0) = 0$, und somit ist $g \circ f$ nicht einmal stetig in 0, also auch nicht differenzierbar.

Trotz dieser Defekte des Begriffs der partiellen Ableitung kann man aber für partiell differenzierbare Abbildungen mit Werten in $\mathbb{R}$ zumindest den Mittelwertsatz beweisen.

Satz 2.20 *Sei $U \subset X$ eine offene Teilmenge eines Banach-Raumes X und seien $x, \tilde{x} \in U$ Punkte, für die die Strecke $C := \{x + t(\tilde{x} - x) \,|\, t \in [0,1]\}$ in U enthalten ist. Ist die Funktion $f : U \to \mathbb{R}$ in jedem Punkt der Strecke C partiell differenzierbar in Richtung $\tilde{x} - x$, dann gibt es ein $\theta \in (0,1)$ mit $f(\tilde{x}) - f(x) = \partial_{(\tilde{x}-x)} f(x + \theta(\tilde{x} - x))$.*

Beweis: Nach Vorraussetzung ist die Funktion $\phi : [0,1] \to \mathbb{R}$, $\phi(t) := f(x + t(\tilde{x} - x))$, differenzierbar. Somit gibt es nach dem Mittelwertsatz der Analysis I ein $\theta \in (0,1)$ mit $\phi(1) - \phi(0) = \phi'(\theta)$, also mit $f(\tilde{x}) - f(x) = \partial_{(\tilde{x}-x)} f(x + \theta(\tilde{x} - x))$. □

Wie wir schon in Beispiel 2.4 gesehen haben, gilt der Mittelwertsatz für Abbildungen mit Werten in einem mehrdimensionalen Raum in dieser Form nicht. Wir werden jedoch in Satz 2.32 eine Verallgemeinerung kennenlernen, die auch in diesem Fall anwendbar ist.

Ableitungen höherer Ordnung Ganz analog zur Analysis I kann man für eine partiell differenzierbare Funktion $f : U \to Y$, $U \subset X$ offen und X, Y Banach-Räume, partielle Ableitungen höherer Ordnung definieren: Ist f in jedem Punkt aus U partiell differenzierbar in Richtung h und ist auch die partielle Ableitung $\partial_h f$ als Funktion von U nach Y in jedem Punkt aus U partiell differenzierbar in Richtung $\tilde{h}$, dann heißt f zweimal partiell differenzierbar und $\partial_{\tilde{h}} \partial_h f(a)$ wird die zweite partielle Ableitung im Punkt $a \in U$ in Richtung $(h, \tilde{h})$ genannt. Per Induktion kann man dann analog partielle Ableitungen beliebig hoher Ordnungen definieren.

Leitet man dabei eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ in Koordinatenrichtungen ab, so verwendet man üblicherweise das Symbol $\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}$ für die k-te partielle Ableitung in die Richtungen $(e_{i_1}, \ldots, e_{i_k})$.

Bei Ableitungen höherer Ordnung offenbart sich ein weiterer Defekt des Begriffs der partiellen Ableitung, denn i.a. gilt $\frac{\partial^2 f}{\partial x_j \partial x_i} \neq \frac{\partial^2 f}{\partial x_i \partial x_j}$.

Beispiel 2.21 *Die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) := \frac{x^3 y - x y^3}{x^2 + y^2}$ für $(x,y) \neq (0,0)$ und $f(0,0) := 0$, ist Gâteaux-differenzierbar mit partiellen Ableitungen $\frac{\partial f}{\partial x}(x,y) = \frac{(x^4 + 4x^2y^2 - y^4)y}{(x^2+y^2)^2}$ und $\frac{\partial f}{\partial y}(x,y) = \frac{(x^4 - 4x^2y^2 - y^4)x}{(x^2+y^2)^2}$ in Koordinatenrichtungen für $(x,y) \neq (0,0)$*

sowie $\frac{\partial f}{\partial x}(0,0) = 0 = \frac{\partial f}{\partial y}(0,0)$. *Die gemischten zweiten partiellen Ableitung im Nullpunkt existieren, sie sind aber wegen*

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = \lim_{t \to 0} \frac{\frac{\partial f}{\partial x}(0,t) - 0}{t} = -1$$

und

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{t \to 0} \frac{\frac{\partial f}{\partial y}(t,0) - 0}{t} = 1$$

voneinander verschieden.

Notwendiges Kriterium für lokale Extrema Sei $U \subset X$ eine offene Teilmenge eines Banach-Raumes X und $f : U \to \mathbb{R}$ eine Funktion. Man sagt, f hat in $x^* \in U$ ein lokales Minimum, wenn es eine Umgebung V von x^* mit $f(x^*) \leq f(x)$ für alle $x \in V$ gibt. Analog definiert man lokale Maxima.

Satz 2.22 *Hat eine Funktion* $f : U \to \mathbb{R}$ *auf einer offenen Teilmenge* $U \subset X$ *eines Banach-Raumes* X *in* $x^* \in U$ *ein lokales Extremum und ist* f *in* x^* *partiell differenzierbar in Richtung* h*, dann gilt* $\partial_h f(x^*) = 0$.

Beweis: Die Funktion $\phi(t) := f(x^* + th)$ auf $\mathbb{R}$ hat in $t = 0$ ein lokales Exremum und ist dort differenzierbar, also gilt $0 = \phi'(0) = \partial_h f(x^*)$. □

Jedoch ist das Verschwinden aller Ableitungen $\partial_h f(x^*)$, $h \in X$, nur ein notwendiges und kein hinreichendes Kriterium für das Vorliegen eines Extremums. Hinreichende Kriterien lernen wir in Satz 2.47 kennen.

Allgemein nennt man einen Punkt x^* mit $\partial_h f(x^*) = 0$ für alle $h \in X$ einen stationären Punkt, egal ob in x^* ein lokales Extremum vorliegt oder nicht.

2.3 Differenzierbare Abbildungen

Wie wir im vorigen Abschnitt gesehen haben, folgt weder aus der partiellen Differenzierbarkeit noch aus der Gâteaux-Differenzierbarkeit einer Abbildung $f : X \to Y$ zwischen Banach-Räumen X, Y die Stetigkeit von f, und auch die Kettenregel ist i.a. nicht gültig. Deswegen führen wir im folgenden einen stärkeren Begriff von Differenzierbarkeit ein.

Definition 2.23 *Eine Abbildung* $f : X \to Y$ *zwischen Banach-Räumen* X, Y *heißt differenzierbar*[2] *im Punkt* $a \in X$*, wenn es eine lineare und stetige Abbildung* $A : X \to Y$ *gibt mit*

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - Ah}{\|h\|} = 0 \quad .$$

[2] In manchen Büchern sagt man zur besseren Unterscheidung statt differenzierbar auch total differenzierbar oder Fréchet-differenzierbar.

Man beachte, dass hierbei h in X beliebig gegen Null konvergiert und nicht entlang von Geraden, wie dies noch bei der partiellen Differenzierbarkeit der Fall war.

Anschaulich bedeutet die Definition, dass es eine affin-lineare Abbildung von X nach Y gibt, nämlich $x \mapsto f(a) + A(x-a)$, die sich an den Graphen von f in dem Sinne anschmiegt, als dass die Differenz von $f(x)$ und dieser affin-linearen Abbildung für $x \to a$ schneller abnimmt als die Norm $\|x-a\|$.

Man nennt deshalb A auch die Linearisierung von f im Punkt a. Üblicherweise bezeichnet man A jedoch als die (totale oder Fréchet-) Ableitung von f im Punkt a und symbolisiert Sie durch $df(a) := A$. Diese Bezeichnung macht natürlich nur Sinn, wenn A eindeutig durch f bestimmt ist und mit der partiellen Ableitung übereinstimmt, was der folgende Satz garantiert.

Satz 2.24 *Ist die Abbildung $f : X \to Y$ zwischen Banach-Räumen X, Y differenzierbar im Punkt $a \in X$, dann ist f in a partiell differenzierbar in jede Richtung $h \in X$ und mit A aus Definition 2.23 gilt $Ah = \partial_h f(a)$.*

Beweis: Ist $A : X \to Y$ eine lineare und stetige Abbildung mit

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - Ah}{\|h\|} = 0 \quad ,$$

dann gilt insbesondere für $0 \neq v \in X$ die Gleichung

$$\lim_{t\searrow 0} \frac{f(a+tv) - f(a) - A(tv)}{\|tv\|} = 0 \quad ,$$

wobei $t \in \mathbb{R}^+$ ist, und somit

$$\lim_{t\searrow 0} \frac{f(a+tv) - f(a)}{t} = Av \quad .$$

Analog folgt $\lim_{t\nearrow 0} \frac{f(a+tv)-f(a)}{t} = Av$, und dies beweist die Existenz der partiellen Ableitung von f im Punkt a in Richtung v sowie die Gültigkeit der Gleichung

$$\partial_v f(a) = \lim_{t\to 0} \frac{f(a+tv) - f(a)}{t} = Av \quad .$$

□

Insbesondere ist nach diesem Satz A eindeutig bestimmt (denn A ist durch die partiellen Ableitungen von f gegeben) und f auch Gâteaux-differenzierbar in a (denn $h \mapsto \partial_h f(a) = Ah$ ist linear und stetig). Wir wollen aus dem vorigen Satz noch einmal explizit die Gleichung

$$df(a)h = \partial_h f(a)$$

zwischen der (totalen) Ableitung $df(a) = A$ und den partiellen Ableitungen $\partial_h f(a)$ festhalten.

Der Satz liefert darüberhinaus auch noch eine direkte Methode, Differenzierbarkeit einer Abbildung nachzuweisen: Um zu überprüfen, ob eine Abbildung f im Punkt a

differenzierbar ist, muss man zunächst die partiellen Ableitungen von f berechnen und nachschauen, ob die Abbildung $A : h \mapsto \partial_h f(a)$ linear und stetig ist, d.h. f Gâteaux-differenzierbar ist. Dies reicht aber für die Differenzierbarkeit noch nicht, sondern abschließend muss man noch überprüfen, ob mit der so gewonnenen linearen und stetigen Abbildung $A : X \to Y$ auch die Grenzwertbeziehung

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - Ah}{\|h\|} = 0$$

gilt.

Angemerkt sei, dass Definition 2.23 natürlich auch für nur auf offenen Teilmengen $U \subset X$ definierte Abbildungen $f : U \to Y$ Sinn macht und alles zuvor Gesagte entsprechend für solche Abbildungen gilt. Außerdem nennt man eine Abbildung $f : U \to Y$ auf einer offenen Teilmenge $U \subset X$ differenzierbar, wenn sie in jedem Punkt $a \in U$ differenzierbar ist.

Beispiel 2.25

- *Jede affin-lineare Abbildung* $f : X \to Y$, $f(x) := Ax + b$ *mit einer linearen und stetigen Abbildung* $A : X \to Y$ *und einem Vektor* $b \in Y$, *ist differenzierbar mit* $df(x) = A$.
- *Ist* $(X, \langle\cdot,\cdot\rangle)$ *ein Hilbert-Raum und* $A : X \to X$ *eine lineare und stetige Abbildung, dann ist* $f : X \to \mathbb{R}$, $f(x) := \langle Ax, x\rangle$, *differenzierbar mit Ableitung* $df(x)h = \langle Ax, h\rangle + \langle Ah, x\rangle$.
- *Jede differenzierbare Kurve* $c : I \to X$ *auf einem offenen Intervall* I *ist auch nach Definition 2.23 differenzierbar mit Ableitung* $dc(x)h = h\dot{c}(x)$. *Denn es gilt* $\dot{c}(x) = \lim\limits_{t\to 0} \frac{c(x+t)-c(x)}{t}$ *genau dann, wenn* $\lim\limits_{h\to 0} \frac{c(x+h)-c(x)-h\dot{c}(x)}{|h|} = 0$ *gilt.*

Hauptsächlich wollen wir uns im folgenden mit der Ableitung von Abbildungen $f : U \to \mathbb{R}^n$ auf offenen Teilmengen $U \subset \mathbb{R}^m$ beschäftigen, d.h. mit der Ableitung von Abbildungen zwischen endlich-dimensionalen Räumen. Man bemerke, dass in diesem Fall der Begriff der Differenzierbarkeit unabhängig von der Norm ist, da alle Normen äquivalent sind. Ist f differenzierbar im Punkt $a \in U$, dann ist die Ableitung $df(a)$ eine lineare und stetige Abbildung von $\mathbb{R}^m$ nach $\mathbb{R}^n$. Solche linearen und stetigen Abbildungen können mittels Basen bekanntlich durch Matrizen repräsentiert werden. Wählt man die Einheitsvektoren als Basis und ist $f = (f_1, \ldots, f_n)$, so wird $df(a)$ wegen

$$(df(a)e_j)_i = \left(\frac{\partial f}{\partial x_j}(a)\right)_i = \frac{\partial f_i}{\partial x_j}(a) \quad (i = 1, \ldots, n\,,\; j = 1, \ldots, m)$$

durch die $(m \times n)$-Matrix

$$Jf(a) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_m}(a) \\ \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_1}(a) & \cdots & \frac{\partial f_n}{\partial x_m}(a) \end{pmatrix}$$

repräsentiert, die Jacobi-Matrix von f im Punkt a genannt wird.

Für in a differenzierbare Funktionen $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^m$ offen, ist die Jacobi-Matrix $Jf(a)$ einfach der Zeilenvektor $(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_m}(a))$. Häufig betrachtet man aber auch den zugehörigen Spaltenvektor, den man den Gradienten von f nennt und mit grad $f(a)$ symbolisiert. Mathematisch präziser entsteht der Gradient grad $f(a)$ aus der Ableitung $df(a)$ einer Funktion durch die Gleichung

$$df(a) = \langle \operatorname{grad} f(a), \cdot \rangle \quad ,$$

die es einem mittels des Euklidischen Skalarproduktes $\langle \cdot, \cdot \rangle$ erlaubt, lineare Funktionen auf dem $\mathbb{R}^m$ (wie $df(a)$ eine ist) mit Vektoren im $\mathbb{R}^m$ zu identifizieren.

Soviel zunächst zur Notation, nun kommen wir zu den Sätzen, die die Stärke des Begriffs der Differenzierbarkeit zeigen.

Satz 2.26 *Seien X, Y Banach-Räume und sei $f : U \to Y$ eine Abbildung auf einer offenen Teilmenge $U \subset X$. Ist f in einem Punkt $a \in U$ differenzierbar, dann ist f auch stetig in a.*

Beweis: Definiert man den Rest $R(h)$ einer in a differenzierbaren Abbildung f mit Ableitung A durch die Gleichung $f(a + h) = f(a) + Ah + R(h)$, dann besagt die Differenzierbarkeitsbedingung gerade $\lim\limits_{h\to 0} \frac{R(h)}{\|h\|} = 0$. Insbesondere gilt $R(h) \to 0$ für $h \to 0$, und da wegen der Stetigkeit von A auch $Ah \to 0$ für $h \to 0$ gilt, folgt aus $f(a+h) = f(a)+Ah+R(h)$ auch $\lim\limits_{h\to 0} f(a+h) = f(a)$, d.h. die Stetigkeit von f in a. $\square$

Aus Differenzierbarkeit folgt also sowohl partielle Differenzierbarkeit als auch Stetigkeit, aber jede andere Implikation zwischen diesen drei Begriffen ist i.a. falsch.

Ganz analog zum Fall partiell differenzierbarer Abbildungen gelten auch für differenzierbare Abbildungen die üblichen algebraischen Rechenregeln. Beginnen wollen wir aber mit dem Analogon von Lemma 2.17.

Lemma 2.27 *Seien X, Y, Z Banach-Räume. Eine Abbildung $f = (f_1, f_2) : U \to Y \times Z$ auf einer offenen Teilmenge $U \subset X$ ist genau dann im Punkt $a \in U$ differenzierbar, wenn die Abbildungen f_1 und f_2 im Punkt a differenzierbar sind, und es gilt $df(a) = (df_1(a), df_2(a))$.*

Beweis: Die Abbildung f ist genau dann differenzierbar in $a \in U$, wenn es eine stetige lineare Abbildung $A : X \to Y \times Z$ gibt, für die der durch $f(a+h) = f(a) + Ah + R(h)$ definierte Rest R die Beziehung $\lim\limits_{h\to 0} \frac{R(h)}{\|h\|} = 0$ in $X \times Y$ erfüllt. Schreibt man $A = (A_1, A_2)$ und $R = (R_1, R_2)$, so ist dies aber äquivalent dazu, dass $f_i(a + h) = f_i(a) + A_i h + R_i(h)$ und $\lim\limits_{h\to 0} \frac{R_i(h)}{\|h\|} = 0$ für $i = 1, 2$ gilt (letzteres nach Lemma 1.24). Dies beweist die Hauptaussage des Lemmas, und wegen $df(a) = A$ sowie $df_i(a) = A_i$ gilt auch die Formel $df(a) = (df_1(a), df_2(a))$. $\square$

Insbesondere ist eine Abbildung $f = (f_1, \ldots, f_n) : U \to \mathbb{R}^n$, $U \subset \mathbb{R}^m$ offen, genau dann in $a \in U$ differenzierbar, wenn alle Komponentenfunktionen $f_i : U \to \mathbb{R}$ differenzierbar sind.

Lemma 2.28 *Seien X, Y Banach-Räume und sei $U \subset X$ offen.*

(a) Sind $f, g : U \to Y$ differenzierbar in $a \in U$ und sind $\lambda, \mu \in \mathbb{R}$ Konstanten, dann ist auch $\lambda f + \mu g$ differenzierbar in a mit Ableitung $d(\lambda f + \mu g)(a) = \lambda\, df(a) + \mu\, dg(a)$.

(b) Sind $f : U \to Y$ und $\lambda : U \to \mathbb{R}$ differenzierbar in $a \in U$, dann ist λf differenzierbar in a mit Ableitung $d(\lambda f)(a)h = d\lambda(a)h \cdot f(a) + \lambda(a) \cdot df(a)h$.

Beweis:

(a) Sind die Abbildungen f, g differenzierbar in $a \in X$, dann gilt für die durch $f(a+h) = f(a) + df(a)h + R_1(h)$ und $g(a+h) = g(a) + dg(a)h + R_2(h)$ definierten Reste R_i die Beziehung $\lim\limits_{h \to 0} \frac{R_i(h)}{\|h\|} = 0$ für $i = 1, 2$. Insbesondere gilt auch $\lim\limits_{h \to 0} \frac{(\lambda R_2 + \mu R_2)(h)}{\|h\|} = 0$, also hat $\lambda f + \mu g$ wegen

$$(\lambda f + \mu g)(a+h) = \lambda f(a) + \mu g(a) + \lambda df(a)h + \mu dg(a)h + \lambda R_1(h) + \mu R_2(h)$$

die Ableitung $h \mapsto \lambda df(a)h + \mu dg(a)h$ in a.

(b) Sind die Abbildungen f und λ differenzierbar in $a \in X$, dann gilt für die durch $f(a+h) = f(a) + df(a)h + R_1(h)$ und $\lambda(a+h) = \lambda(a) + d\lambda(a)h + R_2(h)$ definierten Reste R_i die Beziehung $\lim\limits_{h \to 0} \frac{R_i(h)}{\|h\|} = 0$ für $i = 1, 2$. Ausmultiplizieren liefert

$$\begin{gathered}(\lambda f)(a+h) = \lambda(a)f(a) + d\lambda(a)h \cdot f(a) + \lambda(a) \cdot df(a)h + \\ d\lambda(a)h \cdot df(a)h + (\lambda(a) + d\lambda(a)h)\, R_1(h) + R_2(h)\, (f(a) + df(a)h) + R_1(h)R_2(h) \quad .\end{gathered}$$

Bezeichnet man die zweite Zeile als $R(h)$, so gilt $\lim\limits_{h \to 0} \frac{R_i(h)}{\|h\|} = 0$, weil die Ungleichung $|d\lambda(a)h \cdot df(a)h| \leq \|d\lambda(a)\| \|df(a)\| \|h\|^2$ gilt und die einzelnen R_i dieser Beziehung genügen. Also ist λf differenzierbar in a, und die Ableitung in a ist die lineare und stetige Abbildung $h \mapsto d\lambda(a)h \cdot f(a) + \lambda(a) \cdot df(a)h$.

□

Nun kommen wir zur Kettenregel, die – wie schon erwähnt – für nur partiell differenzierbare oder Gâteaux-differenzierbare Abbildungen nicht allgemein gilt, wohl aber für differenzierbare Abbildungen.

Satz 2.29 *Seien X, Y, Z Banach-Räume und $U \subset X$, $V \subset Y$ offen. Ist $f : U \to V$ differenzierbar in $a \in U$ und $g : V \to Z$ differenzierbar in $f(a)$, dann ist $g \circ f$ differenzierbar in a mit Ableitung $d(g \circ f)(a) = dg(f(a)) \circ df(a)$.*

Beweis: Ist f in a und g in $f(a)$ differenzierbar, dann gilt für die durch $f(a+h) = f(a) + df(a)h + R_1(h)$ und $g(f(a) + \tilde{h}) = g(f(a)) + dg(f(a))\tilde{h} + R_2(\tilde{h})$ definierten Reste R_1, R_2 die Beziehung $\lim\limits_{h \to 0} \frac{R_1(h)}{\|h\|_X} = 0$ und $\lim\limits_{\tilde{h} \to 0} \frac{R_2(\tilde{h})}{\|\tilde{h}\|_Y} = 0$.

Mit $\tilde{h} := df(a)h + R_1(h)$ folgt

$$g(f(a+h)) = g(f(a) + \tilde{h}) = g(f(a)) + dg(f(a))\tilde{h} + R_2(\tilde{h}) =$$
$$g(f(a)) + dg(f(a))(df(a)h) + dg(f(a))(R_1(h)) + R_2(df(a)h + R_1(h)) \quad ,$$

und zu zeigen ist nur noch, dass $R(h) := dg(f(a))(R_1(h)) + R_2(df(a)h + R_1(h))$ die Beziehung $\lim\limits_{h\to 0} \frac{R(h)}{\|h\|_X} = 0$ erfüllt. Nun ist aber einerseits $dg(f(a))$ stetig und daher folgt

$$\lim_{h\to 0} \frac{dg(f(a))(R_1(h))}{\|h\|_X} = dg(f(a)) \left(\lim_{h\to 0} \frac{(R_1(h))}{\|h\|_X} \right) = 0 \quad .$$

Andererseits folgt wegen $\|\tilde{h}\|_Y \leq \|df(a)\| \|h\|_X + \|R_1(h)\|_Y = \|h\|_X \left(\|df(a)\| + \frac{\|R_1(h)\|_Y}{\|h\|_X} \right)$ auch

$$\frac{\|R_2(df(a)h + R_1(h))\|_Z}{\|h\|_X} \leq \left(\|df(a)\| + \frac{\|R_1(h)\|_Y}{\|h\|_X} \right) \frac{\|R_2(\tilde{h})\|}{\|\tilde{h}\|_Y} \to 0$$

bei $h \to \infty$. □

Insbesondere gilt bei $X = \mathbb{R}^k$, $Y = \mathbb{R}^m$ und $Z = \mathbb{R}^n$ für die Jacobi-Matrizen einer in $a \in U$ differenzierbaren Abbildung $f : U \subset V$ und einer in $f(a)$ differenzierbaren Abbildung $g : V \to Z$ die Gleichung

$$J(g \circ f)(a) = Jg(f(a)) \cdot Jf(a)$$

mit der Multiplikation $\cdot$ von Matrizen.

Beispiel 2.30 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine differenzierbare Funktion und $c : I \to \mathbb{R}^n$ eine differenzierbare Kurve, die innerhalb der Niveaufläche $M := \{x \in \mathbb{R}^n \,|\, f(x) = C\}$ verläuft. Dann gilt für jedes $t \in I$ also $f(c(t)) = C$ und somit nach der Kettenregel $df(c(t))\dot{c}(t) = 0$ für jedes $t \in I$. Also gilt nach Definition des Gradienten mit dem Euklidischen Skalarprodukt $\langle (\mathrm{grad}\ f)(c(t)), \dot{c}(t) \rangle = 0$, d.h. der Gradient von f steht senkrecht auf jedem Tangentialvektor an die Niveaufläche M.*

Tatsächlich zeigt der Gradient sogar in Richtung des (steilsten) Anstiegs der Funktion, denn ist c eine Kurve, für die $t \mapsto f(c(t))$ monoton wächst, dann ist der Winkel zwischen $(\mathrm{grad}\ f)(c(t))$ und $\dot{c}(t)$ wegen $\langle (\mathrm{grad}\ f)(c(t)), \dot{c}(t) \rangle \geq 0$ spitz.

Aufgabe:

- *Beweisen Sie für allgemeine stetige bilineare Abbildungen $B : X_1 \times X_2 \to Y$ (X_1, X_2, Y Banach-Räume) die Produktregel $dB(a_1, a_2)(h_1, h_2) = B(a_1, h_2) + B(h_1, a_2)$.*
- *Beweisen Sie für in $a \in U$ differenzierbare Funktionen $f, g : U \to \mathbb{R}$, $U \subset X$ offen und $g(a) \neq 0$, die Quotientenregel $d\left(\frac{f}{g}\right)(a) = \frac{g(a) \cdot df(a) - f(a) \cdot dg(a)}{g(a)^2}$ durch Anwendung der der Produktregel und der Kettenregel auf $h \circ g$ bei $h : \mathbb{R} \setminus \{0\} \to \mathbb{R}$, $h(y) := \frac{1}{y}$.*

2.4 Stetig differenzierbare Abbildungen

Ein noch stärkerer Begriff als die Differenzierbarkeit ist die stetige Differenzierbarkeit.

Definition 2.31 *Eine Abbildung $f : X \to Y$ zwischen Banach-Räumen X, Y heißt stetig differenzierbar, wenn f (total) differenzierbar ist und die Ableitung $df : X \to L(X, Y)$ stetig bzgl. der Operatornorm auf $L(X, Y)$ ist.*

Natürlich macht diese Definition wiederum auch für nur auf offenen Teilmengen $U \subset X$ definierte Abbildungen $f : U \to Y$ Sinn, dann muß eben $f : U \to Y$ differenzierbar und $df : U \to L(X, Y)$ stetig sein. Den Vektorraum aller auf der offenen Teilmenge $U \subset X$ stetig differenzierbaren Abbildungen $f : U \to Y$ bezeichnet man mit $C^1(U, Y)$.

Darüberhinaus übertragen sich alle Regeln für differenzierbare Abbildungen auf stetig differenzierbare Abbildungen. Beispielsweise ist die Verknüpfung $g \circ f$ stetig differenzierbarer Abbildungen f und g nach der Kettenregel 2.29 nicht nur differenzierbar, sondern sogar stetig differenzierbar, da die Ableitung $x \mapsto dg(f(x)) \circ df(x)$ von $g \circ f$ aufgrund der Stetigkeit von $dg \circ f$, der Stetigkeit von df und der Stetigkeit der Verknüpfung von stetigen linearen Abbildungen ($\|BA\| \leq \|B\|\|A\|$) selbst wieder stetig ist.

Stetige Differenzierbarkeit kann man wesentlich leichter überprüfen als Differenzierbarkeit, denn eine Abbildung ist genau dann stetig differenzierbar, wenn sie stetig partiell differenzierbar ist, d.h. f Gâteaux-differenzierbar ist und die Gâteaux-Ableitung als Abbildung $X \mapsto L(X, Y)$, $x \mapsto (h \mapsto \partial_h f(x))$, stetig ist.

Um dies zu zeigen, beweisen wir zunächst eine integrale Form des Mittelwertsatzes, die im Gegensatz zu Satz 2.20 auch für Abbildungen mit Werten in mehrdimensionalen Räumen gilt. Dabei nutzen wir die im Beweis von Satz 2.13 eingeführten Integrale von Kurven.

Satz 2.32 *Seien X, Y Banach-Räume, sei $U \subset X$ eine offene Teilmenge und seien $x, \tilde{x} \in U$ Punkte, für die die Strecke $C := \{x + t(\tilde{x} - x) \,|\, t \in [0, 1]\}$ in U enthalten ist. Ist die Abbildung $f : U \to Y$ in jedem Punkt der Strecke C partiell differenzierbar in Richtung $\tilde{x} - x$ und ist $t \mapsto \partial_{\tilde{x}-x} f(x + t(\tilde{x} - x))$ stetig auf $[0, 1]$, dann gilt*

$$f(\tilde{x}) - f(x) = \int_0^1 \partial_{\tilde{x}-x} f(x + t(\tilde{x} - x))\, dt \quad .$$

Beweis: Sei $\phi : Y \to \mathbb{R}$ ein stetiges lineares Funktional auf Y, dann ist die Funktion $t \to \phi(f(x + t(\tilde{x} - x)))$ differenzierbar und ihre Ableitung $t \mapsto \phi(\partial_{\tilde{x}-x} f(x + t(\tilde{x} - x)))$ stetig, da f auf C partiell differenzierbar mit stetiger partieller Ableitung ist. Also gilt nach dem Hauptsatz der Differential- und Integralrechnung

$$\phi(f(\tilde{x}) - f(x)) = \phi(f(\tilde{x})) - \phi(f(x)) = \int_0^1 \phi(\partial_{\tilde{x}-x} f(x + t(\tilde{x} - x)))\, dt = \phi\left(\int_0^1 \partial_{\tilde{x}-x} f(x + t(\tilde{x} - x))\, dt\right)$$

Da diese Gleichung für alle stetigen linearen Funktionale $\phi : Y \to \mathbb{R}$ gilt (und somit bei $Y = \mathbb{R}^n$ insbesondere für die Projektionen auf die einzelnen Komponenten), folgt wie gewünscht $f(\tilde{x}) - f(x) = \int_0^1 \partial_{\tilde{x}-x} f(x + t(\tilde{x} - x))\, dt$ (wobei man für unendlich-dimensionale Räume wiederum den schon im Beweis von Satz 2.13 erwähnten Satz von Hahn-Banach anwendet). □

Eine wichtige Konsequenz aus dieser integralen Form des Mittelwertsatzes ist die schon angekündigte Äquivalenz von stetiger Differenzierbarkeit und stetiger partieller Differenzierbarkeit.

Satz 2.33 *Seien X, Y Banach-Räume und $U \subset X$ eine offene Teilmenge. Eine Abbildung $f : U \to Y$ ist genau dann stetig differenzierbar, wenn sie Gâteaux-differenzierbar mit stetiger Gâteaux-Ableitung $X \mapsto L(X,Y)$, $x \mapsto (h \mapsto \partial_h f(x))$, ist.*

Beweis: Sei $a \in U$, dann gilt nach dem Mittelwertsatz in der integralen Form für genügend kleine[3] $h \in X$ wegen $\partial_h f(a) = \int_0^1 \partial_h f(a)\, dt$ die Gleichung

$$f(a+h) - f(a) - \partial_h f(a) = \int_0^1 \partial_h f(a + th) - \partial_h f(a)\, dt \quad .$$

Aufgrund der Stetigkeit von $x \mapsto \partial_{.} f(x)$ gibt es zu jedem $\epsilon > 0$ ein $\delta > 0$ mit $\|\partial_{.} f(x + h) - \partial_{.} f(x)\|_{L(X,Y)} \leq \epsilon$ für alle h mit $\|h\| \leq \delta$. Somit gilt

$$\|f(a+h) - f(a) - \partial_h f(a)\|_Y = \| \int_0^1 \partial_h f(a+th) - \partial_h f(a)\, dt\|_Y \leq$$

$$\int_0^1 \|\partial_h f(a+th) - \partial_h f(a)\|_Y\, dt \leq \int_0^1 \|\partial_{.} f(a+th) - \partial_{.} f(a)\|_{L(X,Y)} \|h\|_X\, dt \leq \epsilon \|h\|_X$$

für alle h mit $\|h\| \leq \delta$, d.h. $\lim_{h\to 0} \frac{f(a+h) - f(a) - \partial_h f(a)}{\|h\|} = 0$, und daher ist f in a differenzierbar. Da außerdem $df(a)h = \partial_h f(a)$ gilt für jedes $a \in U$ gilt und somit nach Voraussetzung $df : U \to L(X,Y)$ stetig ist, ist f sogar stetig differenzierbar auf U. □

Für endlich-dimensionale X, Y ist es sogar noch leichter, die stetige Differenzierbarkeit nachzuprüfen, denn man muß nur die Stetigkeit der partiellen Ableitungen in die Koordinatenrichtungen überprüfen.

Korollar 2.34 *Sei $U \subset \mathbb{R}^m$ offen und $f : U \to \mathbb{R}^n$ eine Abbildung. Existieren die partiellen Abbildungen $\frac{\partial f_j}{\partial x_i}$ in jedem Punkt von U und ist $x \mapsto \frac{\partial f_j}{\partial x_i}(x)$ stetig für jedes $i = 1, \ldots, m$, $j = 1, \ldots, n$, so ist f stetig differenzierbar.*

Beweis: Sei e_i die Standardbasis des $\mathbb{R}^m$, sei $a \in U$ ein beliebiger Punkt und betrachte die Komponente $f_j : U \to \mathbb{R}$. Zu einem genügend kleinen Vektor $h \in \mathbb{R}^m$ definiere $a_0 := a$ und $a_i := a_{i-1} + h_i e_i$. Dann gilt $f_j(a+h) - f_j(a) = \sum\limits_{i=1}^{m} (f_j(a_i) - f_j(a_{i-1}))$. Da sich a_i und a_{i-1} nur um $h_i e_i$ unterscheiden, gibt es nach dem Mittelwertsatz 2.20 ein θ_i mit

$$f_j(a_i) - f_j(a_{i-1}) = \frac{\partial f_j}{\partial x_i}(a_{i-1} + \theta_i h_i e_i) h_i \quad .$$

[3] Bei Vektoren bedeutet „genügend klein“ immer von genüngend kleiner Norm.

Also gilt

$$f_j(a+h) - f_j(a) - \sum_{i=1}^{n} \frac{\partial f_j}{\partial x_i}(a)h_i = \sum_{i=1}^{n} \left(\frac{\partial f_j}{\partial x_i}(a_{i-1} + \theta_i h_i e_i) - \frac{\partial f_j}{\partial x_i}(a) \right) h_i \leq$$
$$\|h\|_\infty \sum_{i=1}^{n} \left| \frac{\partial f_j}{\partial x_i}(a_{i-1} + \theta_i h_i e_i) - \frac{\partial f_j}{\partial x_i}(a) \right| \quad .$$

Aus $h \to 0$ folgt nun $a_{i-1} + \theta_i h_i e_i \to a$ und daher wegen der Stetigkeit der partiellen Ableitungen auch $\sum_{i=1}^{n} \left| \frac{\partial f_j}{\partial x_i}(a_{i-1} + \theta_i h_i e_i) - \frac{\partial f_j}{\partial x_i}(a) \right| \to 0$, so dass sich mit der durch $A_j h := \sum_{i=1}^{n} \frac{\partial f_j}{\partial x_i}(a)h_i$ definierten stetigen linearen Abbildung $A_j : \mathbb{R}^m \to \mathbb{R}$ die Beziehung

$$\lim_{h \to 0} \frac{f_j(a+h) - f_j(a) - A_j h}{\|h\|} = 0$$

ergibt. Somit ist jede Komponente f_j in jedem Punkt aus U differenzierbar, und wegen Lemma 2.27 ist also auch f differenzierbar auf U.

Nach Lemma 1.52 ist darüberhinaus $x \mapsto df(x)$ bereits dann stetig, wenn $x \mapsto df(x)(e_i)$ stetig ist, dies ist aber hier wegen $df(x)(e_i) = \frac{\partial f}{\partial x_i}(x)$ der Fall. Also ist f nicht nur differenzierbar, sondern sogar stetig differenzierbar. □

Dieses Korollar kann man insbesondere dazu nutzen, die (stetige) Differenzierbarkeit von Abbildungen zu erschließen, ohne explizit die Grenzwertbeziehung in Definition 2.23 überprüfen zu müssen.

Beispiel 2.35 *Die Polarkoordinaten-Abbildung* $f : \mathbb{R}^+ \times (-\pi, \pi) \to \mathbb{R}^2$, $f(r,\phi) := (r\cos(\phi), r\sin(\phi))$, *ist stetig differenzierbar, denn jede der vier Funktionen*

$$(r,\phi) \mapsto \frac{\partial f_1}{\partial r} = \cos(\phi) \quad , \quad (r,\phi) \mapsto \frac{\partial f_1}{\partial \phi} = -r\sin(\phi)$$
$$(r,\phi) \mapsto \frac{\partial f_2}{\partial r} = \sin(\phi) \quad , \quad (r,\phi) \mapsto \frac{\partial f_2}{\partial \phi} = r\cos(\phi)$$

ist stetig.

Insbesondere kann man für eine beliebige stetig differenzierbare Funktion $g : \mathbb{R}^2 \to \mathbb{R}$, $(x,y) \mapsto g(x,y)$, *die Kettenregel anwenden und daraus für* $h := g \circ f$ *die Formeln*

$$\frac{\partial h}{\partial r} = \frac{\partial g}{\partial x}\cos(\phi) + \frac{\partial g}{\partial y}\sin(\phi)$$
$$\frac{\partial h}{\partial \phi} = -\frac{\partial g}{\partial x} r\sin(\phi) + \frac{\partial g}{\partial y} r\cos(\phi)$$

gewinnen.

Ähnlich wie Satz 2.33 kann man auch den folgenden Schrankensatz beweisen.

Satz 2.36 *Unter den Voraussetzungen von Satz 2.32 gilt*

$$\|f(\tilde{x}) - f(x)\|_Y \leq \sup_{t\in[0,1]} \|\partial_{\tilde{x}-x} f(x + t(\tilde{x} - x))\|_Y \quad .$$

Ist darüberhinaus $f : U \to Y$ sogar stetig differenzierbar, so gilt

$$\|f(\tilde{x}) - f(x)\|_Y \leq \left(\max_{t\in[0,1]} \|df(x + t(\tilde{x} - x))\|_{L(X,Y)}\right) \|\tilde{x} - x\|_X \quad .$$

Beweis: Aufgrund von Satz 2.32 gilt

$$\|f(\tilde{x}) - f(x)\|_Y = \| \int_0^1 \partial_{\tilde{x}-x} f(x + t(\tilde{x} - x))\, dt\|_Y \leq$$
$$\int_0^1 \|\partial_{\tilde{x}-x} f(x + t(\tilde{x} - x))\|_Y\, dt \leq \sup_{t\in[0,1]} \|\partial_{\tilde{x}-x} f(x + t(\tilde{x} - x))\|_Y \quad .$$

Die zweite Aussage folgt dann nach Definition der Operatornorm einfach aus $\|\partial_{\tilde{x}-x} f(x+t(\tilde{x} - x))\|_Y \leq \|df(x + t(\tilde{x} - x))\|_{L(X,Y)}\|\tilde{x} - x\|_X$ und der Tatsache, dass $t \mapsto \|df(x + t(\tilde{x} - x))\|_{L(X,Y)}$ als stetige Funktion auf dem kompakten Intervall $[0, 1]$ ihr Supremum annimmt. □

Eine wichtige Konsequenz aus dem Schrankensatz ist die lokale Lipschitz-Stetigkeit stetig differenzierbarer Abbildungen.

Korollar 2.37 *Jede stetig differenzierbare Abbildung f ist lokal Lipschitz-stetig, d.h. es gibt zu jedem Punkt eine Umgebung, auf der f Lipschitz-stetig ist.*

Beweis: Sei $B_\delta(a)$ eine Kugel um a mit $\|df(x) - df(a)\|_{L(X,Y)} \leq \epsilon$ für alle $x \in B_\delta(a)$. Dann gilt nach dem Schrankensatz und der Dreiecksungleichung auch

$$\|f(\tilde{x}) - f(x)\|_Y \leq \left(\|df(a)\|_{L(X,Y)} + \epsilon\right) \|\tilde{x} - x\|_X$$

für alle $x, \tilde{x} \in B_\delta(a)$. □

Taylor-Formeln Seien X, Y Banach-Räume und $U \subset X$ offen. Ist die Ableitung $df : U \to L(X, Y)$ einer differenzierbaren Abbildung $f : U \to Y$ selbst wieder differenzierbar, so heißt f zweimal differenzierbar. Man beachte, dass dann df auch stetig ist und somit f automatisch stetig differenzierbar ist. Noch stärker ist der folgende Begriff, bei dem man auch noch die Stetigkeit der Ableitung von df verlangt.

Definition 2.38 *Seien X, Y Banach-Räume und $U \subset X$ eine offene Teilmenge. Ist die Ableitung $df : U \to L(X, Y)$ einer stetig differenzierbaren Abbildung $f : U \to Y$ selbst wieder stetig differenzierbar, so heißt f zweimal stetig differenzierbar. Induktiv definiert man so k-mal stetig differenzierbare Abbildungen, und den Vektorraum aller solcher k-mal stetig differenzierbaren Abbildungen bezeichnet man mit $C^k(U, Y)$.*

Man bemerke, dass die Ableitung von df eine Abbildung $d^2f : U \mapsto L(X, L(X,Y))$ ist, die natürlich die zweite Ableitung von f genannt wird [4]. Da man die Abbildungen $B \in L(X, L(X,Y))$ durch $(x, \tilde{x}) \mapsto B(x)(\tilde{x})$ mit bilinearen stetigen Abbildungen $X \times X \to Y$ identifizieren kann, stellt sich die Frage, ob die bilinearen Abbildungen $d^2f(x)$ symmetrisch sind. Für zweimal partiell differenzierbare Funktionen hatten wir bereits gesehen, dass $\frac{\partial^2 f}{\partial x_i \partial x_j} \neq \frac{\partial^2 f}{\partial x_j \partial x_i}$ auftreten kann. Bei zweimal stetig differenzierbaren Abbildungen ist die zweite Ableitung nach dem folgenden Satz von Schwarz aber automatisch symmetrisch.

Satz 2.39 *Seien X, Y Banach-Räume und $U \subset X$ eine offene Teilmenge. Ist $f : U \to Y$ zweimal stetig differenzierbar, so gilt $d^2f(x)(h, \tilde{h}) = d^2f(x)(\tilde{h}, h)$ für jedes $x \in U$.*

Beweis: Sei $x \in U$ und seien $h, \tilde{h}$ genügend klein. Wir schreiben den Term

$$f(x + h + \tilde{h}) - f(x + h) - f(x + \tilde{h}) + f(x)$$

einerseits mit $g(x) := f(x + h) - f(x)$ als $g(x + \tilde{h}) - g(x)$ und andererseits mit $\tilde{g}(x) := f(x + \tilde{h}) - f(x)$ als $\tilde{g}(x + h) - \tilde{g}(x)$.

Durch zweimalige Anwendung des Mittelwertsatzes 2.32 erhält man dann einerseits

$$\begin{gathered} f(x + h + \tilde{h}) - f(x + h) - f(x + \tilde{h}) + f(x) = g(x + \tilde{h}) - g(x) = \\ \int_0^1 dg(x + t\tilde{h})\tilde{h}\, dt = \int_0^1 \Big(df(x + h + t\tilde{h}) - df(x + t\tilde{h}) \Big) \tilde{h}\, dt = \\ \int_0^1 \left(\int_0^1 d^2f(x + sh + t\tilde{h})(h, \tilde{h})\, ds \right) dt \end{gathered}$$

und somit mit der Hilfsfunktion $\phi(s,t) := d^2f(x + sh + t\tilde{h})(h, \tilde{h}) - d^2f(x)(h, \tilde{h})$ die Gleichung

$$f(x + h + \tilde{h}) - f(x + h) - f(x + \tilde{h}) + f(x) = d^2f(x)(h, \tilde{h}) + \int_0^1 \left(\int_0^1 \phi(s,t)\, ds \right) dt \quad .$$

Nun gibt es aufgrund der Stetigkeit von d^2f zu jedem $\epsilon > 0$ ein $\delta > 0$, so dass für $\|h\|_X, \|\tilde{h}\|_X \leq \delta$ das Doppelintegral in der Norm auf Y durch $\frac{\epsilon}{2}\|h\|_X\|\tilde{h}\|_X$ abgeschätzt werden kann.

Andererseits erhält man mit $\tilde{g}$ ebenso

$$f(x + h + \tilde{h}) - f(x + h) - f(x + \tilde{h}) + f(x) = d^2f(x)(\tilde{h}, h) + \int_0^1 \left(\int_0^1 \psi(s,t)\, ds \right) dt \quad .$$

mit $\psi(s,t) := d^2f(x + s\tilde{h} + th)(\tilde{h}, h) - d^2f(x)(\tilde{h}, h)$ und kann wiederum dass Doppelintegral in der Norm auf Y durch $\frac{\epsilon}{2}\|h\|_X\|\tilde{h}\|_X$ abschätzen.

[4] Für diejenigen, die sich mit der Analysis auf Mannigfaltigkeiten auskennen, sei angemerkt, dass hier d^2 nichts mit dem Differential d auf Formen zu tun hat.

Daher gibt es zu jedem $\epsilon > 0$ ein $\delta > 0$, so dass für $\|h\|, \|\tilde{h}\| \leq \delta$ die Abschätzung

$$\|d^2 f(x)(h, \tilde{h}) - d^2 f(x)(\tilde{h}, h)\|_Y \leq \epsilon \|h\|_X \|\tilde{h}\|_X$$

Die bilineare und stetige Abbildung $d^2 f(x)(h, \tilde{h}) - d^2 f(x)(\tilde{h}, h)$ ist dann aber schon die Nullabbildung, und daher gilt $d^2 f(x)(h, \tilde{h}) = d^2 f(x)(\tilde{h}, h)$. □

Induktiv folgt aus dem Satz von Schwarz, dass die k-te Ableitung $d^k f(x)$ an jedem Punkt $x \in U$ eine symmetrische k-lineare Abbildung $d^k f(x) : X \times \cdots \times X \to Y$ ist. Die Abbildungen $d^k f(x)$ kann man wegen der k-Linearität als Monome vom Grad k bezeichnen. Symbolisiert man für einen Vektor $h \in X$ darüberhinaus mit h^k den Vektor $(h, \ldots, h) \in X \times \cdots \times X$, so macht auch die folgende Definition Sinn.

Definition 2.40 *Die Abbildung* $T_k f(x; a) := \sum_{l=0}^{k} \frac{1}{l!} d^l f(a)(x - a)^l$ *heißt das Taylor-Polynom k-ten Grades von f im Entwicklungspunkt a.*

Im Fall einer k-mal stetig differenzierbaren Funktion $f : U \to \mathbb{R}$ auf einer offenen Teilmenge $U \subset \mathbb{R}^n$ ist die k-te Ableitung in Richtung h^k durch

$$d^k f(x) h^k = \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_1} \ldots \partial x_{i_k}}(x) h_{i_1} \cdot \cdots \cdot h_{i_k}$$

gegeben, und analog kann man die Taylor-Polynome von f berechnen.

Beispiel 2.41 *Die auf der bzgl. der 1-Norm offenen Kugel im $\mathbb{R}^2$ beliebig oft stetig differenzierbare Funktion $f(x, y) := \frac{1}{1-x-y}$ hat im Punkt $(0, 0)$ das Taylor-Polynom k-ten Grades $T_k f((x, y); (0, 0)) = \sum_{l=0}^{k} (x + y)^l$ wegen $f = h \circ g$ mit $h(z) := \frac{1}{1-z} = \sum_{l=0}^{\infty} z^l$ und $g(x, y) := x + y$.*

Das Taylor-Polynom k-ten Grades $T_k f(x; a)$ ist das eindeutige Polynom, dessen k-te Ableitungen in a mit denen von f übereinstimmen. Es nähert f im Sinne des folgenden Satzes von Taylor an.

Satz 2.42 *Seien X, Y Banach-Räume und $U \subset X$ eine offene Teilmenge. Sei $T_k f(x; a)$ das Taylor-Polynom k-ten Grades der Abbildung $f \in C^{k+1}(U, Y)$ im Punkt $a \in U$ und sei die Strecke von a nach x komplett in U enthalten. Dann gilt*

$$f(x) - T_k f(x; a) = \int_0^1 \frac{(1-t)^k}{k!} d^{k+1} f(a + t(x - a))(x - a)^{k+1} \, dt \quad .$$

Beweis: Mit einem beliebigen stetigen linearen Funktional $\phi : Y \to \mathbb{R}$ kann man auf die Funktion $g(t) := \phi(f(a + t(x - a)))$ den eindimensionalen Satz von Taylor aus der Analysis I anwenden und erhält wegen $d^l(\phi \circ f) = \phi \circ d^l f$, der Vertauschbarkeit von ϕ

mit der Integration sowie wegen $g^{(l)}(t) = d^l(\phi \circ f)(a+t(x-a))(x-a)^l$ für alle $l \leq k+1$ die Formel

$$\phi\left(f(x) - T_k f(x;a)\right) = g(1) - T_k g(1,0) = \int_0^1 \frac{(1-t)^k}{k!} g^{(k+1)}(t)\,dt =$$
$$\phi\left(\int_0^1 \frac{(1-t)^k}{k!} d^{k+1} f(a+t(x-a))(x-a)^{k+1}\,dt\right) \quad .$$

Da diese Gleichung für alle stetigen linearen Funktionale $\phi : Y \to \mathbb{R}$ gilt (und somit bei $Y = \mathbb{R}^n$ insbesondere für die Projektionen auf die einzelnen Komponenten), folgt die Behauptung des Satzes (wobei man für unendlich-dimensionale Räume wiederum den schon im Beweis von Satz 2.13 erwähnten Satz von Hahn-Banach anwendet). □

Korollar 2.43 *Seien X, Y Banach-Räume und $U \subset X$ eine offene Teilmenge. Sei $T_k f(x;a)$ das Taylor-Polynom k-ten Grades der Abbildung $f \in C^k(U,Y)$ im Punkt $a \in U$ und sei die Strecke von a nach x komplett in U enthalten. Dann gilt*

$$\lim_{x\to a} \frac{f(x) - T_k f(x;a)}{\|x-a\|^k} = 0 \quad .$$

Beweis: Die Formel aus Satz 2.42 für $k-1$ statt k kann man zu

$$f(x) - T_{k-1} f(x;a) = \int_0^1 \frac{(1-t)^{k-1}}{(k-1)!} d^k f(a+t(x-a))(x-a)^k\,dt =$$
$$\int_0^1 \frac{(1-t)^{k-1}}{(k-1)!} \left(d^k f(a)(x-a)^k + \phi(t(x-a))(x-a)^k\right)\,dt$$

umschreiben, wobei es für $\phi(h) := d^k f(a+h) - d^k f(a)$ zu jedem $\epsilon > 0$ ein $\delta > 0$ gibt mit $\|h\| \leq \delta \Rightarrow \|\phi(h)\| \leq \epsilon$. Wegen

$$\int_0^1 \frac{(1-t)^{k-1}}{(k-1)!} d^k f(a)(x-a)^k\,dt = \frac{1}{k!} d^k f(a)(x-a)^k$$

erhält man

$$f(x) - T_k f(x;a) = \int_0^1 \frac{(1-t)^{k-1}}{(k-1)!} \phi(t(x-a))(x-a)^k\,dt \quad ,$$

und da man für $\|x-a\| \leq \delta$ die Norm des Integrals in Y durch $\frac{\epsilon}{k!}\|x-a\|^k$ abschätzen kann, folgt die Behauptung. □

Führt man als Abkürzung für $\lim\limits_{h\to 0} \frac{g(h)}{\|h\|^k} = 0$ die Schreibweise $g(h) = o(\|h\|^k)$ bei $h \to 0$ ein [5], so läßt sich die Aussage des vorigen Korollars prägnant als $f(x) - T_k f(x;a) = o(\|x-a\|^k)$ bei $x \to a$ schreiben. Mit anderen Worten fällt für $x \to a$ der Fehler, den man bei der Approximation von f durch $T_k f(x;a)$ macht, schneller ab als $\|x-a\|^k$.

Nach diesen allgemeinen Resultaten über Taylor-Polynome für Abbildungen von $U \subset X$ nach Y interessieren wir uns nun für Taylor-Polynome von Funktionen $f : U \to \mathbb{R}$. Für solche kann man das Restglied der Taylorschen Formel auch in der Lagrangeschen Form angeben.

[5] Hierbei wird $o(\|h\|^k)$ als „klein o von h hoch k" gesprochen und Landau-Symbol genannt.

Korollar 2.44 *Sei $U \subset X$ eine offene Teilmenge eines Banach-Raumes X. Sei $T_k f(x;a)$ das Taylor-Polynom k-ten Grades der Funktion $f \in C^{k+1}(U,\mathbb{R})$ im Punkt $a \in U$ und sei die Strecke von a nach x komplett in U enthalten. Dann gibt es ein $\theta \in [0,1]$ mit*

$$f(x) - T_k f(x;a) = \frac{1}{(k+1)!} d^{k+1} f(a+\theta(x-a))(x-a)^{k+1} \quad .$$

Beweis: Da das Integral aus Satz 2.42 ein $\mathbb{R}$-wertiges Integral ist, gibt es nach dem Mittelwertsatz der Integralrechnung aus der Analysis I ein $\theta \in [0,1]$ mit

$$\int_0^1 \frac{(1-t)^k}{k!} d^{k+1} f(a+t(x-a))(x-a)^{k+1}\, dt = d^{k+1} f(a+\theta(x-a))(x-a)^{k+1} \int_0^1 \frac{(1-t)^k}{k!}\, dt =$$
$$\frac{1}{(k+1)!} d^{k+1} f(a+\theta(x-a))(x-a)^{k+1} \quad ,$$

und somit folgt die Behauptung aus Satz 2.42 . □

Insbesondere gilt für zweimal stetig differenzierbare Funktionen $f : U \to \mathbb{R}$ auf einer offenen Teilmenge $U \subset \mathbb{R}^n$ wegen $df(a)h = \langle \text{grad } f(a), h \rangle$ und

$$d^2 f(a)(h,h) = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(a) h_i h_j$$

mit der Hesse-Matrix $\operatorname{Hess} f(a) := \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right)_{i,j}$ die Formel

$$T_2 f(x;a) = f(a) + \langle \text{grad } f(a), x-a \rangle + \frac{1}{2} \langle \operatorname{Hess} f(a)(x-a), (x-a) \rangle \quad .$$

Da die Funktion $h \mapsto f(a) + \langle \text{grad } f(a), h \rangle + \frac{1}{2} \langle \operatorname{Hess} f(a)h, h \rangle$ als Polynom zweiten Grades eine Quadrik als Graphen besitzt und sich in a an den Graphen von f anschmiegt, bezeichnet man den Graphen von $T_2 f(x;a)$ auch als Schmiegquadrik an f im Punkt a.

Hinreichendes Kriterium für lokale Extrema Mittels des Taylor-Polynoms zweiten Grades einer zweimal stetig differenzierbaren Funktion $f : U \to \mathbb{R}$ auf einer offenen Teilmenge $U \subset \mathbb{R}^n$ kann man insbesondere hinreichende Kriterien für das Vorliegen eines lokalen Extremums in einem Punkt $a \in U$ gewinnen. Dazu sei an Satz 2.22 erinnert, nach dem $\text{grad } f(a) = 0$ gilt, falls in a ein lokales Extremum von f liegt.

Bei Verschwinden von grad f in a gilt aber $T_2 f(x;a) = f(a) + \frac{1}{2} \langle \operatorname{Hess} f(a)(x-a), (x-a) \rangle$, und somit wird das Verhalten der Funktion f nahe a in zweiter Näherung durch die Hesse-Matrix $\operatorname{Hess} f$ bestimmt. Man beachte, dass $\operatorname{Hess} f$ nach dem Satz von Schwarz 2.39 symmetrisch ist. Mittels der folgenden Begriffe können wir die verschiedenen auftretenden Fälle unterscheiden.

Definition 2.45 *Eine symmetrische $n \times n$-Matrix A oder die ihr zugeordnete quadratische Form $h \mapsto \langle Ah, h \rangle$ heißt*

- *positiv definit, falls* $\langle Ah, h\rangle > 0$ *für alle* $h \neq 0$ *gilt.*
- *negativ definit, falls* $\langle Ah, h\rangle < 0$ *für alle* $h \neq 0$ *gilt.*
- *indefinit, falls es sowohl Vektoren* h *mit* $\langle Ah, h\rangle > 0$ *als auch Vektoren* $\tilde{h}$ *mit* $\langle A\tilde{h}, \tilde{h}\rangle < 0$ *gibt.*
- *positiv bzw. negativ semidefinit, falls* $\langle Ah, h\rangle \geq 0$ *bzw.* $\langle Ah, h\rangle \leq 0$ *für alle* h *gilt.*

Definitheit kann man durch die Eigenwerte von A charakterisieren. Dazu sei daran erinnert, dass eine Zahl λ Eigenwert der $n \times n$-Matrix A heißt, wenn es einen Vektor $h \in \mathbb{R}^n$ gibt mit $Ah = \lambda h$. Die Eigenwerte von A sind die Nullstellen des charakteristischen Polynoms $\lambda \mapsto \det(A - \lambda \,\mathrm{Id})$, und für symmetrische Matrizen A ist jeder Eigenwert reell.

Lemma 2.46 *Für eine symmetrische* $n \times n$*-Matrix* A *gilt:*

- A *ist positiv definit* $\Leftrightarrow$ *jeder Eigenwert von* A *ist positiv*
- A *ist negativ definit* $\Leftrightarrow$ *jeder Eigenwert von* A *ist negativ*
- A *ist indefinit* $\Leftrightarrow$ *es gibt sowohl positive als auch negative Eigenwerte von* A
- A *ist positiv bzw. negativ semidefinit* $\Leftrightarrow$ *jeder Eigenwert von* A *ist nicht-negativ bzw. nicht-positiv*

Beweis: Für symmetrische Matrizen gibt es eine Basis des $\mathbb{R}^n$ aus orthonormalen Eigenvektoren und daher eine orthogonale $n \times n$-Matrix S mit $A = SDS^T$, wobei $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ eine Diagonalmatrix ist, auf deren Diagonaler die Eigenwerte λ_i von A stehen (Hauptachsentransformation). Mit $\tilde{h} = S^T h$ gilt daher

$$\langle Ah, h\rangle = \langle DS^T h, S^T h\rangle = \langle D\tilde{h}, \tilde{h}\rangle = \sum_{i=1}^{n} \lambda_i \tilde{h}_i^2$$

und daraus ergeben sich sofort alle Behauptungen. □

Satz 2.47 *Ist* $f : U \to \mathbb{R}$ *zweimal stetig differenzierbar auf einer offenen Teilmenge* $U \subset \mathbb{R}^n$ *und ist* $a \in U$ *ein Punkt, für den* $\mathrm{grad}\, f(a) = 0$ *gilt und* $\mathrm{Hess}\, f(a)$ *positiv bzw. negativ definit ist, dann liegt in* a *ein lokales Minimum bzw. Maximum.*

Beweis: Ist $\mathrm{Hess}\, f(a)$ positiv definit, dann hat die stetige Funktion $h \mapsto \langle \mathrm{Hess}\, f(a)h, h\rangle$ auf der kompakten Sphäre $S^{n-1} := \{h \in \mathbb{R}^n \mid \|h\|_2 = 1\}$ ein Minimum $m > 0$ wegen $\langle \mathrm{Hess}\, f(a)h, h\rangle > 0$ für $h \in S^{n-1}$. Also gilt $\langle \mathrm{Hess}\, f(a)h, h\rangle \geq m\|h\|_2^2$ für alle $h \in \mathbb{R}^n$.

Andererseits gibt es für den durch $f(a+h) = f(a) + \frac{1}{2}\langle \mathrm{Hess}\, f(a)h, h\rangle + R(h)$ definierten Rest $R(h)$ wegen $\lim\limits_{h \to 0} \frac{R(h)}{\|h\|^2} = 0$ zu $m > 0$ ein $\delta > 0$ mit $|R(h)| \leq \frac{m}{4}\|h\|^2$ für alle h mit $\|h\| \leq \delta$. Insbesondere gilt $R(h) \geq -\frac{m}{4}\|h\|^2$ und daher

$$f(a+h) = f(a) + \frac{1}{2}\langle \mathrm{Hess}\, f(a)h, h\rangle + R(h) \geq f(a) + \frac{m}{2}\|h\|^2 - \frac{m}{4}\|h\|^2 = f(a) + \frac{m}{4}\|h\|^2 \quad ,$$

für alle h mit $\|h\| \leq \delta$. Daher nimmt f auf $B_\delta(a)$ im Punkt a sein eindeutiges Minimum an. $\square$

Beispiel 2.48 *Die Funktion* $f(x, y) := x^2 + xy + y^2$ *hat den Gradienten* $\operatorname{grad} f(x, y) = (2x + y, x + 2y)^T$, *und dieser verschwindet nur im Punkt* $(0, 0)$. *Die Hesse-Matrix* $\operatorname{Hess} f(x, y) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ *hat (im Punkt* $(0, 0)$*) das charakteristische Polynom* $(\lambda - 2)^2 - 1$ *und daher die Eigenwerte* $\lambda = 1$ *und* $\lambda = 3$. *Also liegt in* $(0, 0)$ *ein lokales Minimum von* f.

Beispiel 2.49 *Sei* $m > n$, $A \in \mathbb{R}^{m \times n}$ *eine Matrix vollen Ranges und* $b \in \mathbb{R}^m$. *Dann löst der Vektor* $x \in \mathbb{R}^n$ *das lineare Ausgleichsproblem* $\|Ax - b\|_2 \stackrel{!}{=} \min$ *genau dann, wenn er die Normalengleichungen* $A^T A x = A^T b$ *löst.*

Denn die Funktion $f(x) := \|Ax - b\|_2^2$ *hat wegen* $f(x) = \langle Ax - b, Ax - b\rangle$ *die Ableitung* $df(x)h = 2\langle Ax - b, Ah\rangle$, *und daher ist* $\operatorname{grad} f(x) = 0$ *äquivalent zu* $A^T(Ax - b) = 0$. *Darüberhinaus ist die zweite Ableitung* $\operatorname{Hess} f(x) = A^T A$ *positiv definit, da* A *vollen Rang hat, also ist die Lösung von* $A^T A x = A^T b$ *auch wirklich ein Minimum.*

Differentiation parameterabhängiger Integrale In Satz 1.66 haben wir die Stetigkeit von parameterabhängigen Integralen $F(x) := \int_a^b f(x, t)\, dt$ gezeigt. Nun wollen wir diskutieren, unter welchen Umständen die Funktion F stetig differenzierbar ist.

Satz 2.50 *Sei* $U \subset \mathbb{R}^n$ *eine offene Teilmenge und* $f : U \times [a, b] \to \mathbb{R}$ *eine stetige Funktion* $f(x, t)$, *deren partielle Ableitungen* $\frac{\partial f}{\partial x_i}(x, t)$ *nach den Variablen* x_i, $i = 1, \ldots, n$, *als Funktionen auf* $U \times [a, b]$ *stetig sind. Dann ist die Funktion* $F : U \to \mathbb{R}$, $F(x) := \int_a^b f(x, t)\, dt$ *stetig differenzierbar mit partiellen Ableitungen* $\dfrac{\partial F}{\partial x_i}(x) := \displaystyle\int_a^b \frac{\partial f}{\partial x_i}(x, t)\, dt$.

Beweis: Sei $x^* \in U$ fest gewählt. Wir müssen zeigen, dass der Kandidat $\int_a^b \frac{\partial f}{\partial x_i}(x^*, t)\, dt$ für die partielle Ableitung $\frac{\partial F}{\partial x_i}(x^*)$ wirklich die partielle Ableitung von F nach x_i im Punkt x^* ist, dann folgt wegen der Stetigkeit von $\frac{\partial f}{\partial x_i}(x, t)$ durch Anwendung von Satz 1.66 auch die stetige Differenzierbarkeit von F.

Dazu betrachte zu vorgegebenem $\epsilon > 0$ die Differenz $(x, t) \mapsto \frac{\partial f}{\partial x_i}(x, t) - \frac{\partial f}{\partial x_i}(x^*, t)$, die nach Voraussetzung stetig auf $U \times [a, b]$ ist und auf der Faser $\{x^*\} \times [a, b]$ verschwindet. Also ist

$$W := \{(x, t) \in U \times [a, b] \mid |\frac{\partial f}{\partial x_i}(x, t) - \frac{\partial f}{\partial x_i}(x^*, t)| < \epsilon\}$$

eine offene Umgebung von $\{x^*\} \times [a, b]$ in $U \times [a, b]$, und nach dem Tubenlemma 1.64 gibt es somit eine offene Umgebung U' von x^* mit $U' \times [a, b] \subset W$. Bezeichne e_i die Standardbasis des $\mathbb{R}^n$ und sei s so klein, dass $x^* + se_i \in U'$ gilt, dann gibt es nach dem Mittelwertsatz zu jedem $t \in [a, b]$ eine Stelle $s(t)$ zwischen 0 und s mit

$$\frac{F(x^* + se_i) - F(x^*)}{s} = \int_a^b \frac{f(x^* + se_i, t) - f(x^*, t)}{s}\, dt = \int_a^b \frac{\partial f}{\partial x_i}(x^* + s(t)e_i, t)\, dt$$

Aufgrund der Definition von U' und wegen $x^* + se_i \in U'$ gilt daher

$$\left| \frac{F(x^* + se_i) - F(x^*)}{s} - \int_a^b \frac{\partial f}{\partial x_i}(x^*, t)\, dt \right| \leq \epsilon(b-a)$$

für s so klein, dass $x^* + se_i \in U'$ gilt. Also ist $\int\limits_a^b \frac{\partial f}{\partial x_i}(x^*, t)\, dt$ wirklich die partielle Ableitung von F nach x_i im Punkt x^*. □

Euler-Lagrange-Gleichungen Eine große Rolle innerhalb der Mathematik und den sie anwendenden Wissenschaften spielt auch das Auffinden stationärer Punkte von Funktionen auf unendlich-dimensionalen Banach-Räumen, sogenannten Funktionalen. Wir wollen dies am Beispiel der Euler-Lagrange-Gleichungen verdeutlichen.

Als Banach-Raum betrachten wir die Menge $C^1([a,b],\mathbb{R})$ der stetig differenzierbaren Funktionen $u : [a,b] \to \mathbb{R}$ mit der Norm

$$\|u\| := \sup_{x\in[a,b]} |u(x)| + \sup_{x\in[a,b]} |u'(x)| = \|u\|_\infty + \|u'\|_\infty \quad ,$$

wobei der Leser selbst nachvollziehen möge, dass $C^1([a,b],\mathbb{R})$ mit dieser Norm ein Banach-Raum ist.

Sei nun eine stetig differenzierbare Funktion $L : \mathbb{R}^2 \to \mathbb{R}$ auf dem $\mathbb{R}^2$ vorgegeben, die in diesem Kontext Lagrange-Funktion genannt wird. Dann kann man mit ihrer Hilfe das Funktional $I : C^1([a,b],\mathbb{R}) \to \mathbb{R}$, $I(u) := \int_a^b L(u(t), u'(t))\, dt$, bilden, das einer stetig differenzierbaren Funktion $u : [a,b] \to \mathbb{R}$ den Wert des Integrals $\int_a^b L(u(t), u'(t))\, dt$ zuordnet.

Das Funktional I ist partiell differenzierbar, denn für eine Funktion $h \in C^1([a,b],\mathbb{R})$ ist die Funktion $(s,t) \mapsto L(u(t) + sh(t), u'(t) + sh'(t))$ stetig sowie stetig partiell differenzierbar nach s, also folgt aus Satz 2.50

$$\partial_h I(u) = \frac{\partial}{\partial s} I(u+sh)|_{s=0} = \frac{\partial}{\partial s}\left(\int_a^b L(u(t)+sh(t), u'(t)+sh'(t))\, dt\right)|_{s=0} =$$
$$\int_a^b \frac{\partial}{\partial s} L(u(t)+sh(t), u'(t)+sh'(t))|_{s=0}\, dt$$

Wir suchen nun stationäre Punkte des Funktionals I, d.h. Funktionen $u \in C^1([a,b],\mathbb{R})$ mit $\partial_h I(u) = 0$ für alle $h \in C^1([a,b],\mathbb{R})$ mit $h(a) = 0 = h(b)$. Mit der Bezeichnung $L = L(q,v)$ gilt nach der Kettenregel

$$\frac{\partial}{\partial s} L(u(t)+sh(t), u'(t)+sh'(t))|_{s=0} = \frac{\partial L}{\partial q}(u(t), u'(t)) \cdot h(t) + \frac{\partial L}{\partial v}(u(t), u'(t)) \cdot h'(t) \quad ,$$

und somit folgt mittels partieller Integration

$$\partial_h I(u) = \int_a^b \left(\frac{\partial L}{\partial q}(u(t), u'(t)) - \frac{d}{dt}\frac{\partial L}{\partial v}(u(t), u'(t))\right) \cdot h(t)\, dt$$

unter der Annahme, dass $t \mapsto \frac{\partial L}{\partial v}(u(t), u'(t))$ stetig differenzierbar ist. Daraus ergibt sich der folgende Satz.

Satz 2.51 *Ist $L : \mathbb{R}^2 \to \mathbb{R}$ stetig differenzierbar und $u \in C^1([a,b],\mathbb{R})$ so, dass auch $t \mapsto \frac{\partial L}{\partial v}(u(t),u'(t))$ stetig differenzierbar ist, dann gilt die Euler-Lagrange-Gleichung*

$$\frac{d}{dt}\frac{\partial L}{\partial v}(u(t),u'(t)) = \frac{\partial L}{\partial q}(u(t),u'(t))$$

auf (a,b) genau dann, wenn u ein stationärer Punkt von $I(u) := \int_a^b L(u(t),u'(t))\,dt$ ist.

Beweis: Wegen der Formel für $\partial_h I(u)$ haben wir nur noch zu zeigen, dass für eine stetige Funktion $f : (a,b) \to \mathbb{R}$ aus $\int_a^b f(t)\cdot h(t)\,dt = 0$ für alle $h \in C^1([a,b],\mathbb{R})$ die Gleichung $f \equiv 0$ auf (a,b) folgt.

Angenommen, es gibt einen Punkt $t^* \in (a,b)$ mit $f(t^*) \neq 0$. Gilt $f(t^*) > 0$, dann gibt es aufgrund der Stetigkeit von f ein Intervall $[a^*,b^*] \subset (a,b)$, das den Punkt t^* enthält und auf dem $f(t) \geq \frac{1}{2}f(t^*)$ für alle $t \in [a^*,b^*]$ gilt. Sei nun $h \in C^1([a,b],\mathbb{R})$ eine Funktion, die auf (a^*,b^*) positiv ist und die außerhalb von (a^*,b^*) verschwindet (der Leser möge solch eine Funktion selbst konstruieren). Dann gilt

$$\int_a^b f(t)\cdot h(t)\,dt = \int_{a^*}^{b^*} f(t)\cdot h(t)\,dt \geq \frac{1}{2}f(t^*)\int_{a^*}^{b^*} h(t)\,dt > 0$$

im Widerspruch zu $\int_a^b f(t)\cdot h(t)\,dt = 0$ für alle $h \in C^1([a,b],\mathbb{R})$. Den Fall $f(t^*) < 0$ handhabt man völlig analog. □

Beispiel 2.52 *Sei $L(q,v) := \frac{m}{2}v^2 - \frac{m\alpha}{2}q^2$ mit Konstanten $m,\alpha > 0$, dann lautet die Euler-Lagrange-Gleichung*

$$\frac{d}{dt}(mu') = -m\alpha u$$

und ist mit der Substitution $v := u'$ äquivalent zu den linearen Differentialgleichungen

$$u' = v \quad , \quad v' = \alpha u \quad .$$

Diese Differentialgleichungen beschreiben lineare Schwingungen (vergleiche Beispiel 2.5), und mittels der Exponentialfunktion kann man die Lösung zum Anfangswert (u_0,v_0) als $\begin{pmatrix} u \\ v \end{pmatrix} = \exp\left(t\begin{pmatrix} 0 & 1 \\ \alpha & 0 \end{pmatrix}\right)\begin{pmatrix} u_0 \\ v_0 \end{pmatrix}$ schreiben.

2.5 Diffeomorphismen

In diesem Abschnitt wollen wir uns mit k-mal stetig differenzierbaren Abbildungen beschäftigen, die bijektiv sind und deren Umkehrabbildung auch wieder k-mal stetig differenzierbar ist, sogenannten C^k-Diffeomorphismen.

Definition 2.53 *Seien X,Y Banach-Räume. Eine Abbildung $f : U \to V$ nennt man einen C^k-Diffeomorphismus von der offenen Teilmenge $U \subset X$ auf die offene Teilmenge $V \subset Y$, wenn f eine bijektive C^k-Abbildung ist, deren Umkehrabbildung f^{-1} wiederum eine C^k-Abbildung ist.*

Diffeomorphismen sind also Abbildungen, die nicht nur die topologische Struktur erhalten (sowohl f als auch f^{-1} sind stetig), sondern die auch noch die differenzierbare Struktur erhalten (sowohl f als auch f^{-1} sind k-mal stetig differenzierbar). Will man nur lokal über Diffeomorphie sprechen, so ist die folgende Definition hilfreich.

Definition 2.54 *Seien X, Y Banach-Räume und sei $U \subset X$ offen. Eine Abbildung $f : U \to Y$ heißt lokal C^k-invertierbar bei $x \in U$, wenn es offene Umgebungen $U' \subset U$ von x und $V' \subset Y$ von $f(x)$ gibt, so dass $f : U' \to V'$ ein C^k-Diffeomorphismus ist.*

Während man nur schwer beweisen kann, dass $\mathbb{R}^m$ und $\mathbb{R}^n$ für $m \neq n$ nicht homöomorph sind, kann man dagegen leicht beweisen, dass offene Teilmengen endlich-dimensionaler Räume nur dann diffeomorph sein können, wenn die zugrundeliegenden Räume dieselbe Dimension besitzen.

Lemma 2.55 *Seien X, Y Banach-Räume und sei $U \subset X$ offen. Ist $f : U \to Y$ lokal C^k-invertierbar bei $x \in U$, so ist die stetige lineare Abbildung $df(x) \in L(X, Y)$ invertierbar mit Inversem $df^{-1}(f(x)) \in L(Y, X)$. Insbesondere gilt bei $X = \mathbb{R}^m$ und $Y = \mathbb{R}^n$ auch $m = n$.*

Beweis: Ist $f : U' \to V'$ ein C^k-Diffeomorphismus, so gelten wegen $f^{-1} \circ f = \mathrm{Id}_{U'}$ und $f \circ f^{-1} = \mathrm{Id}_{V'}$ nach der Kettenregel die Gleichungen $df^{-1}(f(x)) \cdot df(x) = \mathrm{Id}_X$ und $df(x) \cdot df^{-1}(f(x)) = \mathrm{Id}_Y$, also hat die stetige lineare Abbildung $df(x) \in L(X, Y)$ das Inverse $df^{-1}(f(x)) \in L(Y, X)$. Insbesondere muß bei $X = \mathbb{R}^m$ und $Y = \mathbb{R}^n$ auch $m = n$ gelten, denn lineare Abbildungen zwischen endlich-dimensionalen Räumen können nur in diesem Fall invertierbar sein. $\square$

Im folgenden beweisen wir den Satz über lokale Umkehrbarkeit, der besagt, dass auch die Umkehrung des vorigen Lemmas gilt. Da man relativ leicht nachprüfen kann, ob eine stetige lineare Abbildung ein stetiges lineares Inverses besitzt, ist dieser Satz ein sehr mächtiges Werkzeug der nichtlinearen Analysis. Bevor wir mit dem Beweis beginnen, müssen wir uns in Verallgemeinerung von Satz 1.55 aber noch klar machen, dass die Inversion auf $GL(X)$ beliebig oft differenzierbar ist.

Lemma 2.56 *Sei X ein Banach-Raum, dann ist die Inversion $A \mapsto A^{-1}$ auf $GL(X)$ beliebig oft stetig differenzierbar.*

Beweis: Aus dem Beweis von Satz 1.55 erinnere man sich daran, dass die Inversion lokal eine Verkettung der Neumannschen Reihe $f(A) := \sum_{k=0}^{\infty} A^k$ und der Abbildungen $B \mapsto \mathrm{Id} - A^{-1}B$ sowie $C \mapsto CA^{-1}$ war. Letztere sind offensichtlich beliebig oft stetig differenzierbar, und auch die Neumannsche Reihe ist innerhalb ihres Konvergenzkreises beliebig of stetig differenzierbar (was man für Potenzreihen auf der Banach-Algebra $L(X, X)$ genauso zeigen kann wie für reelle Potenzreihen). Also folgt die Behauptung aus der Kettenregel. $\square$

Nach dieser Vorbemerkung kommen wir nun zum Satz über lokale Umkehrbarkeit.

Satz 2.57 *Seien X, Y Banach-Räume und $U \subset X$ offen. Eine C^k-Abbildung $f : U \to Y$ ist genau dann lokal C^k-invertierbar bei $x \in U$, wenn $df(x) \in L(X,Y)$ invertierbar ist.*

Beweis: Die eine Richtung zeigt das vorige Lemma, die andere schauen wir uns jetzt an: OE dürfen wir $X = Y$, $x = 0$, $f(x) = 0$ und $df(x) = \mathrm{Id}_X$ annehmen (betrachte einfach $h \mapsto df(x)^{-1}(f(x+h) - f(x))$ statt f).

Wir wollen f nahe Null invertieren, also die Gleichung $f(x) = y$ lösen. Diese schreiben wir mittels $h_y(x) := x + y - f(x)$ um in die Fixpunktgleichung $h_y(x) = x$. Fixpunkte x von h_y sind dann also Lösungen von $f(x) = y$.

Wegen $dh_y(0) = \mathrm{Id} - \mathrm{Id} = 0$ und der stetigen Differenzierbarkeit finden wir eine Kugel $\overline{B_{2r}(0)}$ im Definitionsbereich von f mit $\|dh_y(x)\| \leq 1/2$ für alle $x \in \overline{B_{2r}(0)}$. Somit ist h_y auf $\overline{B_{2r}(0)}$ nach dem Schrankensatz Lipschitz-stetig mit Lipschitz-Konstanter $\leq 1/2$.

Insbesondere gilt für $\|y\| < r$ und $\|x\| \leq 2r$ wegen $h_y(0) = y$ dann auch

$$\|h_y(x)\| \leq \|h_y(x) - h_y(0)\| + \|y\| < \frac{1}{2}2r + r = 2r \quad .$$

Also ist h_y für jedes $y \in B_r(0)$ eine kontrahierende Selbstabbildung von $\overline{B_{2r}(0)}$.

Nach dem Fixpunktsatz gibt es daher zu jedem Punkt $y \in B_r(0)$ genau einen Fixpunkt $x \in \overline{B_{2r}(0)}$ von h_y, und da sogar die echte Ungleichung $\|h_y(x)\| < 2r$ gilt, liegt der Fixpunkt sogar in der offenen Kugel $B_{2r}(0)$. Setze $g(y) := x$, dann ist die Abbildung $f : U' \to B_r(0) =: V'$ mit Definitionsbereich $U' := f^{-1}(B_r(0)) \cap B_{2r}(0)$ und Bild V' also bijektiv mit Inversem g, und insbesondere ist U' wegen der Stetigkeit von f offen.

Wir müssen nur noch zeigen, dass g eine C^k-Abbildung ist. Zunächst einmal zeigen wir dazu, dass g stetig ist: Sei $y, y' \in B_r(0)$, dann gilt für $x := g(y)$, $x' := g(y')$ natürlich $x - x' = h_0(x) - h_0(x') + f(x) - f(x')$, und da h_0 Lipschitz-stetig mit Lipschitz-Konstante $\leq 1/2$ ist, folgt $\|x - x'\| \leq \|x - x'\|/2 + \|f(x) - f(x')\|$. Wegen $f(x) = y$, $f(x) = y'$ gilt also $\|g(y) - g(y')\| \leq 2\|y - y'\|$, d.h. g ist Lipschitz-stetig.

Desweiteren ist g sogar C^k. Dazu ein paar Vorbemerkungen: Die Ableitung df ist in jedem Punkt $x \in \overline{U'}$ invertierbar, da nach Wahl von r ja $\|\mathrm{Id} - df(x)\| = \|dh_y(x)\| \leq 1/2$ gilt, also $df(x)$ nach dem ersten Teil des Beweises von Satz 1.55 (Neumannsche Reihe) invertierbar ist. Darüberhinaus ist wegen der Stetigkeit von df und der Stetigkeit der Inversion auf $GL(X)$ die Operatornorm von df^{-1} auf $\overline{U'}$ beschränkt, d.h. es gibt ein $C < \infty$ mit $\|df(x)^{-1}\| \leq C$.

Um zu zeigen, dass g eine C^k-Abbildung ist, rechnen wir nun einfach nach, dass der Kandidat $df(g(y))^{-1}$ für die Ableitung $dg(y)$ wirklich diese Ableitung ist: Sei $g(y) = x$, $g(y') = x'$, dann gilt

$$\begin{aligned}\|g(y') - g(y) - df(g(y))^{-1}(y' - y)\| = \|x' - x - df(x)^{-1}(f(x') - f(x))\| \leq \\ C\|f(x') - f(x) - df(x)(x' - x)\| \quad ,\end{aligned}$$

also ist mit f auch g differenzierbar. Aber g ist sogar C^k, denn aus der Gleichheit $dg(y) = df(g(y))^{-1}$, aus der Glattheit der Inversion auf $GL(X)$ und da df eine C^{k-1}-

Abbildung ist, folgt induktiv, dass auch dg eine C^{k-1}-Abbildung ist [6]. □

Bevor wir Beispiele für die Anwendbarkeit des Satzes über lokale Umkehrbarkeit diskutieren, beweisen wir noch kurz eine wichtige Folgerung.

Korollar 2.58 *Seien X, Y Banach-Räume, sei $U \subset X$ offen, und sei $f : U \to Y$ stetig differenzierbar mit der Eigenschaft, dass die Ableitung $df(x)$ in jedem Punkt $x \in U$ invertierbar ist. Dann ist $f(U)$ offen, und falls f zusätzlich injektiv ist, dann ist f sogar ein Diffeomorphismus von U auf $f(U)$.*

Beweis: Nach dem Satz über lokale Invertierbarkeit findet man zu jedem Punkt $x \in U$ eine Umgebung $U(x) \subset U$, auf der f lokal C^1-invertierbar ist. Da insbesondere die inverse Abbildung $g := f^{-1}$ stetig ist, ist $g^{-1}(U(x)) = f(U(x))$ offen, und aus der Gleichung $f(U) = \bigcup_{x \in U} f(U(x))$ folgt die Offenheit von $f(U)$. Bei zusätzlicher Injektivität von f ist die Umkehrabbildung $g : f(U) \to U$ sogar global auf $f(U)$ definiert und als lokal stetig differenzierbare Abbildung auch global stetig differenzierbar. □

Beispiel 2.59 *Die Polarkoordinaten-Abbildung $P_n : \mathbb{R}^n \to \mathbb{R}^n$, die für $n \geq 2$ rekursiv durch*

$$P_{n+1}(r, \phi_1, \dots, \phi_n) := \begin{pmatrix} P_n(r, \phi_1, \dots, \phi_{n-1}) \cos(\phi_n) \\ r \sin(\phi_n) \end{pmatrix}$$

und $P_2(r, \phi) := \begin{pmatrix} r\cos(\phi) \\ r\sin(\phi) \end{pmatrix}$ *definiert ist, ist an jedem Punkt $(r, \phi_1, \dots, \phi_{n-1})$ mit $r > 0$ und $(\phi_2, \dots, \phi_{n-1}) \in (-\pi/2, \pi/2)^{n-2}$ ein lokaler C^k-Diffeomorphismus, wobei k beliebig groß ist.*

Denn für die Jacobi-Matrizen gilt $JP_2(r, \phi) = \begin{pmatrix} \cos(\phi) & -r\sin(\phi) \\ \sin(\phi) & r\cos(\phi) \end{pmatrix}$ *sowie*

$$JP_{n+1}(r, \phi_1, \dots, \phi_n) = \begin{pmatrix} JP_n(r, \phi_1, \dots, \phi_{n-1}) \cos(\phi_n) & -P_n(r, \phi_1, \dots, \phi_{n-1}) \sin(\phi_n) \\ \sin(\phi_n), 0, \dots, 0 & r\cos(\phi_n) \end{pmatrix} ,$$

also gilt $\det(JP_2) = r > 0$ *und durch Entwicklung nach der letzten Zeile erhält man wegen* $r \frac{\partial P_n}{\partial r} = P_n$ *die Rekursionsformel*

$$\det(JP_{n+1}) = -(-1)^n (-1)^{n-1} r \sin^2(\phi_n) \cos^{n-1}(\phi_n) \det(JP_n) + r \cos^{n+1}(\phi_n) \det(JP_n) = r \cos^{n-1}(\phi_n) \det(JP_n)$$

und damit $\det(JP_n) = r^{n-1} \cos(\phi_2) \dots \cos^{n-2}(\phi_{n-1})$*. Also ist die Matrix JP_n in jedem Punkt mit $r > 0$ und $(\phi_2, \dots, \phi_{n-1}) \in (-\pi/2, \pi/2)^{n-2}$ invertierbar.*

Tatsächlich ist $P_n : \mathbb{R}^+ \times (-\pi, \pi) \times (-\pi/2, \pi/2)^{n-2} \to \mathbb{R}^n \setminus \{y \in \mathbb{R}^n \,|\, y_1 \leq 0, y_2 = 0\}$ ein globaler C^k-Diffeomorphismus, wie man durch die explizite Angabe einer Umkehrabbildung beweisen kann.

---

[6] Behauptung: Aus $df \in C^{k-1}$ folgt $dg \in C^{k-1}$.
Induktionsanfang: Ist df stetig, so wegen $dg(y) = df(g(y))^{-1}$ auch dg.
Induktionsschluß: Ist $df \in C^{k-1}$, so ist $df \in C^{k-2}$ und nach Induktionsvoraussetzung daher auch $dg \in C^{k-2}$, also $g \in C^{k-1}$. Aus der Gleichung $dg(y) = df(g(y))^{-1}$ folgt daher nach der Kettenregel, dass auch $dg \in C^{k-1}$ ist.

Beispiel 2.60 *Die Abbildung* $\exp : L(X,X) \to L(X,X)$ *hat im Punkt* 0 *den Wert* $\exp(0) = \mathrm{Id}_X$ *und die Ableitung* $d\exp(0) = \mathrm{Id}_{L(X,X)}$. *Somit gibt es eine offene Umgebung* $U \subset L(X,X)$ *der Nullabbildung und eine offene Umgebung* $V \subset L(X,X)$ *der Identität* Id_X, *so dass* $\exp : U \to V$ *ein* C^k*-Diffeomorphismus ist.*

Insbesondere kann man jede genügend nah an der Einheitsmatrix E_n *gelegene Matrix* B *als Bild* $B = \exp(A)$ *mit einer* $n \times n$*-Matrix* A *darstellen.*

2.6 Implizit definierte Abbildungen

Will man Gleichungen wie beispielsweise die Kreisgleichung $f(x,y) := x^2 + y^2 = r^2$ für festes $r > 0$ lokal in der Nähe einer schon gefundenen Lösung (x^*, y^*) nach einer der Variablen auflösen, so können unterschiedliche Fälle auftreten. Denn nahe des Punktes $(0, r)$ kann man die Kreisgleichung nur nach y auflösen (durch $y = \sqrt{r^2 - x^2}$), aber nicht nach x, da die Gleichung $x^2 + y^2 = r^2$ für $y > r$ keine Lösung x besitzt, während sie für $y < r$ immer schon zwei Lösungen x hat. Es gibt also keine auf einer Umgebung von r stetige Funktion $g(y)$ mit $g(r) = 0$ und $g(y)^2 + y^2 = r^2$ für alle y aus der Umgebung von r. Ebenso verhält es sich im Punkt $(0, -r)$.

Umgekehrt kann man die Kreisgleichung nahe der Punkte $(\pm r, 0)$ nur nach x auflösen (durch $x = \pm\sqrt{r^2 - y^2}$), aber nicht nach y. In der Nähe aller anderen Punkte (x^*, y^*) des Kreises kann man sie dagegen nach jeder der beiden Variablen auflösen (aus $x^2 + y^2 = r^2 = (x^*)^2 + (y^*)^2$ ergibt sich $x = \pm\sqrt{(x^*)^2 + (y^*)^2 - y^2}$ und $y = \pm\sqrt{(x^*)^2 + (y^*)^2 - x^2}$). Die so durch $f(x, g(x)) = r^2$ bzw. $f(g(y), y) = r^2$ definierten Funktionen $g(x)$ bzw. $g(y)$ nennt man implizit definiert.

Es kann aber sogar der Fall auftreten, dass man eine Gleichung $f(x,y) = 0$ nahe einer Lösung (x^*, y^*) nach keiner der beiden Variablen auflösen kann. Beispielsweise gilt für die Gleichung $f(x,y) := xy = 0$, dass es nahe der Lösung $(0,0)$ keine Funktionen $g(x)$ bzw. $g(y)$ mit $f(x, g(x)) = 0$ bzw. $f(g(y), y) = 0$ gibt.

Der folgende Satz über implizit definierte Abbildungen (oder kurz implizite Funktionen) gibt ein sehr nützliches und leicht überprüfbares Kriterium dafür an, wann man eine Gleichung $f(x,y) = z^*$ in der Nähe einer Lösung (x^*, y^*) nach y auflösen kann. Er ist damit ein sehr mächtiges Werkzeug, um Aussagen über Lösungen von Gleichungen zu gewinnen, insbesondere in dem Fall, dass man die Gleichungen analytisch nicht explizit lösen kann.

Satz 2.61 *Seien* X, Y, Z *Banach-Räume, sei* $U \subset X \times Y$ *offen und sei* $f : U \to Z$ *eine* C^k*-Abbildung. Löst zu vorgegebenem* $z^* \in Z$ *der Punkt* $(x^*, y^*) \in U$ *die Gleichung* $f(x^*, y^*) = z^*$ *und ist* $df(x^*, y^*)|_Y =: d_Y f(x^*, y^*) \in L(Y, Z)$ *invertierbar, dann gibt es offene Umgebungen* U^* *von* x^* *und* V^* *von* y^* *sowie eine* C^k*-Abbildung* $g : U^* \to V^*$ *mit* $f(x,y) = z^* \Leftrightarrow y = g(x)$ *für* $(x,y) \in U^* \times V^*$.

Beweis: Die Abbildung $\phi : U \to X \times Z$, $\phi(x,y) := (x, f(x,y))$ mit $\phi(x^*, y^*) = (x^*, z^*)$ ist wegen der Invertierbarkeit von $d\phi(x^*, y^*) = \begin{pmatrix} \mathrm{Id}_X & 0 \\ df(x^*, y^*)|_X & df(x^*, y^*)|_Y \end{pmatrix}$ lokal bei

(x^*, y^*) ein C^k-Diffeomorphismus. Also gibt es offene Umgebungen $U' \subset U \subset X \times Y$ von (x^*, y^*) und $V' \subset X \times Z$ von (x^*, z^*) derart, dass $\phi : U' \to V'$ ein C^k-Diffeomorphismus ist. Wegen der speziellen Form von ϕ hat die Umkehrabbildung ϕ^{-1} die Form $\phi^{-1}(x, z) = (x, h(x, z))$ mit einer C^k-Abbildung $h : V' \to Y$. Somit gilt $f(x, y) = z^* \Leftrightarrow y = h(x, z^*)$ für $(x, y) \in U'$. Wähle nun abschließend offene Umgebungen U^* von x^* und V^* von y^* mit $U^* \times V^* \subset U'$ und $h(U^* \times \{z^*\}) \subset V^*$, dann hat die C^k-Abbildung $g(x) := h(x, z^*)$ die gewünschte Eigenschaft. $\square$

Beispiel 2.62 *Die Gleichung* $f(x,y) := x^2 + y^3 - 2xy = 0$ *kann man nahe der Lösung* $(1,1)$ *zwar nach* y *auflösen durch eine (beliebig oft) stetig differenzierbare und nahe* 1 *definierte Funktion* $g(x)$ *mit* $g(1) = 1$, *denn es gilt* $\frac{\partial f}{\partial y} = 3y^2 - 2x$ *und somit ist* $\frac{\partial f}{\partial y}(1,1) = 3 - 2 = 1$ *invertierbar. Aber man kann* $f(x,y) = 0$ *nicht lokal bei* $(1,1)$ *nach* x *auflösen, denn es gilt* $\frac{\partial f}{\partial x} = 2x - 2y$ *und* $\frac{\partial f}{\partial x}(1,1) = 2 - 2 = 0$ *ist nicht invertierbar.*

Die Ableitung von $g(x)$ *(und auch höhere Ableitungen) in* $x = 1$ *kann man aus der Formel* $f(x,g(x)) = 0$ *gewinnen, denn Ableiten dieser Formel liefert*

$$\frac{\partial f}{\partial x}(x,g(x)) + \frac{\partial f}{\partial y}(x,g(x))g'(x) = 0 \Rightarrow g'(1) = -\frac{0}{1} = 0$$

Beispiel 2.63 *Das Gleichungssystem*

$$\begin{aligned} f_1(x,y,z) &:= x^2 + y^2 + z^2 = 3 \\ f_2(x,y,z) &:= xy - yz + xz = 1 \end{aligned}$$

kann man nahe der Lösung $(1,1,1)$ *nach* (x,y) *auflösen, denn mit der Abbildung* $f := (f_1, f_2) : \mathbb{R}^3 \to \mathbb{R}^2$ *ist die durch die Matrix*

$$\begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix}(1,1,1) = \begin{pmatrix} 2 & 2 \\ 2 & 0 \end{pmatrix}$$

repräsentierte lineare Abbildung $d_Y f(1,1,1)$ *offensichtlich invertierbar. Ist* $g = (g_1, g_2)$ *die durch* $f(g_1(z), g_2(z), z) = (3,1)^T$ *implizit definierte Kurve mit* $g(1) = (1,1)$, *dann hat* $g(z)$ *wegen*

$$\begin{aligned} &\frac{\partial f_1}{\partial x}(g_1(z), g_2(z), z))g_1'(z) + \frac{\partial f_1}{\partial y}(g_1(z), g_2(z), z))g_2'(z) + \frac{\partial f_1}{\partial z}(g_1(z), g_2(z), z)) = 0 \\ &\qquad \Longrightarrow 2g_1'(1) + 2g_2'(1) + 2 = 0 \\ &\frac{\partial f_2}{\partial x}(g_1(z), g_2(z), z))g_1'(z) + \frac{\partial f_2}{\partial y}(g_1(z), g_2(z), z))g_2'(z) + \frac{\partial f_2}{\partial z}(g_1(z), g_2(z), z)) = 0 \\ &\qquad \Longrightarrow 2g_1'(1) = 0 \end{aligned}$$

die Ableitung $\dot{g}(1) = (0,-1)^T$.

Beispiel 2.64 *Sei* $f : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}$, $f(x,y) := y^n + a_{n-1}(x)y^{n-1} + \cdots + a_1(x)y + a_0(x)$, *ein Polynom in* y *mit stetig differenzierbar von* x *abhängigen Koeffizienten* $a_0, a_1, \ldots, a_{n-1}$. *Ist für festes* x^* *die Zahl* y^* *eine einfache Nullstelle des Polynoms* $y \mapsto f(x^*, y)$, *d.h. gilt* $f(x^*, y^*) = 0$ *und* $\frac{\partial f}{\partial y}(x^*, y^*) \neq 0$, *dann gibt es also eine auf einer Umgebung* U^* *von* x^* *definierte stetig differenzierbare Funktion* $g : U^* \to \mathbb{R}$ *mit* $g(x^*) = y^*$ *und* $f(x,y) = 0 \Leftrightarrow y = g(x)$ *nahe* (x^*, y^*).

Ist insbesondere $A(x)$ *eine von Parametern* $x \in \mathbb{R}^m$ *stetig differenzierbar abhängige* $n \times n$*-Matrix und hat* $A(x^*)$ *den algebraisch einfachen Eigenwert* $\lambda(x^*)$, *dann gibt es eine auf einer Umgebung* U^* *von* x^* *definierte stetig differenzierbare Funktion* $\lambda(x)$, *mit der* $\lambda(x)$ *für* $x \in U^*$ *der einzige Eigenwert von* $A(x)$ *nahe* $\lambda(x^*)$ *ist. Denn das*

Polynom $\lambda \mapsto \det(\lambda \operatorname{Id} - A(x))$ *hat in* x^* *die einfache Nullstelle* $\lambda(x^*)$*, also ergibt sich das Resultat direkt aus der zuvor betrachteten Situation. Etwas verkürzt ausgedrückt hängen also die Eigenwerte einer von* x *stetig differenzierbar abhängigen quadratischen Matrix auch stetig differenzierbar von* x *ab.*

2.7 Untermannigfaltigkeiten

Interessiert man sich nicht so sehr für die Parametrisierung $y = g(x)$ der Lösungen einer Gleichung $f(x, y) = z^*$ zu festem z^* nahe einer Lösung (x^*, y^*), sondern eher für die Gestalt der Lösungsmenge $M := \{(x, y) \,|\, f(x, y) = z^*\}$, so stellt sich die Frage, ob diese Lösungsmenge M nahe einer Lösung (x^*, y^*) von $f(x, y) = z^*$ irgendwelche Ecken, Kanten oder sonstige nicht-Glattheiten aufweist.

Wie wir schon in den Vorbemerkungen zum Satz über implizite Funktionen am Beispiel der Gleichung $f(x, y) := xy = 0$ und der Lösung $(x^*, y^*) = (0, 0)$ gesehen haben, kann M sehr wohl Ecken besitzen, selbst wenn f beliebig oft stetig differenzierbar ist. Wieder ist der Schlüssel zur Glattheit von M eine Bedingung, die uns die Invertierbarkeit eines Teils der Ableitung von f garantiert.

Definition 2.65 *Sei* $f : \mathbb{R}^n \to \mathbb{R}^m$ *eine differenzierbare Abbildung. Ein Punkt* $x \in \mathbb{R}^n$ *heißt regulärer Punkt von* f*, wenn* $df(x)$ *surjektiv ist (insbesondere muss dann* $n \geq m$ *sein). Ein Wert* $z^* \in \mathbb{R}^m$ *heißt regulärer Wert von* f*, wenn jedes* x *mit* $f(x) = z^*$ *ein regulärer Punkt ist, oder mit anderen Worten, wenn das Urbild* $f^{-1}(\{z^*\})$ *nur aus regulären Punkten besteht.*

Mittels dieser Definition können wir nun glatte d-dimensionale Flächen im $\mathbb{R}^n$ – sogenannte d-dimensionale Untermannigfaltigkeiten – durch den folgenden Satz charakterisieren, in dessen Beweis der Satz über implizite Funktionen essentiell eingeht.

Satz 2.66 *Für eine nichtleere Teilmenge* $M \subset \mathbb{R}^n$ *sind äquivalent:*

(a) *Lokal ist* M C^k*-diffeomorph zum* $\mathbb{R}^d$*, d.h. für jeden Punkt aus* M *gibt es eine offene Umgebung* $U \subset \mathbb{R}^n$ *und einen* C^k*-Diffeomorphismus* $\phi : U \to V$ *auf eine offene Teilmenge* $V \subset \mathbb{R}^n$ *mit* $\phi(M \cap U) = \mathbb{R}^d \cap V$*, wobei wir* $\mathbb{R}^d$ *als den Unterraum* $\{x \in \mathbb{R}^n \,|\, x_{d+1} = \cdots = x_n = 0\}$ *von* $\mathbb{R}^n$ *auffassen („*ϕ *biegt die* d*-dimensionale Fläche* M *gerade“).*

(b) *Lokal ist* M *Urbild eines regulären Wertes einer* C^k*-Funktion vom* $\mathbb{R}^n$ *in den* $\mathbb{R}^{n-d}$*, d.h. für jeden Punkt aus* M *gibt es eine Umgebung* $U \subset \mathbb{R}^n$*, eine* C^k*-Funktion* $f : U \to \mathbb{R}^{n-d}$ *und einen regulären Wert* $z^* \in \mathbb{R}^{n-d}$ *von* f *mit* $M \cap U = f^{-1}(\{z^*\})$*.*

(c) *Lokal ist* M *der Graph einer* C^k*-Funktion auf dem* $\mathbb{R}^d$*, d.h. für jeden Punkt* $a \in M$ *gibt es eine Zerlegung* $\mathbb{R}^n \cong X \times Y$ *in einen* d*-dimensionalen linearen Unterraum* X *und einen* $(n-d)$*-dimensionalen linearen Unterraum* Y *von* $\mathbb{R}^n$ *sowie bei* $a = (x^*, y^*)$ *offene Umgebungen* $U^* \subset X$ *von* x^**,* $V^* \subset Y$ *von* y^* *und eine* C^k*-Abbildung* $g : U^* \to V^*$ *mit* $M \cap U^* \times V^* = \operatorname{Graph}(g)$*, wobei* $\operatorname{Graph}(g) := \{(x, g(x)) | x \in U^*\}$ *den Graphen von* g *bezeichnet.*

Beweis:

(a) $\Rightarrow$ (b): Ist ϕ wie angegeben, dann ist 0 ein regulärer Wert von $f := (\phi_{d+1}, \dots, \phi_n)|_U$ und $f^{-1}(\{0\}) = M \cap U$.

(b) $\Rightarrow$ (c): Sei f wie angegeben und $a \in M$. Definiere $X := \operatorname{Kern}(df(a)) \cong \mathbb{R}^d$ und wähle ein Komplement $Y \cong \mathbb{R}^{n-d}$ von X in $\mathbb{R}^n$, dann gilt $X \times Y \cong \mathbb{R}^n$ mittels des stetigen linearen Isomorphismus $(x, y) \mapsto x + y$.

Da X gerade der Kern von $df(a)$ ist, ist $df(a)|_Y$ ein Isomorphismus von Y nach $Z := \mathbb{R}^{n-d}$. Also kann man den Satz über implizite Funktionen anwenden und erhält für den $a \in \mathbb{R}^n$ entsprechenden Punkt $(x^*, y^*) \in X \times Y$ Umgebungen U^* von x^* und V^* von y^* sowie eine C^k-Abbildung $g : U^* \to V^*$ mit $f(x, y) = z^* \Leftrightarrow y = g(x)$ für $(x, y) \in U^* \times V^*$. Somit ist $f^{-1}(z^*) \cap (U^* \times V^*) = \operatorname{Graph}(g)$, was zu zeigen war.

(c) $\Rightarrow$ (a): Ist g wie angegeben, so bildet $\psi(x, y) := x + g(x) - y$ den Graphen von g in den d-dimensionalen Unterraum X von $\mathbb{R}^n$ hinein ab und ist wegen $d\psi(x, y) = \begin{pmatrix} \mathrm{Id}_X & 0 \\ dg(x) & -\mathrm{Id}_Y \end{pmatrix}$ ein lokaler C^k-Diffeomorphismus. Schaltet man nun noch hinter ψ einen linearen Isomorphismus des $\mathbb{R}^n$, der X auf $\{x \in \mathbb{R}^n \mid x_{d+1} = \dots = x_n = 0\}$ abbildet, so erhält man den gewünschten Diffeomorphismus ϕ.

$\square$

Man bemerke, dass wir den vorigen Satz nur für den $\mathbb{R}^n$ formuliert haben, man könnte ihn aber natürlich ganz analog auch für allgemeine endlich-dimensionale Vektorräume angeben. Tatsächlich muß man ihn aber für unendlich-dimensionale Banach-Räume modifizieren, da es dort nicht selbstverständlich ist, dass man zu einem linearen Unterraum X eine komplementären Unterraum Y und einen stetigen linearen Isomorphismus von $X \times Y$ in den ursprünglichen Raum findet.

Nun aber zur wichtigen Definition einer Untermannigfaltigkeit.

Definition 2.67 *Eine Teilmenge $M \subset \mathbb{R}^n$, die eine (und somit jede) der Eigenschaften aus dem vorigen Satz besitzt, heißt d-dimensionale C^k-Untermannigfaltigkeit des $\mathbb{R}^n$.*

Untermannigfaltigkeiten sind also d-dimensionale Flächen, die sich lokal geradebiegen lassen, oder äquivalenterweise lokal Urbilder regulärer Punkte sind, oder äquivalenterweise lokal Graphen sind.

Beispiel 2.68 *Jede offene Teilmenge $U \subset \mathbb{R}^n$ ist nach dem ersten Punkt von Satz 2.66 mit $\phi := \mathrm{Id}_U$ eine n-dimensionale glatte Untermannigfaltigkeit des $\mathbb{R}^n$.*

Beispiel 2.69 *Die Einheits-Sphäre $S^n := \{x \in \mathbb{R}^{n+1} | \, \|x\|_2^2 = 1\}$ im Euklidischen $\mathbb{R}^{n+1}$ (also bzgl. der Euklidischen Norm $\|x\|_2^2 := \sum_{i=0}^{n} x_i^2$) ist das Urbild des regulären Wertes Eins unter der C^∞-Funktion $f : \mathbb{R}^{n+1} \to \mathbb{R}$, $f(x) := \|x\|_2^2$, und somit eine n-dimensionale C^∞-Untermannigfaltigkeit. Tatsächlich ist Eins ein regulärer Wert, denn* *ist $f(x) = 1$, so gilt $df(x) = 2x^T \neq 0$ wegen $\|x\|_2 = 1$, und somit ist $df(x)$ als lineare Abbildung von $\mathbb{R}^{n+1}$ nach $\mathbb{R}$ surjektiv.*

Beispiel 2.70 *Der Konfigurationsraum eines Stabes, d.h. zweier Punkte im $\mathbb{R}^3$, die in einem festem Abstand r miteinander verbunden sind, ist die 5-dimensionale Untermannigfaltigkeit $M := \{(x,y) \in \mathbb{R}^6 \,|\, \|x-y\|_2^2 = r^2\}$ des $\mathbb{R}^6$. Tatsächlich ist M eine 5-dimensionale C^∞-Untermannigfaltigkeit des $\mathbb{R}^6$, denn r^2 ist regulärer Wert von $f(x,y) := \|x-y\|_2^2$, da $df(x,y) = \begin{pmatrix} 2(x-y)^T & -2(x-y)^T \end{pmatrix}$ bei $\|x-y\|_2^2 = r^2$ ungleich Null und somit surjektiv ist. Allgemeiner hat jeder Roboter, der aus Stäben und Drehgelenken zusammengebaut ist, eine C^∞-Untermannigfaltigkeit als Konfigurationsraum.*

Beispiel 2.71 *Die Menge der orthogonalen Matrizen $O(n) := \{A \in \mathbb{R}^{n\times n} \,|\, A^T A = E_n\}$ ist eine $\frac{n(n-1)}{2}$-dimensionale Untermannigfaltigkeit des $\mathbb{R}^{n^2}$. Denn $O(n)$ ist das Urbild der Einheitsmatrix E_n unter der C^∞-Abbildung $f(A) := A^T A$ von $\mathbb{R}^{n\times n} \cong \mathbb{R}^{n^2}$ in den Vektorraum der symmetrischen Matrizen $S(n) \cong \mathbb{R}^{n(n+1)/2}$, und E_n ist ein regulärer Wert dieser Abbildung. Tatsächlich, $A^T A$ ist immer symmetrisch, und die Ableitung $df(A)H = A^T H + H^T A$ ist surjektiv für A mit $A^T A = E_n$, da dann die Gleichung $A^T H + H^T A = S$ für symmetrisches S nämlich die Lösung $H = \frac{1}{2} AS$ hat.*

Tangentialräume Der Tangentialraum $T_a M$ im Punkt $a \in M$ einer d-dimensionalen C^k-Untermannigfaltigkeit M des $\mathbb{R}^n$ ist anschaulich der d-dimensionale Unterraum, der sich an M anschmiegt, wenn man ihn in den Punkt a verschiebt. Die folgende Definition präzisiert dies:

Definition 2.72 *Sei M eine d-dimensionale C^k-Untermannigfaltigkeit des $\mathbb{R}^n$. Der d-dimensionale Unterraum $T_a M$ des $\mathbb{R}^n$ bestehe aus allen möglichen Ableitungen $\dot{c}(0)$ von parametrisierten stetig differenzierbaren Kurven $c : I \to M$, die innerhalb von M durch den Punkt $c(0) = a$ verlaufen (wobei $0 \in I$ angenommen wurde).*

Tatsächlich ist $T_a M$ für eine d-dimensionale C^k-Untermannigfaltigkeit M des $\mathbb{R}^n$ ein d-dimensionaler Unterraum. Denn ist $\phi : U \to V$ ein Diffeomorphismus von einer offenen Umgebung U von a auf eine offene Teilmenge V mit

$$\phi(M \cap U) = \{x \in V \,|\, x_{d+1} = \cdots = x_n = 0\} \quad ,$$

dann ordnet ϕ jeder Kurve $c : I \to M \cap U$ die Bildkurve $\tilde{c} := \phi \circ c$ zu. Umgekehrt gibt es zu jeder Kurve

$$\tilde{c} : I \to \{x \in V \,|\, x_{d+1} = \cdots = x_n = 0\}$$

eine Kurve $c : I \to M \cap U$ mit $\tilde{c} = \phi \circ c$, nämlich $c := \phi^{-1} \circ \tilde{c}$.

Nach der Kettenregel gilt $\dot{c}(0) = (d\phi(a))^{-1}\dot{\tilde{c}}(0)$, und da $\dot{\tilde{c}}(0)$ ein beliebiger Vektor im d-dimensionalen Raum $\{x \in \mathbb{R}^n \,|\, x_{d+1} = \cdots = x_n = 0\}$ ist, ist $T_a M = T_a(M \cap U)$ als Bild von $\{x \in \mathbb{R}^n \,|\, x_{d+1} = \cdots = x_n = 0\}$ unter der linearen Abbildung $d\phi(a))^{-1}$ ein d-dimensionaler Unterraum.

Der folgende Satz ist nützlich, um den Tangentialraum $T_a M$ in $a \in M$ auszurechnen, falls M lokal bei a das Urbild eines regulären Wertes z^* unter f ist.

Satz 2.73 *Ist die d-dimensionale C^k-Untermannigfaltigkeit M des $\mathbb{R}^n$ lokal bei $a \in M$ das Urbild eines regulären Wertes $z^* \in \mathbb{R}^{n-d}$ unter der C^k-Abbildung $f : \mathbb{R}^n \to \mathbb{R}^{n-d}$, dann gilt $T_aM = \mathrm{Kern}(df(a))$.*

Beweis: Ist $c : I \to M$ eine Kurve in M mit $c(0) = a$, dann gilt $f(c(t)) = z^*$ für alle genügend nah bei 0 liegenden $t \in I$. Ableiten der Gleichung nach t ergibt bei $t = 0$ nach der Kettenregel $df(a)\dot{c}(0) = 0$, also gilt $\dot{c}(0) \in \mathrm{Kern}(df(a))$ und somit nach Definition des Tangentialraumes $T_aM \subset \mathrm{Kern}(df(a))$. Nun ist $\mathrm{Kern}(df(a))$ aber höchstens d-dimensional, da $df(a)$ als lineare Abbildung von $\mathbb{R}^n$ nach $\mathbb{R}^{n-d}$ surjektiv ist, denn a ist ja ein regulärer Punkte (weil z^* ein regulärer Wert ist). Also gilt sogar $T_aM = \mathrm{Kern}(df(a))$. □

Beispiel 2.74 *Die Einheits-Sphäre $S^n := \{x \in \mathbb{R}^{n+1} | \, \|x\|_2^2 = 1\}$ im Euklidischen $\mathbb{R}^{n+1}$ ist das Urbild des regulären Wertes Eins unter der durch $f(x) := \|x\|_2^2$ gegebenen C^∞-Funktion $f : \mathbb{R}^{n+1} \to \mathbb{R}$, und da f in einem Punkt $a \in M$ die Ableitung $df(a) = 2a^T$ besitzt, gilt $T_aS^n = \{x \in \mathbb{R}^{n+1} \,|\, a^Tx = 0\}$.*

Beispiel 2.75 *Die Menge der orthogonalen Matrizen $O(n) := \{A \in \mathbb{R}^{n\times n} \,|\, A^TA = E_n\}$ ist das Urbild der Einheitsmatrix E_n unter der C^∞-Abbildung $f(A) := A^TA$ von $\mathbb{R}^{n\times n}$ in die symmetrischen Matrizen $S(n)$. Speziell in $A := E_n$ gilt $df(E_n)H = H + H^T$ und daher besteht $T_{E_n}O(n) = \{A \in \mathbb{R}^{n\times n} \,|\, A = -A^T\}$ aus den schiefsymmetrischen Matrizen.*

2.8 Extrema unter Nebenbedingungen

Das folgende Problem spielt eine große Rolle sowohl innerhalb der Mathematik als auch in den sie anwendenden Wissenschaften:

Gesucht sind lokale Maxima bzw. Minima einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$, wobei jedoch nicht alle Punkte $x \in \mathbb{R}^n$ zugelassen sind, sondern nur diejenigen, die den Nebenbedingungen $g(x) = 0$ für eine vorgegebene Abbildung $g : \mathbb{R}^n \to \mathbb{R}^k$ mit $k < n$ genügen.

Genauer sagt man, dass in $x^* \in U$ mit $g(x^*) = 0$ ein lokales Maximum von f unter den Nebenbedingungen $g = 0$ liegt, wenn es eine Umgebung $U \subset \mathbb{R}^n$ von x^* gibt mit $f(x) \le f(x^*)$ für alle $x \in U$ mit $g(x) = 0$ (und analog für Minima).

Mit anderen Worten sucht man lokale Extrema der Einschränkung $f : M \to \mathbb{R}$ von f auf die Teilmenge $M := \{x \in \mathbb{R}^n \,|\, g(x) = 0\}$. Die Multiplikatorregel von Lagrange ist eine notwendiges Kriterium für das Vorliegen eines solchen Extremums unter Nebenbedingungen in einem Punkt x^*, in dessen Nähe M eine $(n-k)$-dimensionale C^1-Untermannigfaltigkeit ist.

Satz 2.76 *Sei $U \subset \mathbb{R}^n$ offen, $f : U \to \mathbb{R}$ differenzierbar und $g = (g_1, \ldots, g_k) : U \to \mathbb{R}^k$ stetig differenzierbar. Liegt in x^* ein lokales Extremum von f unter den Nebenbedingungen $g = 0$ und ist die lineare Abbildung $dg(x^*) : \mathbb{R}^n \to \mathbb{R}^k$ surjektiv, d.h. sind die*

Gradienten grad $g_1(x^*), \ldots,$ grad $g_k(x^*)$ *linear unabhängig, dann gibt es einen Vektor* $\lambda = (\lambda_1, \ldots, \lambda_k) \in \mathbb{R}^k$ *mit* grad $f(x^*) = \sum_{i=1}^{k} \lambda_i \operatorname{grad} g_i(x^*)$.

Beweis: Da $dg(x^*)$ surjektiv ist, gibt es auch eine Umgebung U' von x^*, so dass $dg(x)$ für jedes $x \in U$ surjektiv ist. Denn ist $A(x^*)$ eine $(k \times k)$-Untermatrix von $Jg(x^*)$ mit $\det(A(x^*)) \neq 0$, dann gibt es aufgrund der Stetigkeit der Ableitung von g auch eine Umgebung U', auf der die entsprechende $(k \times k)$-Untermatrix $A(x)$ für alle Punkte $x \in U$ die Bedingung $\det(A(x)) \neq 0$ erfüllt.

Also ist $M \cap U' = \{x \in U' \mid g(x) = 0\}$ eine $(n-k)$-dimensionale C^1-Untermannigfaltigkeit des $\mathbb{R}^n$, denn 0 ist regulärer Wert, da für alle Punkte $x \in U'$ die Ableitung $dg(x)$ surjektiv ist, und insbesondere gilt dies also für die $x \in U'$ mit $g(x) = 0$.

Um den Satz zu beweisen, müssen wir nur zeigen, dass grad $f(x^*)$ von den Vektoren grad $g_i(x^*)$, $i = 1, \ldots, k$, linear abhängt, dann gibt es nämlich $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ mit grad $f(x^*) = \sum_{i=1}^{k} \lambda_i \operatorname{grad} g_i(x^*)$. Die lineare Abhängigkeit ist aber gleichbedeutend damit, dass für jeden Vektor $v \in \mathbb{R}^n$ aus $dg(x^*)v = 0$ schon $df(x^*)v = 0$ folgt.

Nun ist $v \in \mathbb{R}^n$ wegen $T_{x^*}(M \cap U') = \operatorname{Kern}(dg(x^*))$ genau dann ein Vektor mit $dg(x^*)v = 0$, wenn $v \in T_{x^*}M$ im Tangentialraum an $M \cap U'$ im Punkt x^* liegt. Dann gibt es aber eine Kurve $c : I \to M \cap U'$ mit $c(0) = x^*$ und $\dot{c}(0) = v$, und die reelle Funktion $f \circ c : I \to \mathbb{R}$ hat somit in 0 ein lokales Extremum, da f in x^* ein lokales Extremum unter den Nebenbedingungen $g = 0$ hat, die ja gerade $M \cap U'$ definieren. Also gilt nach der Kettenregel $0 = \frac{d}{dt}(f \circ c)(0) = df(c(0))\dot{c}(0) = df(x^*)v$, was zu beweisen war. □

Man kann Satz 2.76 auch folgendermaßen umformulieren: Ist x^* ein lokales Extremum von f unter den Nebenbedingungen $g = 0$ und hat $dg(x^*) : \mathbb{R}^n \to \mathbb{R}^k$ maximalen Rang, so gibt es einen Vektor λ, für den (x^*, λ) ein stationärer Punkt der erweiterten Funktion $F(x, \lambda) := f(x) - \sum_{i=1}^{k} \lambda_i g_i(x)$ ist, d.h. $dF(x^*, \lambda) = 0$ gilt. Den Vektor λ nennt man Lagrange-Multiplikator und seine Komponenten λ_i Lagrange-Multiplikatoren.

Sind f und g zweimal stetig differenzierbar, dann kann man sogar ein hinreichendes Kriterium für das Vorliegen eines Extremums unter Nebenbedingungen formulieren: Hat $dg(x^*)$ für einen Punkt x^* maximalen Rang, gilt mit einem Vektor λ die Gleichung $dF(x^*, \lambda) = 0$, und ist die Einschränkung der Hesse-Matrix Hess $F_\lambda(x^*)$ (wobei hier bei festem λ nur nach x abgeleitet wird) auf den Tangentialraum $T_{x^*}M = \operatorname{Kern}(dg(x^*))$ positiv bzw. negativ definit, so liegt in x^* ein lokales Minimum bzw. Maximum.

Beispiel 2.77 *Sei M eine Untermannigfaltigkeit des $\mathbb{R}^n$ und liege der Punkt a außerhalb von M. Unter allen Punkten $x \in M$ sei der Euklidische Abstand von x und a im Punkt $x^* \in M$ am kleinsten (solch ein Punkt existiert nach dem auf Korollar 1.60 folgenden Beispiel jedenfalls dann, wenn M kompakt ist). Dann steht die Verbindungsgerade von a und x^* im Punkt x^* senkrecht auf M.*

Sei nämlich M lokal nahe x^* das Urbild von 0 unter der Abbildung g mit surjektiver Ableitung dg nahe x^*. Dann ist also x^* ein (lokales) Minimum der Funktion $f(x) := \|x - a\|_2^2$ unter den Nebenbedingungen $g = 0$. Daher gibt es Zahlen $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ mit

$2(x^*-a) = \operatorname{grad} f(x^*) = \sum_{i=1}^{k} \lambda_i \operatorname{grad} g_i(x^*)$. Nun besteht wegen $v \in T_{x^*}M \Leftrightarrow dg(x^*)v = 0$ und $dg(x^*)v = (\langle \operatorname{grad} g_i(x^*), v\rangle)_{i=1,\dots,k}$ der Tangentialraum $T_{x^*}M$ aber gerade aus den Vektoren, die senkrecht auf den Gradienten $\operatorname{grad} g_i(x^*)$ stehen (und damit auch auf jeder Linearkombination davon). Also steht auch der Vektor $2(x^* - a)$ senkrecht auf $T_{x^*}M$, oder mit anderen Worten steht die Verbindungsgerade von a und x^* im Punkt x^* senkrecht auf M.

Beispiel 2.78 *Man kann auch das Problem, Eigenwerte und Eigenvektoren einer Matrix zu finden, als ein Extremwertproblem unter Nebenbedingungen auffassen. Dazu suche man zu einer $n \times n$-Matrix A nach dem Maximum von $f(x) := \langle Ax, x\rangle$ unter der Nebenbedingung $\|x\|_2^2 = 1$, d.h. dem Maximum von f auf der $(n-1)$-Sphäre $S^{n-1} = \{x \in \mathbb{R}^n \mid \|x\|_2^2 = 1\}$. Solch eines existiert nach Korollar 1.60 aufgrund der Kompaktheit von S^{n-1} und der Stetigkeit von f immer.*

Die Funktion $g(x) := \|x\|_2^2 - 1$ besitzt den Gradienten $\operatorname{grad} g(x) = 2x$, und dieser verschwindet nur bei $x = 0$, insbesondere also auf einer Umgebung der Sphäre S^{n-1} nicht. Darüberhinaus besitzt f zumindest für symmetrische Matrizen A den Gradienten $2Ax$. Liegt also in einem Punkt x das Maximum von f auf S^{n-1}, so gibt es nach der Lagrangeschen Multiplikatorregel 2.76 ein $\lambda \in \mathbb{R}$ mit $2Ax = 2\lambda x$, und wegen $\|x\|_2^2 = 1$ ist der Wert λ aufgrund von

$$f(x) = \langle Ax, x\rangle = \langle \lambda x, x\rangle = \lambda\langle x, x\rangle = \lambda$$

der Maximalwert von f auf S^{n-1}.

Für symmetrische Matrizen A ist jede Maximalstelle x von f auf S^{n-1} also ein Eigenvektor von A zum Eigenwert $\lambda = f(x)$.

Kapitel 3

Integralrechnung

Zum Abschluß der Analysis II wird in diesem Kapitel die mehrdimensionale Integration diskutiert. Beginnen wollen wir mit dem iterierten Riemann-Integral einer stetigen Funktion auf einem abgeschlossenen Quader $Q \subset \mathbb{R}^n$ bzw. einer stetigen Funktion mit kompaktem Träger. Dieser Integral-Begriff ist zwar ausreichend für elementare differentialgeometrische Rechnungen, jedoch ist der mit der Norm $\|f\|_1 := \int_{\mathbb{R}^n} |f(x)|\, dx$ versehene Raum stetiger Funktionen mit kompaktem Träger nicht vollständig.

Um wesentlich mehr Funktionen integrieren zu können, müssen wir daher den Integral-Begriff erweitern, was aufgrund von Satz 1.31 einfach durch das Bilden der Vervollständigung geschehen kann. Jedoch ist die Vervollständigung zunächst nur ein abstrakter Banach-Raum, und insbesondere ist nicht klar, ob die bei der Vervollständigung dazugewonnenen Elemente wiederum als Funktionen interpretiert werden können.

Um dies garantieren zu können, beobachten wir zunächst, dass sich die charakteristische Funktion 1_Q eines Quaders Q bzgl. $\|\cdot\|_1$ beliebig genau durch stetige Funktionen mit kompaktem Träger annähern läßt. Daher stimmt die Vervollständigung des Raumes stetiger Funktionen mit kompaktem Träger bzgl. $\|\cdot\|_1$ mit der Vervollständigung des Raumes von Treppenfunktionen $f = \sum_{i=1}^{k} f_i 1_{Q_i}$ zu Quadern Q_i bzgl. der (Halb-)Norm $\|f\|_1 := \sum_{i=1}^{k} |f_i| \operatorname{Vol}(Q_i)$ überein, wobei $\operatorname{Vol}(Q)$ das Elementarvolumen eines Quaders Q ist.

Aufgrund dieser Beobachtung beschäftigen wir uns in Verallgemeinerung des Volumens von Quadern mit Prämaßen und zeigen, wie man diese zu Maßen fortsetzen kann. Dies erlaubt es uns zu beweisen, dass Cauchy-Folgen von Treppenfunktionen bzgl. $\|\cdot\|_1$ auch schon punktweise fast überall konvergieren, die Elemente der Vervollständigung also als Funktionen interpretiert werden können. Darüberhinaus kann man das Integral von Treppenfunktionen auf die Vervollständigung fortsetzen und erhält so das Lebesgue-Integral.

Der hier gewählte Zugang zum Lebesgue-Integral hat den Vorteil, dass er nicht nur die Integration reellwertiger Funktionen erlaubt, sondern auch ohne zusätzliche Schwierigkeiten die Integration von Funktionen mit Werten in einem beliebigen Banach-Raum.

Nachdem wir auf diese Weise das Lebesgue-Integral eingeführt haben, diskutieren wir Konvergenzsätze, führen die L^p-Räume ein und beschäftigen uns mit der kontinuierlichen Fourier-Transformation.

Abschließend sprechen wir noch kurz die Integration über Untermannigfaltigkeiten an und beweisen den Satz von Gauss.

3.1 Iterierte Riemann-Integrale

Sei $Q := [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$ ein (abgeschlossener) n-dimensionaler Quader und $f : Q \to \mathbb{R}$ eine stetige Funktion. Nach Satz 1.66 wird durch die Integration von $f(x_1, x_2, \ldots, x_n)$ nach der Variable x_1 eine stetige Funktion

$$F(x_2, \ldots, x_n) := \int_{a_1}^{b_1} f(x_1, x_2, \ldots, x_n)\, dx_1$$

auf dem (abgeschlossenen) $(n-1)$-dimensionalen Quader $[a_2, b_2] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^{n-1}$ definiert. Diese kann man nun wiederum nach x_2 integrieren, und setzt man dieses Verfahren bis zur Variablen x_n fort, so erhält man abschließend eine reelle Zahl

$$\int_{a_n}^{b_n} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2\, \ldots\, dx_n \quad ,$$

die das iterierte Riemann-Integral von f über Q genannt wird und kurz durch $\int_Q f(x)\, dx$ symbolisiert wird.

Beispiel 3.1 *Das Integral der konstanten Funktion $f := 1$ über den (abgeschlossenen) Quader $Q := [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$ ist der Wert $\mathrm{Vol}(Q) := \int_Q 1\, dx = \prod_{i=1}^{n} (b_i - a_i)$. Diesen bezeichnet man üblicherweise als das n-dimensionale Elementarvolumen von Q. Genauer schreibt man $\mathrm{Vol}_n(Q)$, wenn man die Dimension explizit deutlich machen will.*

Der folgende Begriff dient dazu, sich bei der Integration nicht immer auf einen festen Quader beziehen zu müssen.

Definition 3.2 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ stetig. Die Menge $\mathrm{Tr}(f) := \overline{\{x \in \mathbb{R}^n \mid f(x) \neq 0\}}$ der Punkte, in deren Nähe f nicht verschwindet, wird der Träger von f genannt. Ist der Träger von f eine kompakte Teilmenge (oder äquivalenterweise beschränkt, denn abgeschlossen ist er schon aufgrund seiner Definition), dann bezeichnet man f als eine stetige Funktion mit kompaktem Träger. Die Menge aller stetigen Funktionen mit kompakten Träger symbolisiert man durch $C_c(\mathbb{R}^n, \mathbb{R})$.*

Offensichtlich ist $C_c(\mathbb{R}^n, \mathbb{R})$ ein Vektorraum, und da jedes $f \in C_c(\mathbb{R}^n, \mathbb{R})$ sowohl Maximum als auch Minimum besitzt, ist $C_c(\mathbb{R}^n, \mathbb{R})$ mit $\|f\|_\infty := \sup_{x \in \mathbb{R}^n} |f(x)|$ ein normierter Vektorraum (jedoch nicht vollständig).

Aufgabe: *Geben Sie eine Cauchy-Folge in* $(C_c(\mathbb{R}^n, \mathbb{R}), \|\cdot\|_\infty)$ *an, die nicht konvergiert.*

Darüberhinaus ist der Träger $\mathrm{Tr}(f)$ jeder Funktion $f \in C_c(\mathbb{R}^n, \mathbb{R})$ in einem Quader $Q \subset \mathbb{R}^n$ enthalten. Mit solch einem Quader Q definiere man das Integral von f durch

$$\int_{\mathbb{R}^n} f(x)\, dx := \int_Q f(x)\, dx \quad .$$

Diese Definition ist augenscheinlich unabhängig von der Wahl des den Träger von f enthaltenden Quaders Q.

Beispiel 3.3

- *Mit der durch* $f(x, y) := 1 - x^2 - y^2$ *für* $(x, y) \in B_1(0)$ *und* $f(x, y) := 0$ *für* $(x, y) \notin B_1(0)$ *definierten stetigen Funktion* $f : \mathbb{R}^2 \to \mathbb{R}$ *auf dem Euklidischen* $\mathbb{R}^2$*, deren Träger* $\overline{B_1(0)} \subset [-1, 1] \times [-1, 1]$ *ist, gilt*

$$\int_{\mathbb{R}^n} f(x)\, dx = \int_{-1}^{1} \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} 1 - x^2 - y^2\, dx\, dy =$$
$$\int_{-1}^{1} \left((1 - y^2)x - \frac{1}{3}x^3 \right) |_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}\, dy = \frac{4}{3} \int_{-1}^{1} (1 - y^2)^{3/2}\, dy =$$
$$\frac{4}{3} \int_{-\pi/2}^{\pi/2} \cos^4(t)\, dt = \int_{-\pi/2}^{\pi/2} \cos^2(t)\, dt = \pi/2$$

  wobei in den letzten Schritten Substitution und partielle Integration angewendet wurde.

- *Sind* $f_i : \mathbb{R} \to \mathbb{R}$ *stetige Funktionen mit kompaktem Träger, dann ist* $f(x_1, \ldots, x_n) := \prod_{i=1}^{n} f_i(x_i)$ *eine stetige Funktion* $f : \mathbb{R}^n \to \mathbb{R}$ *mit kompaktem Träger, und es gilt* $\int_{\mathbb{R}^n} f(x)\, dx = \prod_{i=1}^{n} \int_{\mathbb{R}} f_i(x_i)\, dx_i$.

Das so definierte iterierte Integral $I : C_c(\mathbb{R}^n, \mathbb{R}) \to \mathbb{R}$, $I(f) := \int_{\mathbb{R}^n} f(x)\, dx$, hat die Eigenschaften, die man auch von einem Integral erwartet.

Satz 3.4 *Das Integral* $I : C_c(\mathbb{R}^n, \mathbb{R}) \to \mathbb{R}$ *ist*

(a) linear, d.h. $I(\lambda f + \mu g) = \lambda I(f) + \mu I(g)$ *gilt für alle* $f, g \in C_c(\mathbb{R}^n, \mathbb{R})$ *und* $\lambda, \mu \in \mathbb{R}$*,*

(b) monoton, d.h. gilt für jedes $x \in \mathbb{R}^n$ *die Ungleichung* $f(x) \leq g(x)$*, dann auch* $I(f) \leq I(g)$.

Beweis: Dies folgt durch n-malige Anwendung direkt aus der Linearität und Monotonie des Riemannschen Integrals für auf einem Intervall $[a, b]$ definierte Funktionen. $\square$

Man bemerke allerdings, dass es viele Funktionale auf $C_c(\mathbb{R}^n, \mathbb{R})$ gibt, die linear und monoton sind.

Aufgabe: *Für jede stetige nichtnegative Funktion g ist $f \mapsto I(fg)$ linear und monoton.*

Allerdings ist I zusätzlich noch translationsinvariant.

Satz 3.5 *I ist translationsinvariant, d.h. für jedes $f \in C_c(\mathbb{R}^n, \mathbb{R})$ gilt mit der durch $(T_y f)(x) := f(x - y)$ definierten Translation um einen Vektor $y \in \mathbb{R}^n$ die Gleichung $I(f) = I(T_y f)$.*

Beweis: Da I durch n-fache Iteration aus dem Riemann-Integral reeller Funktionen entsteht, brauchen wir nur zu zeigen, dass für eine stetige Funktion $f : \mathbb{R} \to \mathbb{R}$ mit kompaktem Träger $\mathrm{Tr}(f) \subset [a, b]$ die Translation $T_y f$ um $y \in \mathbb{R}$, die selbstverständlich eine stetige Funktion mit kompaktem Träger $\mathrm{Tr}(T_y f) \subset [a + y, b + y]$ ist, das selbe Integral wie f hat. Dies folgt aber aus der Substitutionsregel

$$I(f) = \int_a^b f(x)\, dx = \int_{a+y}^{b+y} f(x - y)\, dx = I(T_y f) \quad .$$

□

Die Translationsinvarianz zeichnet I unter allen linearen und monotonen Funktionalen auf $C_c(\mathbb{R}^n, \mathbb{R})$ aus. Tatsächlich ist jedes lineare, monotone und translationsinvariante Funktional auf $C_c(\mathbb{R}^n, \mathbb{R})$ ein Vielfaches von I (siehe [Forster III, 1, Satz 3]). Statt dies hier zu zeigen, beweisen wir später in Satz 3.39 die Aussage, dass jedes translationsinvariante Maß ein Vielfaches des Lebesgue-Maßes ist.

Abschließend wollen wir die sogenannte L^1-Norm $\|f\|_1 := \int_{\mathbb{R}^n} |f(x)|\, dx$ auf $C_c(\mathbb{R}^n, \mathbb{R})$ definieren. Leider ist $C_c(\mathbb{R}^n, \mathbb{R})$ mit dieser Norm nicht vollständig.

Um eine größere Klasse von integrierbaren Funktionen zu gewinnen, könnte man daher den normierten Vektorraum $(C_c(\mathbb{R}^n, \mathbb{R}), \|\cdot\|_1)$ einfach vervollständigen. Das Integral läßt sich nämlich auf die Vervollständigung fortsetzen, denn ist $f_k \in C_c(\mathbb{R}^n, \mathbb{R})$ bzgl. $\|\cdot\|_1$ eine Cauchy-Folge, dann ist wegen

$$\left| \int_{\mathbb{R}^n} f_k(x)\, dx - \int_{\mathbb{R}^n} f_l(x)\, dx \right| \leq \int_{\mathbb{R}^n} |f_k - f_l|\, dx = \|f_k - f_l\|_1$$

auch die Folge der Integrale $\int_{\mathbb{R}^n} f_k(x)\, dx$ eine Cauchy-Folge in $\mathbb{R}$. Diese konvergiert also gegen einen Wert in $\mathbb{R}$, den man als das Integral des Grenzwertes von f_k in der Vervollständigung ansehen könnte. Allerdings können wir mit unserer bisherigen Begriffsbildung nicht klären, ob (bzw. in welchem Sinne) man die Elemente des durch die Vervollständigung entstandenen Banach-Raumes als Funktionen auffassen kann. Um dies leichter diskutieren zu können, werden wir im folgenden nicht den Raum stetiger Funktionen mit kompaktem Träger vervollständigen, sondern den Raum der Treppenfunktionen zu Quadern. Vorher wollen wir uns aber vergewissern, dass bei der Vervollständigung in beiden Fällen derselbe Banach-Raum entsteht.

Approximation von Quadern Das folgende Lemma zeigt, dass man die charakteristische Funktion $1_Q : \mathbb{R}^n \to \mathbb{R}$ eines abgeschlossenen Quaders $Q \subset \mathbb{R}^n$, die durch $1_Q(x) := 1$ bei $x \in Q$ und $1_Q(x) := 0$ bei $x \notin Q$ definiert ist, beliebig genau durch stetige Funktionen mit kompaktem Träger im Sinne der L^1-Norm annäheren kann. Dann kann man natürlich auch die charakteristische Funktion eines beliebigen Quaders (also eines Produktes aus n abgeschlossenen, offenen oder halboffenen beschränkten, möglicherweise sogar nur aus einem Punkt bestehenden Intervallen) im Sinne der L^1-Norm beliebig genau annäheren, da bei der Integration über ein Intervall der Wert in einem Randpunkt keine Rolle spielt.

Man bemerke dazu, dass für eine stetige Funktion mit kompaktem Träger das iterierte Riemann-Integral $\int_{\mathbb{R}^n} f(x) - 1_Q(x)\, dx$ Sinn macht, denn der Integrand hat kompakten Träger und ist für jedes einzelne Integral immerhin noch stückweise stetig.

Lemma 3.6 *Zu jedem Quader $Q \subset \mathbb{R}^n$ und jedem $\epsilon > 0$ gibt es eine stetige Funktion $f \in C_c(\mathbb{R}^n, \mathbb{R})$ mit kompaktem Träger, die $0 \leq f \leq 1$, $f(x) = 1$ für $x \in Q$ und $\int_{\mathbb{R}^n} f(x) - 1_Q(x)\, dx \leq \epsilon$ erfüllt.*

Beweis: Nach der Vorbemerkung zeigen wir den Satz nur für abgeschlossene Quader. Sei $Q := [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$. Zunächst einmal gibt es zu jedem $\epsilon > 0$ und jedem Intervall $[a_i, b_i]$ eine Funktion $\phi_i \in C_c(\mathbb{R}, \mathbb{R})$ mit $0 \leq \phi_i \leq 1$, $\phi_i(x) = 1$ für alle $x \in [a_i, b_i]$ und $\phi_i(x) = 0$ für alle $x \notin [a_i - \epsilon, b_i + \epsilon]$, z.B. indem man zusätzlich $\phi_i(x) := 1 - \frac{1}{\epsilon}(a_i - x)$ für $x \in [a_i - \epsilon, a_i]$ und $\phi_i(x) := 1 - \frac{1}{\epsilon}(x - b_i)$ für $x \in [b_i, b_i + \epsilon]$ definiert.

Mit diesen Funktionen kann man dann $f(x) := \prod\limits_{i=1}^{n} \phi_i(x_i)$ definieren und erhält so eine Funktion $f \in C_c(\mathbb{R}^n, \mathbb{R})$ mit $0 \leq f \leq 1$ und $f(x) = 1$ für $x \in Q$. Darüberhinaus gilt

$$\int_{\mathbb{R}^n} f(x) - 1_Q(x)\, dx = \prod_{i=1}^{n} \left(\int_{a_i - \epsilon}^{a_i} \phi_i(x_i)\, dx_i + \int_{b_i}^{b_i + \epsilon} \phi_i(x_i)\, dx_i \right) \leq (2\epsilon)^n$$

Ersetzt man also das anfängliche ϵ durch $\sqrt[n]{\epsilon}/2$, so folgt die Aussage des Lemmas. □

Aufgabe: *Beweisen Sie, dass es zu jedem Quader $Q \subset \mathbb{R}^n$ und jedem $\epsilon > 0$ sogar eine C_c^∞-Funktion f gibt mit $0 \leq f \leq 1$, $f(x) = 1$ für $x \in Q$ und $\int_{\mathbb{R}^n} f(x) - 1_Q(x)\, dx \leq \epsilon$.*

Insbesondere kann man auch jede Treppenfunktion zu Quadern bzgl. der L^1-Norm beliebig genau annähern.

Definition 3.7 *Eine Funktion $g : \mathbb{R}^n \to \mathbb{R}$ heißt Treppenfunktion zu Quadern Q_i, $i = 1, \ldots, k$, wenn die Quader Q_i disjunkt sind, g auf Q_i den konstanten Wert g_i hat und außerhalb von $\bigcup_{i=1}^k Q_i$ den Wert Null annimmt. Insbesondere gilt dann mit den charakteristischen Funktionen 1_{Q_i} dieser Quader die Gleichung $g = \sum_{i=1}^k g_i 1_{Q_i}$.*

Nähert nun f_i den Quader Q_i im Sinne von Lemma 3.6 mit $\epsilon/|g_i|$ statt ϵ an, dann nähert die stetige Funktion $f := \sum_{i=1}^k g_i f_i$ mit kompaktem Träger die Treppenfunktion g im

Sinne der L^1-Norm an, denn

$$\int_{\mathbb{R}^n} |f(x) - g(x)|\, dx \leq \sum_{i=1}^{k} |g_i| \int_{\mathbb{R}^n} |f_i(x) - 1_{Q_i}(x)|\, dx \leq \sum_{i=1}^{k} |g_i| \frac{\epsilon}{|g_i|} = \epsilon \quad .$$

Umgekehrt kann man natürlich auch jede stetige Funktion f mit kompaktem Träger beliebig genau durch Treppenfunktionen zu Quadern annähern. Sei Q nämlich ein (halboffener) Quader, der $\mathrm{Tr}(f)$ enthält, dann ist f als stetige Funktion auf der kompakten Menge Q gleichmäßig stetig. Also gibt es zu $\epsilon > 0$ ein $\delta > 0$ mit $\|x - y\|_{\mathbb{R}^n} \leq \delta \Rightarrow |f(x) - f(y)| \leq \epsilon / \operatorname{Vol}(Q)$. Zerlege nun Q in endlich viele (halboffene) disjunkte Quader Q_i mit nichtleerem Inneren der maximalen Seitenlänge δ. Wähle im Inneren jedes dieser Quader Q_i einen Punkt x_i und betrachte mit $g_i := f(x_i)$ die Treppenfunktion $g := \sum_{i=1}^{k} g_i 1_{Q_i}$. Dann gilt $|f(x) - g(x)| \leq |f(x) - f(x_i)| \leq \epsilon / \operatorname{Vol}(Q)$ für jeden Punkt x im Inneren eines Quaders Q_i und somit

$$\int_{\mathbb{R}^n} |f(x) - g(x)|\, dx = \sum_i \int_{Q_i} |f(x) - f(x_i)|\, dx \leq \sum_i \frac{\epsilon}{\operatorname{Vol}(Q)} \operatorname{Vol}(Q_i) = \epsilon \quad .$$

Also ist die Vervollständigung von $C_c(\mathbb{R}^n, \mathbb{R})$ bzgl. $\|\cdot\|_1$ identisch mit der Vervollständigung des Raumes der Treppenfunktionen $g = \sum_{i=1}^{k} g_i 1_{Q_i}$ zu Quadern Q_i bzgl. der (Halb-)Norm $\|g\|_1 := \sum_{i=1}^{k} |g_i| \operatorname{Vol}(Q_i)$, die wegen

$$\int_{\mathbb{R}^n} g(x)\, dx = \sum_{i=1}^{k} g_i \int_{\mathbb{R}^n} 1_{Q_i}(x)\, dx = \sum_{i=1}^{k} g_i \operatorname{Vol}(Q_i)$$

gerade dem iterierte Riemann-Integral der Treppenfunktion $|g|$ entspricht.

Stetigkeit des Integrals Das Integral $I : C_c(\mathbb{R}^n, \mathbb{R}) \to \mathbb{R}$ und auch jedes andere monotone lineare Funktional I auf $C_c(\mathbb{R}^n, \mathbb{R})$ ist im folgenden Sinne stetig.

Satz 3.8 *Sei $I : C_c(\mathbb{R}^n, \mathbb{R}) \to \mathbb{R}$ linear und monoton. Ist $f_k \in C_c(\mathbb{R}^n, \mathbb{R})$ eine Folge von Funktionen, deren Träger $\mathrm{Tr}(f_k)$ in einem gemeinsamen Kompaktum K enthalten sind und die auf K gleichmäßig konvergieren, dann gilt $\lim\limits_{k\to\infty} I(f_k) = I\left(\lim\limits_{k\to\infty} f_k\right)$.*

Beweis: Die kompakte Menge K ist beschränkt, also ist sie insbesondere in einem Quader Q enthalten. Nach Lemma 3.6 gibt es insbesondere eine Funktion $g \in C_c(\mathbb{R}^n, \mathbb{R})$ mit $0 \leq g \leq 1$ und $g(x) = 1$ für alle $x \in Q$.

Sei nun f der gleichmäßige Grenzwert der f_k, dann gilt wegen $\mathrm{Tr}(f_k) \subset K$ auch $\mathrm{Tr}(f) \subset K$, also ist f eine stetige Funktion mit kompaktem Träger in K. Somit gilt auf Q die Ungleichung

$$-\|f_k - f\|_\infty g \leq f_k - f \leq \|f_k - f\|_\infty g \quad .$$

Daher gilt wegen der Monotonie des Integrals auch

$$-\|f_k - f\|_\infty I(g) \leq I(f_k - f) \leq \|f_k - f\|_\infty I(g)$$

und somit $|I(f_k) - I(f)| \leq I(g)\|f_k - f\|_\infty$. Also konvergiert auch $I(f_k)$ gegen $I(f)$. $\square$

Gleichäßige Konvergenz ist jedoch für viele Anwendungen eine viel zu starke Forderung. Glücklicherweise wird die schon angekündigte Fortsetzung des Integrals auf die Vervollständigung unter viel geringeren Voraussetzungen mit der Limesbildung vertauschen.

Integration von Banach-Raum-wertigen Funktionen Auch wenn wir uns hauptsächlich für die Integration von reellwertigen Funktionen interessieren, lohnt es sich anzumerken, dass das im Beweis von Satz 2.13 eingeführte Riemann-Integral für Funktionen auf $[a, b]$ mit Werten in einem Banach-Raum Y es uns erlaubt, auch das iterierte Riemann-Integral $I_Y(f) := \int_{\mathbb{R}^n} f(x)\,dx$ von stetigen Funktionen $f : \mathbb{R}^n \to Y$ mit kompaktem Träger zu definieren. Denn Satz 1.66 über die stetige Abhängigkeit parameterabhängiger Integrale gilt ebenso, im Beweis muß man nur den reellen Betrag durch die Norm auf Y ersetzen.

Das so definierte Integral $I_Y : C_c(\mathbb{R}^n, Y) \to Y$ ist weiterhin linear, stetig und translationsinvariant im Sinne der Sätze 3.4(i), 3.8 und 3.5, und an Stelle der Monotonie gilt die Ungleichung $\|I_Y(f)\|_Y \leq I(\|f\|_Y)$, wobei auf der rechten Seite das Integral I für reellwertige Funktionen gemeint ist.

Aufgabe:

- *Zeigen Sie, dass für das reellwertige Integral $I : C_c(\mathbb{R}^n, \mathbb{R}) \to \mathbb{R}$ Monotonie äquivalent zu $|I(f)| \leq I(|f|)$ ist.*

- *Sei Y ein Banach-Raum und f_k eine Cauchy-Folge in $C_c(\mathbb{R}^n, Y)$ bzgl. der L^1-Norm $\|f\|_1 := \int_{\mathbb{R}^n} \|f(x)\|_Y\,dx$. Zeigen Sie, dass dann auch die Integrale $\int_{\mathbb{R}^n} f_k(x)\,dx$ eine Cauchy-Folge in Y bilden und somit gegen einen Vektor in Y konvergieren.*

Wir halten diese Ergebnisse über das Y-wertige iterierte Riemann-Integral im folgenden Satz fest.

Satz 3.9 *Sei Y ein Banach-Raum, dann ist das Y-wertige iterierte Riemann-Integral $I_Y : C_c(\mathbb{R}^n, Y) \to Y$ linear, stetig, translationsinvariant und erfüllt $\|I_Y(f)\|_Y \leq I(\|f\|_Y)$ mit dem reellwertigen iterierten Riemann-Integral I.*

Im Fall $Y = \mathbb{R}^k$ kann man das Y-wertige iterierte Integral komponentenweise ausrechnen, da dies auch für das Riemann-Integral von Funktionen auf einem Intervall mit Werten in Y galt.

Eine sinnvolle Anwendung des $\mathbb{R}^n$-wertigen Integrals ist die Berechnung des Schwerpunktes eines Körpers mit stetiger Massendichte. Solch einen Körper kann man durch eine kompakte Teilmenge $K \subset \mathbb{R}^n$ modellieren, und die stetige Massendichte durch eine stetige Funktion $\rho : \mathbb{R}^n \to \mathbb{R}$, die außerhalb von K verschwindet. Den Punkt $\frac{1}{\int_{\mathbb{R}^n} \rho(x)\,dx} \int_{\mathbb{R}^n} \rho(x)x\,dx$ im $\mathbb{R}^n$ bezeichnet man dann als den Schwerpunkt des Körpers K mit stetiger Massendichte ρ.

Beispiel 3.10 *Betrachtet man ein Quadrat $K := [-1,1]\times[-1,1]$ und die Massendichte $\rho(x,y) := 2 - |x| - |y|$ für $(x,y) \in K$ und $\rho(x,y) := 0$ für $(x,y) \notin K$, dann ist ρ stetig mit der kompakten Teilmenge K als Träger, und – wie zu erwarten war – ist der Schwerpunkt wegen*

$$\int_{\mathbb{R}^n} \rho(x,y)\begin{pmatrix} x \\ y \end{pmatrix} d(x,y) =$$
$$\begin{pmatrix} \int_{-1}^0\int_{-1}^0 (2+x+y)x\,dx\,dy + \int_{-1}^0\int_0^1 (2-x+y)x\,dx\,dy \\ \int_{-1}^0\int_{-1}^0 (2+x+y)y\,dx\,dy + \int_{-1}^0\int_0^1 (2-x+y)y\,dx\,dy \end{pmatrix} +$$
$$\begin{pmatrix} \int_0^1\int_{-1}^0 (2+x-y)x\,dx\,dy + \int_0^1\int_0^1 (2-x-y)x\,dx\,dy \\ \int_0^1\int_{-1}^0 (2+x-y)y\,dx\,dy + \int_0^1\int_0^1 (2-x-y)y\,dx\,dy \end{pmatrix} =$$
$$\begin{pmatrix} \int_{-1}^0 -1+\frac{1}{3}-\frac{1}{2}y\,dy + \int_{-1}^0 1-\frac{1}{3}+\frac{1}{2}y\,dy \\ \int_{-1}^0 (2y+y^2)-\frac{1}{2}y\,dy + \int_{-1}^0 (2y+y^2)-\frac{1}{2}y\,dy \end{pmatrix} +$$
$$\begin{pmatrix} \int_0^1 -1+\frac{1}{3}+\frac{1}{2}y\,dy + \int_0^1 1-\frac{1}{3}-\frac{1}{2}y\,dy \\ \int_0^1 (2y-y^2)-\frac{1}{2}y\,dy + \int_0^1 (2y-y^2)-\frac{1}{2}y\,dy \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

der Nullpunkt. Man beachte, dass wir mit unseren bisherigen Mitteln beispielsweise noch nicht in der Lage sind zu zeigen, dass der Schwerpunkt eines Dreiecks K in der Euklidischen Ebene $\mathbb{R}^2$ mit den Eckpunkten $x,y,z \in \mathbb{R}^2$ bzgl. der Massendichte $\rho(x) := 1$ für $x \in K$ und $\rho(x) := 0$ für $x \notin K$ der Punkt $\frac{1}{3}(x+y+z)$ ist, denn die Funktion $\rho : \mathbb{R}^n \to \mathbb{R}$ ist unstetig, da sie am Rand ∂K von 0 auf 1 springt, und die Sprungmengen sind auch nicht nur Ränder von Quadern.

3.2 Maßtheorie

Im vorigen Abschnitt haben wir gesehen, dass man das Integral einer Treppenfunktion $g = \sum_{i=1}^k g_i 1_{Q_i}$ zu Quadern Q_i relativ einfach mit Hilfe des Elementarvolumens $\mathrm{Vol}(Q_i)$ der (offenen, halboffenen oder abgeschlossenen) Quader $Q_i \subset \mathbb{R}^n$ durch $\int_{\mathbb{R}^n} g(x)\,dx := \sum_{i=1}^k g_i \,\mathrm{Vol}(Q_i)$ definieren kann.

Die Vervollständigung des Raumes der Treppenfunktionen konstruieren wir daher explizit in zwei Teilen. In diesem Abschnitt setzen wir das bislang nur für Quader definierte Elementarvolumen Vol auf eine viel größere Menge von Teilmengen des $\mathbb{R}^n$ fort, die Menge der sogenannten Lebesgue-meßbaren Teilmengen des $\mathbb{R}^n$. Im folgenden Abschnitt erklären wir dann, wie man das Integral für Treppenfunktionen zu solchen meßbaren Mengen auf die viel größere Menge der integrierbaren Funktionen fortsetzt.

Da es keinen zusätzlichen Aufwand bereitet, formulieren wir die folgenden Resultate statt nur für das Elementarvolumen Vol auf halboffenen Quadern im $\mathbb{R}^n$ gleich für allgemeine Prämaße auf einem Halbring von Teilmengen einer Menge Ω.

Inhalte auf Halbringen und Ringen Aus Bequemlichkeit gehen wir bei der Konstruktion der Fortsetzung des Elementarvolumens von halboffenen Quadern aus, d.h.

Quadern Q der Form $Q = (a_1, b_1] \times \cdots \times (a_n, b_n]$. Zusammen mit der leeren Menge bilden diese Quader einen Halbring von Teilmengen des $\mathbb{R}^n$.

Definition 3.11 *Eine Menge $\mathcal{H} \subset \mathcal{P}(\Omega)$ von Teilmengen einer Menge Ω heißt ein Halbring, falls*

- *$\emptyset \in \mathcal{H}$ gilt,*
- *für $A, B \in \mathcal{H}$ auch $A \cap B \in \mathcal{H}$ gilt,*
- *es für $A, B \in \mathcal{H}$ disjunkte $C_1, \ldots, C_m \in \mathcal{H}$ mit $A \setminus B = \bigcup\limits_{k=1}^{m} C_k$ gibt.*

Tatsächlich ist der Durchschnitt halboffener Quader wieder ein halboffener Quader oder leer, und entfernt man aus einem halboffenen Quader einen anderen halboffenen Quader, so kann man die übrigbleibende Menge als Vereinigung endlich vieler halboffener Quader schreiben. Dies folgt per Induktion aus der Tatsache, dass $(a, b] \setminus (c, d]$ leer oder ein halboffenes Intervall oder die Vereinigung von zwei halboffenen Intervallen ist, und dass

$$(A \times B) \setminus (C \times D) = ((A \setminus C) \times B) \cup ((A \cap C) \times (B \setminus D))$$

für das Produkt von Mengen gilt.

Das Elementarvolumen Vol ist nun auf dem Halbring aus der leeren Menge und den halboffenen Quadern ein Inhalt.

Definition 3.12 *Sei $\mathcal{H} \subset \mathcal{P}(\Omega)$ ein Halbring von Teilmengen einer Menge Ω. Eine Abbildung $\mu : \mathcal{H} \to [0, \infty]$ heißt Inhalt auf $\mathcal{H}$, wenn*

- *$\mu(\emptyset) = 0$ gilt,*
- *für **disjunkte** Mengen $A_1, \ldots, A_m \in \mathcal{H}$ mit $\bigcup\limits_{k=1}^{m} A_k \in \mathcal{H}$ die Gleichung $\mu\left(\bigcup\limits_{k=1}^{m} A_k\right) = \sum\limits_{k=1}^{m} \mu(A_k)$ gilt ("μ ist additiv").*

Zu dieser Definition sei angemerkt, dass es keinerlei Schwierigkeiten bereitet, ∞ als Wert von μ zuzulassen, da man die Addition in $[0, \infty)$ sinnvoll durch $a + \infty := \infty$, $\infty + \infty := \infty$ und auf $[0, \infty]$ fortsetzen kann.

Bevor wir nachweisen, dass das Elementarvolumen Vol auf dem Halbring aus der leeren Menge und den halboffenen Quadern ein Inhalt ist, seien noch zwei andere Beispiele für Inhalte genannt.

Beispiel 3.13

- *Für eine beliebige Menge Ω ist die Potenzmenge $\mathcal{H} := \mathcal{P}(\Omega)$ ein Halbring. Sei $\mu(A)$ für eine endliche Teilmenge A von Ω die Anzahl der Elemente in A und ansonsten $\mu(A) := \infty$. Dann ist μ ein Inhalt auf $\mathcal{H}$.*

- *Sei Ω abzählbar unendlich und $\mathcal{H}$ bestehe aus den Teilmengen von Ω, die entweder endlich sind oder deren Komplement endlich ist. Dann ist $\mathcal{H}$ ein Halbring, und durch $\mu(A) := 0$ für endliche A bzw. $\mu(A) := \infty$ für A, deren Komplement $\Omega \setminus A$ endlich ist, kann man einen Inhalt μ auf $\mathcal{H}$ definieren.*

Satz 3.14 *Das Elementarvolumen* Vol *ist ein Inhalt auf dem Halbring aus der leeren Menge und den halboffenen Quadern im $\mathbb{R}^n$.*

Beweis:

- Ist $Q = (a_1, b_1] \times \cdots \times (a_n, b_n]$ und wird Q durch die Hyperebene $\{x \mid x_\nu = \alpha\}$ $(a_\nu < \alpha < b_\nu)$ in die zwei disjunkten Quader $Q_1 = (a_1, b_1] \times \cdots \times (a_\nu, \alpha] \times \cdots \times (a_n, b_n]$ und $Q_2 = (a_1, b_1] \times \cdots \times (\alpha, b_\nu] \times \cdots \times (a_n, b_n]$ zerlegt, so gilt $Q = Q_1 \cup Q_2$ und offensichtlich $\mathrm{Vol}(Q) = \mathrm{Vol}(Q_1) + \mathrm{Vol}(Q_2)$ wegen $(\alpha - a_\nu) + (b_\nu - \alpha) = b_\nu - a_\nu$. Per Induktion folgt, dass auch bei einer Zerlegung von Q in $Q_1, \ldots, Q_m$ durch m Hyperebenen $\mathrm{Vol}(Q) = \sum\limits_{k=1}^{m} \mathrm{Vol}(Q_k)$ folgt.

- Sei nun $Q = \bigcup\limits_{k=1}^{m} Q_k$ eine Zerlegung von Q in disjunkte halboffene Quader Q_k. Deren Randflächen liegen in gewissen Hyperebenen, und diese zerteilen Q in disjunkte halboffene Quader Q_{kl}, wobei die Nummerierung so gewählt sei, dass $Q_k = \bigcup\limits_{l=1}^{m_k} Q_{kl}$ gilt. Da sowohl die Zerlegung von Q als auch die jedes Q_k eine durch Hyperflächen induzierte ist und wir für solche Zerlegungen im ersten Punkt schon die Additivität von Vol nachgewiesen haben, folgt die Additivität

$$\mathrm{Vol}(Q) = \sum_{k=1}^{m} \sum_{l=1}^{m_k} \mathrm{Vol}(Q_{kl}) = \sum_{k=1}^{m} \mathrm{Vol}(Q_k) \quad .$$

$\square$

Eine wichtige Eigenschaft von Inhalten ist ihre Monotonie.

Lemma 3.15 *Jeder Inhalt μ auf einem Halbring $\mathcal{H}$ ist monoton, d.h. gilt für $A, B \in \mathcal{H}$ die Inklusion $A \subset B$, so gilt auch $\mu(A) \leq \mu(B)$.*

Beweis: Seien $A, B \in \mathcal{H}$ Mengen mit $A \subset B$. Da $\mathcal{H}$ ein Halbring ist, gibt es disjunkte $C_1, \ldots, C_m$ mit $B \setminus A = \bigcup\limits_{k=1}^{m} C_k$. Also ist $B = A \cup \bigcup_{k=1}^{m} C_k$ eine disjunkte Vereinigung und daher gilt $\mu(B) = \mu(A) + \sum\limits_{k=1}^{m} \mu(C_k) \geq \mu(A)$. $\square$

Wie angekündigt wollen wir nun das Elementarvolumen auf eine größere Menge von Teilmengen fortsetzen. Dazu beweisen wir zunächst, dass jeder Halbring von Teilmengen einen Ring erzeugt, und sich jeder Inhalt eindeutig auf diesen Ring fortsetzen läßt. Neben den üblichen Operationen $\cap$, $\cup$, $\setminus$ auf Mengen verwenden wir dabei zusätzlich die symmetrische Differenz

$$A \Delta B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B) \quad .$$

Definition 3.16 *Eine Menge $\mathcal{R} \subset \mathcal{P}(\Omega)$ von Teilmengen einer Menge Ω heißt ein Ring, falls*

- *$\emptyset \in \mathcal{R}$ gilt,*
- *für $A, B \in \mathcal{R}$ auch $A \Delta B \in \mathcal{R}$ gilt,*
- *für $A, B \in \mathcal{R}$ auch $A \cap B \in \mathcal{R}$ gilt.*

Gilt zusätzlich noch $\Omega \in \mathcal{R}$, dann nennt man $\mathcal{R}$ eine Algebra von Teilmengen von Ω.

Man bemerke, dass die Potenzmenge $\mathcal{P}(\Omega)$ zusammen mit Δ als Addition und $\cap$ als Multiplikation im algebraischen Sinne ein Ring ist (Assoziativ-, Kommutativ- und Distributivgesetze gelten, und das additive Inverse von $A \subset \Omega$ ist A selbst), in dem $\emptyset$ das neutrale Element der Addition und Ω das neutrale Element der Multiplikation ist. Ein wie oben definierter Ring ist dann nichts anderes als ein Unterring von $\mathcal{P}(\Omega)$, und eine wie oben definierte Algebra nichts anderes als ein Unterring von $\mathcal{P}(\Omega)$, der das neutrale Element der Multiplikation enthält.

Aus $A \cup B = (A \Delta B) \Delta (A \cap B)$ und $A \setminus B = A \Delta (A \cap B)$ ergibt sich sofort das folgende Lemma.

Lemma 3.17 *In einem Ring $\mathcal{R}$ von Teilmengen von Ω gilt für $A, B \in \mathcal{R}$ auch $A \cup B \in \mathcal{R}$ und $A \setminus B \in \mathcal{R}$,*

Aufgabe: *Bezeichne $A^c := \Omega \setminus A$ das Komplement von Teilmengen $A \subset \Omega$. Eine Menge $\mathcal{R} \subset \mathcal{P}(\Omega)$ von Teilmengen von Ω ist genau dann eine Algebra, wenn $\Omega \in \mathcal{R}$ gilt und mit $A, B \in \mathcal{R}$ auch $A^c \in \mathcal{R}$ und $A \cup B \in \mathcal{R}$ gilt.*

Nach Lemma 3.17 ist jeder Ring ein Halbring, denn für $A, B \in \mathcal{R}$ liegt sowohl $A \cap B$ als auch $A \setminus B$ selbst in $\mathcal{R}$. Umgekehrt erzeugt jeder Halbring $\mathcal{H}$ einen Ring, man bilde nämlich einfach den Durchschnitt aller $\mathcal{H}$ enthaltenden Ringe in $\mathcal{P}(\Omega)$. Diesen kleinsten $\mathcal{H}$ enthaltenden Ring $\mathcal{R}$ kann man auch ganz konkret angeben.

Lemma 3.18 *Der von einem Halbring $\mathcal{H}$ erzeugte Ring $\mathcal{R}$ ist*

$$\mathcal{R} = \{\bigcup_{k=1}^{m} A_k \mid A_1, \ldots, A_m \in \mathcal{H} \textit{ disjunkt}\} \quad .$$

Beweis: Da Ringe unter endlichen Vereinigungen abgeschlossen sind, muß der von $\mathcal{H}$ erzeugte Ring die Mengen $\bigcup_{k=1}^{m} A_k$ für disjunkte $A_1, \ldots, A_m \in \mathcal{H}$ enthalten. Daher reicht es zu zeigen, dass $\{\bigcup_{k=1}^{m} A_k \mid A_1, \ldots, A_m \in \mathcal{H} \text{ disjunkt}\}$ bereits ein Ring ist.

Offensichtlich enthält diese Menge wegen $\emptyset \in \mathcal{H}$ die leere Menge, und sind $A = \bigcup_{k=1}^{m} A_k$ sowie $B = \bigcup_{l=1}^{n} B_l$ endliche disjunkte Vereinigungen von $A_1, \ldots, A_m \in \mathcal{H}$ bzw. $B_1, \ldots, B_n \in$

$\mathcal{H}$, dann ist auch $A \cap B = \bigcup_{\substack{k=1,\ldots,m \\ l=1,\ldots,n}} A_k \cap B_l$ eine endliche disjunkte Vereinigung der Mengen $A_k \cap B_l \in \mathcal{H}$.

Ebenso ist

$$A \setminus B = \bigcup_{k=1}^{m} \left(A_k \setminus \left(\bigcup_{l=1}^{n} B_l \right) \right)$$

eine endliche disjunkte Vereinigung der Mengen $A_k \setminus \left(\bigcup_{l=1}^{n} B_l \right)$, die ihrerseits wieder endliche disjunkte Vereinigungen von Mengen aus $\mathcal{H}$ sind, wie man durch Induktion nach n beweist [1]. Also ist auch $A \Delta B = (A \setminus B) \cup (B \setminus A)$ eine endliche disjunkte Vereinigung von Mengen aus $\mathcal{H}$. Somit ist $\{ \bigcup_{k=1}^{m} A_k \mid A_1, \ldots, A_m \in \mathcal{H}$ disjunkt$\}$ ein Ring.

$\square$

Der Begriff des Inhalts macht offensichtlich auch für auf Ringen definierte Funktionen $\mu : \mathcal{R} \to [0, \infty]$ Sinn. So sind die in Beispiel 3.13 genannten Inhalte sogar auf Ringen definiert. Für die Additivität eines Inhaltes auf einem Ring muß man auch gar nicht mehr darauf achten, ob die Vereinigung disjunkter $A_1, \ldots, A_m$ wieder in $\mathcal{R}$ liegt, dies ist automatisch der Fall. Ein Inhalt auf einem Ring ist also einfach eine Abbildung $\mu : \mathcal{R} \to [0, \infty]$ mit $\mu(\emptyset) = 0$ und $\mu(A \cup B) = \mu(A) + \mu(B)$ für **disjunkte** $A, B \in \mathcal{R}$. Neben der Monotonie lassen sich dann leicht noch weitere Eigenschaften herleiten:

Lemma 3.19 *Jeder Inhalt μ auf einem Ring $\mathcal{R}$ erfüllt*

- $\mu(A \setminus B) = \mu(A) - \mu(B)$ *bei* $B \subset A$ *und* $\mu(B) < \infty$,
- $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$,
- $\mu(\bigcup_{k=1}^{m} A_k) \leq \sum_{k=1}^{m} \mu(A_k)$ *für (nicht notwendig disjunkte)* $A_1, \ldots, A_m \in \mathcal{R}$ *("μ ist subadditiv").*

Beweis:

- Wegen $B \subset A$ ist $A = B \cup (A \setminus B)$, und da dies eine disjunkte Vereinigung ist, gilt $\mu(A) = \mu(B) + \mu(A \setminus B)$. Da außerdem $\mu(B) < \infty$ gilt, folgt daraus $\mu(A \setminus B) = \mu(A) - \mu(B)$.

---

[1] Ist $A \setminus \left(\bigcup_{l=1}^{n} B_l \right)$ endliche disjunkte Vereinigung der Mengen $C_1, \ldots, C_p \in \mathcal{H}$, dann auch

$$A \setminus \left(\bigcup_{l=1}^{n+1} B_l \right) = \left(A \setminus \left(\bigcup_{l=1}^{n} B_l \right) \right) \setminus B_{n+1} = \left(\bigcup_{i=1}^{p} C_i \right) \setminus B_{n+1} = \bigcup_{i=1}^{p} C_i \setminus B_{n+1} \quad ,$$

, wobei sich die Mengen $C_i \setminus B_{n+1}$ wiederum als endliche disjunkte Vereinigungen von Mengen in $\mathcal{H}$ schreiben lassen.

- Die Zerlegungen $B = (B \setminus A) \cup (A \cap B)$ und $A \cup B = A \cup (B \setminus A)$ sind disjunkt, also gilt
$$\mu(A) + \mu(B) = \mu(A) + \mu(B \setminus A) + \mu(A \cap B) = \mu(A \cup B) + \mu(A \cap B)$$
- Da $\bigcup\limits_{k=1}^{m} A_k = \bigcup\limits_{k=1}^{m} \left(A_k \setminus \left(\bigcup\limits_{l=1}^{k-1} A_l \right) \right)$ eine endliche disjunkte Vereinigung ist, folgt mit Hilfe der Monotonie von μ
$$\mu(\bigcup_{k=1}^{m} A_k) = \sum_{k=1}^{m} \mu \left(A_k \setminus \left(\bigcup_{l=1}^{k-1} A_l \right) \right) \leq \sum_{k=1}^{m} \mu(A_k) \quad .$$

□

Der folgende Satz besagt, dass sich jeder auf einem Halbring $\mathcal{H}$ definierte Inhalt eindeutig auf den von $\mathcal{H}$ erzeugten Ring fortsetzen läßt.

Satz 3.20 *Für einen Inhalt μ auf einem Halbring $\mathcal{H}$ gibt es genau eine Fortsetzung ν zu einem Inhalt auf dem von $\mathcal{H}$ erzeugten Ring $\mathcal{R}$. Diese ist durch $\nu(A) := \sum\limits_{k=1}^{m} \mu(A_k)$ für disjunkte $A_1, \ldots, A_m \in \mathcal{H}$ mit $A = \bigcup\limits_{k=1}^{m} A_k$ gegeben.*

Beweis: Da $\mathcal{R}$ nach Lemma 3.18 gerade aus den endlichen disjunkten Vereinigungen von Mengen in $\mathcal{H}$ besteht, muß jeder Inhalt ν auf $\mathcal{R}$, der μ fortsetzt, die Gleichung $\nu(A) := \sum\limits_{k=1}^{m} \mu(A_k)$ für disjunkte $A_1, \ldots, A_m \in \mathcal{H}$ mit $A = \bigcup\limits_{k=1}^{m} A_k$ erfüllen und ist daher eindeutig bestimmt.

Nachzuweisen bleibt noch, dass durch $\nu(A) := \sum\limits_{k=1}^{m} \mu(A_k)$ für disjunkte $A_1, \ldots, A_m \in \mathcal{H}$ mit $A = \bigcup\limits_{k=1}^{m} A_k$ wirklich ein Inhalt auf $\mathcal{R}$ definiert. Trivialerweise ist ν additiv, also ist nur noch die Wohldefiniertheit zu zeigen: Sind $A = \bigcup\limits_{k=1}^{m} A_k$ und $A = \bigcup\limits_{l=1}^{n} B_l$ zwei Zerlegungen von A in disjunkte Mengen aus $\mathcal{H}$, dann gilt
$$\sum_{k=1}^{m} \mu(A_k) = \sum_{k=1}^{m} \mu \left(A_k \cap \bigcup_{l=1}^{n} B_l \right) = \sum_{k=1}^{m} \mu \left(\bigcup_{l=1}^{n} A_k \cap B_l \right) = \sum_{k=1}^{m} \sum_{l=1}^{n} \mu \left(A_k \cap B_l \right) =$$
$$\sum_{l=1}^{n} \mu \left(\bigcup_{k=1}^{m} A_k \cap B_l \right) = \sum_{l=1}^{n} \mu \left(B_l \cap \bigcup_{k=1}^{m} A_k \right) = \sum_{l=1}^{n} \mu(B_l) \quad ,$$
also ist ν wohldefiniert. □

Insbesondere läßt sich das Elementarvolumen zu einem Inhalt auf dem Ring aus der leeren Menge und den endlichen disjunkten Vereinigungen von halboffenen Quadern fortsetzen. Üblicherweise bezeichnet man diese durch $\mathrm{Vol}(\bigcup\limits_{k=1}^{m} Q_k) := \sum\limits_{k=1}^{m} \mathrm{Vol}(Q_k)$ definierte Fortsetzung wiederum mit Vol, und im wesentlichen hat der vorige Satz nur gezeigt, dass diese Fortsetzung wohldefiniert ist.

Prämaße und σ-Algebren Die im vorigen Abschnitt definierte Fortsetzung des Elementarvolumens auf endliche disjunkte Vereinigungen von halboffenen Quadern reicht bei weitem nicht aus, z.B. können wir mit ihr immer noch nicht das Volumen von beliebigen offenen Teilmengen des $\mathbb{R}^n$ bestimmen. Man beachte aber, dass sich jede offene Teilmenge des $\mathbb{R}^n$ als abzählbare Vereinigung von halboffenen Quadern schreiben läßt, beispielsweise dadurch, dass man um jeden Punkt der offenen Menge mit rationalen Koordinaten eine halboffenen Würfel mit einer so kleinen positiven Seitenlänge wählt, dass der Würfel immer noch komplett in der offenen Menge enthalten ist. Dies läßt den Wunsch aufkommen, das Elementarvolumen auf die von den halboffenen Quadern erzeugte σ-Algebra fortsetzen zu können.

Definition 3.21 *Eine Ring $\mathcal{R} \subset \mathcal{P}(\Omega)$ von Teilmengen einer Menge Ω heißt eine σ-Ring, falls zusätzlich für jede Folge $A_k \in \mathcal{R}$ auch die abzählbar unendliche Vereinigung $\bigcup\limits_{k=1}^{\infty} A_k$ in $\mathcal{R}$ liegt.*

Ein σ-Ring $\mathcal{A}$ mit $\Omega \in \mathcal{A}$ wird σ-Algebra genannt.

Lemma 3.22 *Ist $\mathcal{R}$ ein σ-Ring (oder sogar eine σ-Algebra), dann liegt für jede Folge $A_k \in \mathcal{R}$ auch der abzählbar unendliche Durchschnitt $\bigcap\limits_{k=1}^{\infty} A_k$ in $\mathcal{R}$.*

Beweis: Es gilt mit $A := \bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{R}$ die Beziehung

$$\bigcap_{k=1}^{\infty} A_k = A \setminus \left(\bigcup_{k=1}^{\infty} (A \setminus A_k) \right) \in \mathcal{R}$$

□

Ähnlich wie in Aufgabe 3.2 kann man auch σ-Algebren charakterisieren.

Aufgabe: *Eine Menge $\mathcal{A} \subset \mathcal{P}(\Omega)$ von Teilmengen von Ω ist genau dann eine σ-Algebra, wenn $\Omega \in \mathcal{A}$ gilt, mit $A \in \mathcal{A}$ auch $A^c \in \mathcal{A}$ liegt, und mit jeder Folge $A_k \in \mathcal{A}$ auch die abzählbar unendliche Vereinigung $\bigcup\limits_{k=1}^{\infty} A_k$ in $\mathcal{A}$ liegt.*

Jeder Ring $\mathcal{R}$ erzeugt eine σ-Algebra, man bilde einfach den Durchschnitt aller $\mathcal{R}$ enthaltenden σ-Algebren in $\mathcal{P}(\Omega)$. Diese kleinste den Ring $\mathcal{R}$ enthaltende σ-Algebra bezeichnet man üblicherweise mit $\sigma(\mathcal{R})$. Man kann sie aber nicht mehr so elementar beschreiben, wie dies noch in Lemma 3.18 für den von einem Halbring erzeugten Ring möglich war.

Definition 3.23 *Die vom Ring der disjunkten Vereinigungen von halboffenen Quadern im $\mathbb{R}^n$ erzeugte σ-Algebra, die nach der Vorbemerkung dieselbe ist wie die von den offenen Teilmengen des $\mathbb{R}^n$ erzeugte σ-Algebra, heißt die Borel-σ-Algebra von Teilmengen des $\mathbb{R}^n$ und wird üblicherweise mit $\mathcal{B}(\mathbb{R}^n)$ bezeichnet.*

Man bemerke, dass $\mathcal{B}$ nicht nur die offenen Teilmengen des $\mathbb{R}^n$ enthält, sondern als σ-Algebra auch deren Komplemente, also sind auch alle abgeschlossenen Teilmengen des $\mathbb{R}^n$ in $\mathcal{B}(\mathbb{R}^n)$ enthalten.

Will man nun das Elementarvolumen auf diese (oder eine noch größere) σ-Algebra fortsetzen, so muß man natürlich garantieren, dass bei Darstellung einer schon ursprünglich im Ring vorhandenen Menge durch eine abzählbare disjunkte Vereinigung dasselbe Volumen berechnet wird, oder mit anderen Worten, dass das Elementarvolumen ein Prämaß ist.

Definition 3.24 *Sei $\mathcal{H} \subset \mathcal{P}(\Omega)$ ein Halbring (oder Ring) von Teilmengen einer Menge Ω. Eine Abbildung $\mu : \mathcal{H} \to [0, \infty]$ heißt Prämaß auf $\mathcal{H}$, wenn*

- *$\mu(\emptyset) = 0$ gilt,*
- *für jede Folge **disjunkter** Mengen $A_k \in \mathcal{H}$ mit $\bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{H}$ die Gleichung $\mu\left(\bigcup\limits_{k=1}^{\infty} A_k\right) = \sum\limits_{k=1}^{\infty} \mu(A_k)$ gilt ("μ ist σ-additiv")*

Ein Prämaß auf einer σ-Algebra wird Maß genannt. Dabei ist für eine Folge disjunkter Mengen A_k in einer σ-Algebra automatisch auch $\bigcup\limits_{k=1}^{\infty} A_k$ in der σ-Algebra enthalten.

Zunächst bemerke man, dass es weiterhin kein Problem ist, den Wert ∞ für μ zuzulassen, indem man $\sum_{k=1}^{\infty} a_k := \infty$ für divergente Reihen von Zahlen $a_k \geq 0$ (bzw. falls eines der a_k gleich ∞ ist) setzt. Desweiteren beachte man, dass die σ-Additivität genannte Eigenschaft wesentlich stärker ist als die endliche Additivität, die man von einem Inhalt verlangt. Denn einerseits folgt die endliche Additivität aus der σ-Additivität, indem man $A_k := \emptyset$ für $k > m$ setzt. Andererseits gibt es Inhalte, die keine Prämaße sind.

Beispiel 3.25 *Der im ersten Punkt von Beispiel 3.13 definierte Inhalt auf der σ-Algebra $\mathcal{P}(\Omega)$ ist sogar ein Prämaß, denn ist $A = \bigcup\limits_{k=1}^{\infty} A_k$ eine disjunkte Vereinigung, dann sind entweder alle A_k bis auf endlich viele leer (und damit liegt eine endliche Vereinigung vor) oder A enthält unendlich viele Elemente, und dann gilt auch $\mu(A) = \sum\limits_{k=1}^{\infty} \mu(A_k) = \infty$.*

Dagegen ist der im zweiten Punkt von Beispiel 3.13 genannte Inhalt kein Prämaß, denn jede Teilmenge A der abzählbaren Menge Ω, die ein endliches Komplement hat, läßt sich als abzählbare Vereinigung endlicher Mengen A_k schreiben (z.B. als Vereinigung ihrer Punkte), was zu $\infty = \mu(A) \neq \sum\limits_{k=1}^{\infty} \mu(A_k) = 0$ führt.

Das Elementarvolumen Vol und auch dessen Fortsetzung auf den Ring der endlichen disjunkten Vereinigungen von halboffenen Quadern sind aber glücklicherweise Prämaße. Dies zeigen wir in zwei Schritten.

Lemma 3.26 *Ein Inhalt μ auf einem Halbring $\mathcal{H}$ ist genau dann ein Prämaß, wenn seine Fortsetzung auf den von $\mathcal{H}$ erzeugten Ring ein Prämaß ist.*

Beweis: Aus der σ-Additivität der Fortsetzung ν folgt natürlich auch die σ-Additivität von μ. Sei nun μ ein Prämaß auf dem Halbring $\mathcal{H}$, dann bleibt zu zeigen, dass der als Fortsetzung erhaltene Inhalt ν auf dem Ring $\mathcal{R}$ auch σ-additiv ist. Sei dazu $A_k \in \mathcal{R}$ eine Folge **disjunkter** Mengen mit $A := \bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{R}$. Dann gibt es disjunkte $B_1, \ldots, B_n \in \mathcal{H}$ mit $A = \bigcup\limits_{l=1}^{n} B_l$, und zu jedem k disjunkte $B_{k1}, \ldots, B_{kn_k} \in \mathcal{H}$ mit $A_k = \bigcup\limits_{l=1}^{n_k} B_{kl}$. Da

$$B_l = A \cap B_l = \bigcup_{k=1}^{\infty} A_k \cap B_l = \bigcup_{k=1}^{\infty} \bigcup_{l=1}^{n_k} B_{kl} \cap B_l$$

eine disjunkte Vereinigung von Mengen aus $\mathcal{H}$ ist und μ auf $\mathcal{H}$ ein Prämaß ist, folgt nach Definition von ν

$$\mu(B_l) = \sum_{k=1}^{\infty} \sum_{l=1}^{n_k} \mu(B_{kl} \cap B_l) = \sum_{k=1}^{\infty} \nu(A_k \cap B_l) \quad .$$

Somit ist ν auch σ-additiv wegen

$$\nu(A) = \sum_{l=1}^{n} \mu(B_l) = \sum_{l=1}^{n} \sum_{k=1}^{\infty} \nu(A_k \cap B_l) = \sum_{k=1}^{\infty} \sum_{l=1}^{n} \nu(A_k \cap B_l) = \sum_{k=1}^{\infty} \nu(A_k) \quad .$$

□

Bevor wir zeigen, dass das Elementarvolumen ein Prämaß ist, bemerke man noch kurz, dass aus der in 3.19 gezeigten Subadditivität von Inhalten μ auf Ringen $\mathcal{R}$ die Ungleichung

$$\sum_{k=1}^{\infty} \mu(A_k) \leq \mu\left(\bigcup_{k=1}^{\infty} A_k\right)$$

für eine Folge $A_k \in \mathcal{R}$ disjunkter Mengen mit $\bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{R}$ folgt. Tatsächlich, mit $A := \bigcup\limits_{k=1}^{\infty} A_k$ gilt $\bigcup\limits_{k=1}^{n} A_k \subset A$ und somit aufgrund der Additivität und Monotonie

$$\sum_{k=1}^{n} \mu(A_k) = \mu\left(\bigcup_{k=1}^{n} A_k\right) \leq \mu(A)$$

für jedes n. Durch Grenzübergang $n \to \infty$ folgt daher die zu beweisende Ungleichung.

Satz 3.27 *Das Elementarvolumen* Vol *ist ein Prämaß auf dem Halbring aus der leeren Menge und den halboffenen Quadern im $\mathbb{R}^n$ bzw. auf dem davon erzeugten Ring.*

Beweis: Bezeichne $\mathcal{H}$ den im Satz angesprochenen Halbring. Da Vol zu einem Inhalt auf dem von $\mathcal{H}$ erzeugten Ring fortgesetzt werden kann und somit die zuvor bewiesene Ungleichung $\sum\limits_{k=1}^{\infty} \mathrm{Vol}(Q_k) \leq \mathrm{Vol}\left(\bigcup\limits_{k=1}^{\infty} Q_k\right)$ insbesondere für jede Folge $Q_k \in \mathcal{H}$ disjunkter halboffener Quader mit $\bigcup\limits_{k=1}^{\infty} Q_k \in \mathcal{H}$ gilt, ist für die σ-Additivität nur noch die umgekehrte Ungleichung zu zeigen.

Diese umgekehrte Ungleichung beweisen wir durch ein Kompaktheitsargument: Sei der halboffene Quader $Q = (a_1, b_1] \times \cdots \times (a_n, b_n]$ in abzählbar viele halboffene disjunkte Quader $Q_k = (a_{1k}, b_{1k}] \times \cdots \times (a_{nk}, b_{nk}]$ zerlegt, d.h. $Q = \bigcup\limits_{k=1}^{\infty} Q_k$. Wir bezeichnen im folgenden Quader wie Q kurz durch $(a, b]$ mit Vektoren $a, b \in \mathbb{R}^n$ und nutzen die stetige Abhängigkeit von $\mathrm{Vol}((a, b])$ von den Randpunkten a, b aus.

Zu $\epsilon > 0$ wähle man einen Punkt $a' \in Q$ so nah an a, dass $\mathrm{Vol}(Q) \leq \mathrm{Vol}((a', b]) + \epsilon$ gilt. Für jedes $k \in \mathbb{N}$ wähle man außerdem eine Punkt b'_k mit $b_k \in (a_k, b'_k]$ so nah an b_k, dass $\mathrm{Vol}((a_k, b'_k]) \leq \mathrm{Vol}(Q_k) + 2^{-k}\epsilon$ gilt. Dann wird auch der abgeschlossene und daher kompakte Quader $[a', b]$ als Teilmenge von Q durch die Quader Q_k überdeckt, und daher ebenso durch die größeren offenen Quader (a_k, b'_k), in denen Q_k enthalten ist. Aufgrund der Kompaktheit reichen bei dieser Überdeckung aber endlich viele offene Quader aus. Es gibt also ein N, so dass $[a', b] \subset \bigcup\limits_{k=1}^{N} (a_k, b'_k)$ gilt, und daraus folgt auch $(a', b] \subset \bigcup\limits_{k=1}^{N} (a_k, b'_k]$.

Berechnen wir die Volumina für die letztgenannte Inklusion, so ergibt sich aufgrund der Subadditivität von Vol die Ungleichung $\mathrm{Vol}((a', b]) \leq \sum\limits_{k=1}^{N} \mathrm{Vol}((a_k, b'_k])$. Aufgrund der Wahl der Punkte a' und b' folgt daraus aber

$$\mathrm{Vol}((a, b]) \leq \mathrm{Vol}((a', b]) + \epsilon \leq \sum_{k=1}^{N} \mathrm{Vol}((a_k, b'_k]) + \epsilon \leq \sum_{k=1}^{N} \left(\mathrm{Vol}((a_k, b_k]) + 2^{-k}\epsilon\right) + \epsilon \leq \sum_{k=1}^{\infty} \mathrm{Vol}((a_k, b_k]) + 2\epsilon$$

Da dies für jedes $\epsilon > 0$ gilt, folgt $\mathrm{Vol}(Q) \leq \sum\limits_{k=1}^{\infty} \mathrm{Vol}(Q_k)$, d.h. die zu zeigende Ungleichung. $\square$

Zum Ausrechnen von Volumina kann man die beiden folgenden Resultate verwenden.

Aufgabe:

- *Beweisen Sie, dass ein Inhalt μ auf einem Ring $\mathcal{R}$ genau dann ein Prämaß ist, wenn für jede aufsteigende Folge $A_1 \subset A_2 \subset \ldots$ von Mengen $A_k \in \mathcal{R}$ mit $\bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{R}$ die Gleichung $\lim\limits_{k\to\infty} \mu(A_k) = \mu\left(\bigcup\limits_{k=1}^{\infty} A_k\right)$ gilt ("μ ist stetig von unten").*

- *Ist μ ein Prämaß auf dem Ring $\mathcal{R}$, dann gilt für jede Folge $A_1 \supset A_2 \supset \dots$ von Mengen $A_k \in \mathcal{R}$ mit $\mu(A_1) < \infty$ und $\bigcap\limits_{k=1}^{\infty} A_k \in \mathcal{R}$ die Gleichung $\lim\limits_{k\to\infty} \mu(A_k) = \mu\left(\bigcup\limits_{k=1}^{\infty} A_k\right)$ ("μ ist stetig von oben").*

Der Fortsetzungssatz von Carathéodory Da das Elementarvolumen ein Prämaß auf dem Halbring aus der leeren Menge und den halboffenen Quadern im $\mathbb{R}^n$ bzw. auf dessen erzeugten Ring ist, besteht zumindest die Chance, das Elementarvolumen zu einem eindeutigen Maß auf der von diesem Ring erzeugten σ-Algebra (oder einer noch größeren σ-Algebra) fortsetzen zu können. Der Fortsetzungssatz von Carathéodory, den wir in diesem Abschnitt beweisen werden, besagt nun gerade, dass solch eine Fortsetzung tatsächlich möglich ist.

Zur Konstruktion der Fortsetzung eines auf einem Halbring $\mathcal{H}$ von Teilmengen von Ω definierten Prämaßes $\mu : \mathcal{H} \to [0, \infty]$ betrachte man die Funktion $\mu^* : \mathcal{P}(\Omega) \to [0, \infty]$,

$$\mu^*(A) := \inf\{\sum_{k=1}^{\infty} \mu(A_k) \mid A \subset \bigcup_{k=1}^{\infty} A_k \text{ mit } A_k \in \mathcal{H}\} \quad . \tag{3.1}$$

Diese Funktion ist zwar für alle Teilmengen $A \subset \Omega$ definiert, aber nur ein sogenanntes äußeres Maß.

Definition 3.28 *Eine Funktion $\eta : \mathcal{P}(\Omega) \to [0, \infty]$ heißt äußeres Maß auf Ω, falls*

- *$\eta(\emptyset) = 0$ gilt,*
- *$\eta(A) \leq \eta(B)$ für Teilmengen $A \subset B$ von Ω gilt,*
- *$\eta\left(\bigcup\limits_{k=1}^{\infty} A_k\right) \leq \sum\limits_{k=1}^{\infty} \eta(A_k)$ für jede Folge A_k von Teilmengen von Ω gilt ("η ist σ-subadditiv").*

Lemma 3.29 *Für jeden Inhalt μ auf einem Halbring $\mathcal{H}$ von Teilmengen von Ω (und somit insbesondere für jedes Prämaß auf einem Ring) ist die durch* (3.1) *definierte Funktion $\mu^* : \mathcal{P}(\Omega) \to [0, \infty]$ ein äußeres Maß.*

Beweis: Offensichtlich ist $\mu^*(\emptyset) = 0$ und $\mu^*(A) \leq \mu^*(B)$ bei $A \subset B$, denn jede abzählbare Überdeckung von B mit Mengen aus $\mathcal{H}$ ist auch eine von A. Zu zeigen bleibt also nur noch die σ-Subadditivität.

Sei dazu A_k eine Folge von Teilmengen von Ω. Ist $\mu^*(A_k) = \infty$ für ein $k \in \mathbb{N}$, so ist die Ungleichung $\mu^*\left(\bigcup\limits_{k=1}^{\infty} A_k\right) \leq \sum\limits_{k=1}^{\infty} \mu^*(A_k)$ trivialerweise gültig.

Sei daher $\mu^*(A_k) < \infty$ für alle $k \in \mathbb{N}$ und sei $\epsilon > 0$ vorgegeben. Dann gibt es nach Definition von μ^* als Infimum zu jedem $k \in \mathbb{N}$ eine Folge $B_{kl} \in \mathcal{H}$ mit $A_k \subset \bigcup_{l=1}^{\infty} B_{kl}$ und

$$\sum_{l=1}^{\infty} \mu(B_{kl}) \leq \mu^*(A_k) + 2^{-k}\epsilon \quad .$$

Also überdeckt $\bigcup_{k=1}^{\infty} \bigcup_{l=1}^{\infty} B_{kl}$ die Menge $\bigcup_{k=1}^{\infty} A_k$, und daher gilt

$$\mu^*\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \mu(B_{kl}) \leq \sum_{k=1}^{\infty} \left(\mu^*(A_k) + 2^{-k}\epsilon\right) = \left(\sum_{k=1}^{\infty} \mu^*(A_k)\right) + \epsilon \quad .$$

Da diese Ungleichung für jedes $\epsilon > 0$ gilt, folgt die σ-Subadditivität von μ^*. □

Man bemerke, dass das vorige Lemma sogar dann noch gültig bleibt, falls μ irgendeine Funktion mit Werten in $[0, \infty]$ und $\mu(\emptyset) = 0$ ist.

Nun ist aber ein äußeres Maß η im allgemeinen auf ganz $\mathcal{P}(\Omega)$ nicht σ-additiv. Dies ändert sich jedoch, wenn man das äußere Maß auf die sogenannte σ-Algebra der η-meßbaren Mengen einschränkt. Dazu bezeichne $A^c := \Omega \setminus A$ das Komplement einer Teilmenge $A \subset \Omega$.

Definition 3.30 *Sei η ein äußeres Maß auf Ω. Eine Menge $A \subset \Omega$ heißt η-meßbar, falls $\eta(B) \geq \eta(B \cap A) + \eta(B \cap A^c)$ für alle $B \subset \Omega$ gilt.*

Äquivalenterweise hätte man auch $\eta(B) = \eta(B \cap A) + \eta(B \cap A^c)$ verlangen können, denn $\eta(B) \leq \eta(B \cap A) + \eta(B \cap A^c)$ gilt trivialerweise wegen $B = (B \cap A) \cup (B \cap A^c)$ und der σ-Subadditivität von η (aus der natürlich die endliche Subadditivität folgt). Mit anderen Worten ist also eine Menge η-meßbar, falls sie jede Menge in disjunkte Teilmengen zerlegt, auf denen sich η additiv verhält.

Lemma 3.31 *Die Menge $\mathcal{A}_\eta := \{A \subset \Omega \,|\, A \text{ ist } \eta - \text{meßbar}\}$ ist eine σ-Algebra von Teilmengen von Ω und die Einschränkung $\eta : \mathcal{A}_\eta \to [0, \infty]$ von η auf $\mathcal{A}_\eta$ ein Maß.*

Beweis: Zunächst zeigen wir, dass $\mathcal{A}_\eta$ eine Algebra ist. Nach Aufgabe 3.2 ist dazu nur nachzuweisen, dass $\Omega \in \mathcal{A}_\eta$ gilt und $A \in \mathcal{A}_\eta$ auch $A^c \in \mathcal{A}_\eta$ impliziert (was offensichtlich der Fall ist), sowie dass aus $A, B \in \mathcal{A}_\eta$ auch $A \cup B \in \mathcal{A}_\eta$ folgt. Letzteres folgt aus der für jede Teilmenge $C \subset \Omega$ gültigen Ungleichungskette

$$\eta(C) \geq \eta(C \cap A) + \eta(C \cap A^c) \geq \eta(C \cap A) + \eta(C \cap A^c \cap B) + \eta(C \cap A^c \cap B^c) \geq$$
$$\eta((C \cap A) \cup (C \cap A^c \cap B)) + \eta(C \cap A^c \cap B^c) = \eta(C \cap (A \cup B)) + \eta(C \cap (A \cup B)^c)\,.$$

Nun ist noch noch zu zeigen, dass für jede Folge disjunkter $A_k \in \mathcal{A}_\eta$ die Vereinigung $A := \bigcup_{k=1}^{\infty} A_k$ in $\mathcal{A}_\eta$ liegt und $\eta(A) = \sum_{k=1}^{\infty} \eta(A_k)$ gilt, d.h. die Einschränkung von η auf $\mathcal{A}_\eta$ auch σ-additiv ist.

Dazu bemerke man, dass η auf $\mathcal{A}_\eta$ endlich additiv ist. Denn in der Definition von η-Meßbarkeit kann man nach der Vorüberlegung $\geq$ durch $=$ ersetzen und erhält dann für disjunkte $B, C \in \mathcal{A}_\eta$ und beliebiges $D \subset \Omega$ die Gleichung

$$\eta(D \cap (B \cup C)) = \eta(D \cap (B \cup C) \cap B) + \eta(D \cap (B \cup C) \cap B^c) = \eta(D \cap B) + \eta(D \cap C) \quad ,$$

also nicht nur die endliche Additivität von η auf $\mathcal{A}_\eta$ (durch Setzen von $D := \Omega$), sondern auch noch

$$\eta\left(D \cap \left(\bigcup_{k=1}^m A_k\right)\right) = \sum_{k=1}^m \eta(D \cap A_k)$$

für beliebige Teilmengen $D \subset \Omega$.

Nun haben wir schon bewiesen, dass $\mathcal{A}_\eta$ eine Algebra ist, also gilt $\bigcup_{k=1}^m A_k \in \mathcal{A}_\eta$, und daher für jedes $D \subset \Omega$

$$\eta(D) \geq \eta\left(D \cap \left(\bigcup_{k=1}^m A_k\right)\right) + \eta\left(D \cap \left(\bigcup_{k=1}^m A_k\right)^c\right) \geq \left(\sum_{k=1}^m \eta(D \cap A_k)\right) + \eta(D \cap A^c) \quad ,$$

also bei $m \to \infty$ wegen der σ-Subadditivität von η sogar

$$\eta(D) \geq \left(\sum_{k=1}^\infty \eta(D \cap A_k)\right) + \eta(D \cap A^c) \geq \eta(D \cap A) + \eta(D \cap A^c) \geq \eta(D)$$

(die letzte Ungleichung galt ja trivialerweise für alle Mengen $A \subset \Omega$). Daher ist also A schon η-meßbar, und speziell mit $D := A$ folgt $\eta(D) = \sum\limits_{k=1}^\infty \eta(A_k)$. □

Ist also μ ein Prämaß auf einem Halbring $\mathcal{H}$ von Teilmengen von Ω, so ist die Einschränkung des äußeren Maßes μ^* auf die σ-Algebra der μ^*-meßbaren Mengen ein Maß. Somit ist μ^* ein heißer Kandidat für die Fortsetzung von μ zu einem Maß auf der von $\mathcal{H}$ erzeugten σ-Algebra $\sigma(\mathcal{H})$.

Um dies letztendlich zu beweisen, müssen wir einerseits noch zeigen, dass jede Menge $A \in \mathcal{H}$ schon μ^*-meßbar ist und $\mu^*(A) = \mu(A)$ gilt, und andererseits, dass es höchstens eine Fortsetzung von μ auf $\sigma(\mathcal{H})$ geben kann.

Sei also $A \in \mathcal{H}$ und $B \subset \Omega$ eine beliebige Teilmenge mit $\mu^*(B) < \infty$ (sonst ist nichts zu zeigen). Wegen $\mu^*(B) < \infty$ gibt es insbesondere eine Folge $B_k \in \mathcal{H}$ mit $B \subset \bigcup\limits_{k=1}^\infty B_k$. Aus $\mu(B_k) = \mu(B_k \cap A) + \mu(B_k \setminus A)$ folgt dann

$$\sum_{k=1}^\infty \mu(B_k) = \sum_{k=1}^\infty \mu(B_k \cap A) + \sum_{k=1}^\infty \mu(B_k \setminus A) \geq \mu^*(B \cap A) + \mu^*(B \setminus A)$$

und also $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \setminus A)$, d.h. A ist μ^*-meßbar.

Desweiteren gilt nicht nur $\mu^*(A) \leq \mu(A)$ für alle $A \in \mathcal{H}$ nach Definition von μ^* als Infimum, sondern auch die umgekehrte Ungleichung. Denn μ ist als Prämaß auf dem von $\mathcal{H}$ erzeugten Ring $\mathcal{R}$ σ-subadditiv. Ist nämlich $A \subset \bigcup_{k=1}^{\infty} A_k$ eine Überdeckung von $A \in \mathcal{R}$ mit Mengen $A_k \in \mathcal{R}$, dann ist

$$A = \bigcup_{k=1}^{\infty} A \cap \left(A_k \setminus \left(\bigcup_{l=1}^{k-1} A_l \right) \right)$$

eine disjunkte Zerlegung von A in Mengen aus $\mathcal{R}$, und somit gilt aufgrund der σ-Additivität von μ

$$\mu(A) = \sum_{k=1}^{\infty} \mu \left(A \cap \left(A_k \setminus \left(\bigcup_{l=1}^{k-1} A_l \right) \right) \right) \leq \sum_{k=1}^{\infty} \mu(A_k) \quad .$$

Also gilt auch die Ungleichung $\mu(A) \leq \mu^*(A)$, was zu zeigen war.

Abschließend noch zur Eindeutigkeit der Fortsetzung von μ. Um diese beweisen zu können, nennen wir μ ein σ-endliches Prämaß auf Ω, wenn μ ein Prämaß auf einem Halbring $\mathcal{H}$ von Teilmengen von Ω ist und es eine Folge $A_k \in \mathcal{H}$ von Teilmengen mit $\mu(A_k) < \infty$ gibt, die Ω überdecken.

Beispiel 3.32

- *Das Elementarvolumen* Vol *auf dem* $\mathbb{R}^n$ *ist* σ*-endlich, denn der* $\mathbb{R}^n$ *kann in abzählbar viele halboffene Quader von endlichem Volumen zerlegt werden, z.B. in die Würfel* $(m_{1k}, m_{1k} + 1] \times \cdots \times (m_{nk}, m_{nk} + 1]$ *mit* $m_k \in \mathbb{Z}^n$.
- *Das im ersten Punkt von Beispiel 3.13 genannte Zählmaß ist für eine Grundmenge* Ω *mit überzählbar vielen Elementen nicht* σ*-endlich, denn das Zählmaß hat nur auf endlichen Teilmengen von* Ω *einen endlichen Wert, aber* Ω *läßt sich als überabzählbare Menge nicht in abzählbar viele endliche Teilmengen zerlegen.*

Bezeichne μ die gerade konstruierte Fortsetzung und ν eine weitere Fortsetzung des ursprünglichen Prämaßes zu einem Maß auf der von $\mathcal{H}$ erzeugten σ-Algebra $\sigma(\mathcal{H})$, dann ist für die Eindeutigkeit nur $\mu(B \cap A_k) = \nu(B \cap A_k)$ für alle $B \in \sigma(\mathcal{H})$ zu beweisen. Also reicht es aus, für jede Teilmenge B einer Teilmenge von endlichem Prämaß die Gleichheit $\mu(B) = \nu(B)$ zu zeigen. Nach Definition unserer Fortsetzung gilt aber

$$\mu(B) = \inf\{\sum_{k=1}^{\infty} \mu(B_k) \,|\, B \subset \bigcup_{k=1}^{\infty} B_k \text{ mit } B_k \in \mathcal{H}\}$$

und daher auch

$$\mu(B) = \inf\{\sum_{k=1}^{\infty} \nu(B_k) \,|\, B \subset \bigcup_{k=1}^{\infty} B_k \text{ mit } B_k \in \mathcal{H}\} \geq \nu(B) \quad .$$

Ebenso gilt $\mu(A \setminus B) \geq \nu(A \setminus B)$ für jede B enthaltende Menge $A \in \mathcal{H}$ endlichen Prämaßes, und daher

$$\mu(A) = \nu(A) = \nu(A \setminus B) + \nu(B) \leq \mu(A \setminus B) + \mu(B) = \mu(A) \quad ,$$

was $\nu(B) = \mu(B)$ beweist. Denn wegen $\nu(A \setminus B) \leq \mu(A \setminus B)$ und $\nu(B) \leq \mu(B)$ kann Gleichheit der beiden Seiten nur vorliegen, wenn sowohl $\nu(A \setminus B) = \mu(A \setminus B)$ als auch $\nu(B) = \mu(B)$ gilt.

Wir fassen unsere Ergebnisse im folgenden Satz zusammen:

Satz 3.33 *Jedes Prämaß μ auf einem Halbring von Teilmengen von Ω kann zu einem Maß auf einer σ-Algebra von Teilmengen von Ω fortgesetzt werden. Ist μ auf Ω sogar σ-endlich, dann ist die Fortsetzung eindeutig.*

Bemerkung 3.34 *Auch wenn man dieser Konstruktion nach Carathéodory nicht direkt ansieht, dass es sich bei den μ^*-meßbaren Mengen A mit $\mu^*(A) < \infty$ um die Vervollständigung des Halbrings $\mathcal{H}$ bzgl. des L^1-Abstandes $d(A, B) := \mu(A \setminus B) + \mu(B \setminus A)$ handelt, auf die μ fortgesetzt wurde, ist dies doch der Fall. Der Nachweis dafür wäre aber etwas abstrakter gewesen als die (auch in der Lehrbuch-Literatur übliche) Fortsetzung zu einem Maß nach Carathéodory.*

Wendet man den Fortsetungssatz 3.33 von Carathéodory auf das Prämaß Vol an und nennt die Vol*-meßbaren Mengen einfach Lebesgue-meßbar, so ergibt sich das folgende Korollar.

Korollar 3.35 *Das Volumen* Vol *von halboffenen Quadern im $\mathbb{R}^n$ läßt sich eindeutig zu einem Maß auf der σ-Algebra $\mathcal{L}(\mathbb{R}^n)$ der Lebesgue-meßbaren Mengen fortsetzen, welches Lebesgue-Maß genannt wird.*

Nullmengen Nach Konstruktion ist jede Menge aus der Borel-σ-Algebra $\mathcal{B}(\mathbb{R}^n)$ auch Lebesgue-meßbar. Tatsächlich enthält aber die Lebesgue-σ-Algebra $\mathcal{L}(\mathbb{R}^n)$ sehr viel mehr Mengen als $\mathcal{B}(\mathbb{R}^n)$.

Definition 3.36 *Sei μ ein Maß auf einer σ-Algebra. Dann heißt eine Menge A aus der σ-Algebra μ-Nullmenge (oder einfach Nullmenge, wenn klar ist, welches Maß gemeint ist), falls $\mu(A) = 0$ gilt.*

Beispiel 3.37 *Ist A_k eine Folge von Mengen der σ-Algebra und gilt $\mu(A_k) = 0$, dann wegen der σ-Subadditivität von μ auch $\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = 0$. Insbesondere sind abzählbare Vereinigungen von Nullmengen wieder Nullmengen.*

Speziell ist jede abzählbare Teilmenge des $\mathbb{R}^n$ wie z.B. $\mathbb{Q}^n$ bzgl. des Lebesgue-Maßes eine Nullmenge, da Punkte Lebesgue-Maß Null haben. Angemerkt sei aber, dass es auch überabzählbare Teilmengen des $\mathbb{R}^n$ gibt, die Lebesgue-Nullmengen sind.

Die Lebesgue-σ-Algebra enthält nun (im Gegensatz zur Borel-σ-Algebra, was wir hier aber nicht beweisen wollen) sogar jegliche Teilmengen A mit $\mathrm{Vol}^*(A) = 0$. Denn dann gilt $\mathrm{Vol}^*(B) \geq \mathrm{Vol}^*(B \cap A) + \mathrm{Vol}^*(B \cap A^c)$ für jede Teilmenge $B \subset \mathbb{R}^n$, da $\mathrm{Vol}^*(B \cap A) \leq \mathrm{Vol}^*(A) = 0$ wegen $B \cap A \subset A$ und $\mathrm{Vol}^*(B \cap A^c) \leq \mathrm{Vol}^*(B)$ wegen $B \cap A^c \subset B$ gilt, also ist A schon Lebesgue-meßbar.

Da Nullmengen im Sinne des Maßes kleine Mengen sind, macht die folgende Sprechweise Sinn:

Definition 3.38 *Man sagt, eine Aussage $A(x)$, die von $x \in \Omega$ abhängt, gilt für μ-fast alle $x \in \Omega$ (oder μ-fast überall), falls es eine Nullmenge Z gibt, so dass $A(x)$ für alle $x \in Z^c$ gilt.*

So ist beispielsweise jede Funktion $f : \mathbb{R} \to \mathbb{R}$ mit abzählbar vielen Sprungstellen – wie z.B. die Funktion $f(x) = \lfloor x \rfloor$, die jedem $x \in \mathbb{R}$ die nächstkleinere ganze Zahl zuordnet – fast überall stetig, denn die Punkte, in denen f nicht stetig ist, sind nur abzählbar viele und bilden somit eine Lebesgue-Nullmenge.

Angesichts der Tatsache, dass beliebige offene bzw. abgeschlossene Teilmengen des $\mathbb{R}^n$, deren abzählbare Vereinigungen und auch alle Teilmengen von Nullmengen Lebesgue-meßbar sind, stellt sich die Frage, ob es überhaupt Teilmengen des $\mathbb{R}^n$ gibt, die nicht Lebesgue-meßbar sind (ansonsten wäre die Lebesgue-σ-Algebra nämlich gerade die Potenzmenge $\mathcal{P}(\mathbb{R}^n)$). Tatsächlich gibt es leider solche nicht-meßbaren Teilmengen, wie im folgenden Abschnitt gezeigt wird.

Translations- und Bewegungsinvarianz des Lebesgue-Maßes Das Lebesgue-Maß μ ist translationsinvariant, d.h. ist $A \subset \mathbb{R}^n$ Lebesgue-meßbar und $f : x \mapsto x + y$ die Translation um $y \in \mathbb{R}^n$, dann gilt $\mu(A) = \mu(f(A))$. Tatsächlich charakterisiert die Translationsinvarianz sogar das Lebesgue-Maß [2].

Satz 3.39 *Das Lebesgue-Maß ist translationsinvariant, und jedes translationsinvariante Maß μ auf der Lebesgue-σ-Algebra mit $\mu((0,1]^n) = 1$ ist identisch mit dem Lebesgue-Maß.*

Beweis: Das Elementarvolumen Vol ist translationsinvariant, also auch das äußere Maß Vol* und somit das Lebesgue-Maß als Einschränkung des äußeren Maß Vol* auf die Lebesgue-meßbaren Mengen.

Sei nun umgekehrt μ ein translationsinvariantes Maß auf der Lebesgue-σ-Algebra des $\mathbb{R}^n$ mit $\mu((0,1]) = 1$. Betrachte zu $m_1, \ldots, m_n \in \mathbb{N}$ die Zerlegung

$$(0,1] = \bigcup_{\substack{k_j=0,\ldots,m_j-1\\ j=1,\ldots,n}} \left(\prod_{i=1}^{n} \left(0, \frac{1}{m_i}\right] + (\frac{k_1}{m_1}, \ldots, \frac{k_n}{m_n})^T \right)$$

Die $m_1 \cdots m_n$ Mengen aus der Zerlegung haben wegen der Translationsinvarianz das gleiche Maß wie $\mu(\prod_{i=1}^n \left(0, \frac{1}{m_i}\right])$, und wegen $\mu((0,1]) = 1$ folgt daher $\mu(\prod_{i=1}^n \left(0, \frac{1}{m_i}\right]) = \prod_{i=1}^n \frac{1}{m_i}$. Nochmals aufgrund der Translationsinvarianz hat also μ auf allen Quadern mit rationalen Randpunkten denselben Wert wie Vol. Somit stimmen μ und das von Vol induzierte Lebesgue-Maß aufgrund der Eindeutigkeit der Fortsetzung eines σ-endlichen Prämaßes überein. $\square$

[2] Man beachte aber, dass auch das Zählmaß μ translationsinvariant ist, jedoch $\mu((0,1]^n) = \infty$ gilt.

Aus der Translationsinvarianz des Lebesgue-Maßes folgt auch sofort die Existenz nicht-meßbarer Mengen, zu deren Nachweis man jedoch das Auswahlaxiom benutzen muß.

Korollar 3.40 *Es gibt Teilmengen $M \subset \mathbb{R}^n$ mit $M \notin \mathcal{L}(\mathbb{R}^n)$.*

Beweis: Bezeichne μ das Lebesgue-Maß auf $\mathbb{R}^n$. Wähle in jeder der Äquivalenzklassen $x + \mathbb{Q}^n$ aus $\mathbb{R}^n/\mathbb{Q}^n$ einen Vertreter aus dem Einheitswürfel $[0,1]^n$. Zu der Menge $A \subset [0,1]^n$ dieser Vertreter betrachte man mit einer Abzählung q_k von $\mathbb{Q}^n \cap [-1,1]$ die Menge $B := \bigcup\limits_{k=1}^{\infty}(q_k + A)$. Dann enthält B offensichtlich den Einheitswürfel $[0,1]^n$ und müsste somit bei Meßbarkeit von A ein Volumen ≥ 1 haben, andererseits müsste wegen der σ-Additivität und Translationsinvarianz auch $\mu(B) = \sum\limits_{k=1}^{\infty}\mu(q_k + A) = \sum\limits_{k=1}^{\infty}\mu(A)$ gelten, was aufgrund der Beschränktheit von B schon $\mu(A) = 0$ und auch dann auch $\mu(B) = 0$ impliziert. Aufgrund dieses Widerspruches muss also A nicht-Lebesgue-meßbar sein. □

Das Lebesgue-Maß ist nicht nur translationsinvariant, sondern auch invariant unter Bewegungen, d.h. zusätzlich gilt für orthogonale lineare Abbildungen $f \in O(n)$ die Gleichung $\mu(A) = \mu(f(A))$. Tatsächlich kann man für jede lineare Abbildung f berechnen, wie das Volumen unter solch einer Abbildung verzerrt wird.

Satz 3.41 *Das Lebesgue-Maß μ auf dem $\mathbb{R}^n$ ist bewegungsinvariant, es gilt sogar $\mu(f(A)) = |\det(f)|\mu(A)$ für jede lineare Abbildung $f : \mathbb{R}^n \to \mathbb{R}^n$.*

Beweis: Das Maß $\nu(A) := \mu(f(A))$ ist wohldefiniert, denn für eine lineare Abbildung $f : \mathbb{R}^n \to \mathbb{R}^n$ ist mit A auch $f(A)$ Lebesgue-meßbar, und translationsinvariant, denn $f(A+y) = f(A) + f(y)$. Also muss ν nach Satz 3.39 ein Vielfaches des Lebesgue-Maßes sein, d.h. $\nu(A) = C\mu(A)$ gilt mit einer Konstanten $C \in \mathbb{R}_0^+$.

Zu zeigen bleibt $C = |\det(f)|$. Ist f eine orthogonale Abbildung, so gilt $f(B_1(0)) = B_1(0)$ für die Einheitskugel $B_1(0) \subset \mathbb{R}^n$ bzgl. der Euklidischen Norm und daher $\nu(B_1(0)) = \mu(B_1(0))$, also $C = 1 = |\det(f)|$.

Ist f eine beliebige lineare Abbildung, dann gibt es orthogonale Abbildungen v, w und eine durch eine Diagonalmatrix mit nichtnegativen Einträgen d_i repräsentierte lineare Abbildung d mit $f = v \circ d \circ w$. Nun gilt aufgrund der Orthogonalität von v, w wie eben gerade bewiesen $\nu(A) = \mu(v(d(w(A)))) = \mu(d(A))$ und daher $\nu((0,1]^n) = (\prod_{i=1}^n d_i)\mu((0,1]^n) = \det(d)\mu((0,1]^n)$, sowie $|\det f| = |\det(v)| \cdot \det(d) \cdot |\det(w)| = \det(d)$, also wirklich $C = |\det f|$. □

3.3 Integration bzgl. eines Maßes

In diesem gesamten Abschnitt sei $(\Omega, \mathcal{A}, \mu)$ ein σ-endlicher Maßraum, d.h. $\mathcal{A}$ sei eine σ-Algebra von Teilmengen von Ω, deren Mengen wir meßbare Teilmengen nennen, und $\mu : \mathcal{A} \to [0, \infty]$ sei ein σ-endliches Maß.

Speziell sind wir natürlich an dem Fall interessiert, bei dem $\Omega = \mathbb{R}^n$, $\mathcal{A}$ die σ-Algebra der Lebesgue-meßbaren Teilmengen und μ das Lebesgue-Maß ist. Aber die im folgenden geschilderte Integrationstheorie ist weitgehend unabhängig davon, ob wir das Lebesgue-Maß oder irgendein anderes Maß betrachten.

Es war ja unserer Ziel, das Integral allgemeiner Funktionen aus dem elementaren Integral von Treppenfunktionen durch Fortsetzung auf die Vervollständigung zu gewinnen. Da es uns im letzten Kapitel gelungen ist, das Elementarvolumen von Quadern zu einem Maß auf der viel größeren Klasse der meßbaren Teilmengen des $\mathbb{R}^n$ fortzusetzen, können wir zunächst schon einmal das Integral von Treppenfunktionen $g = \sum_{i=1}^k g_i 1_{A_i}$ zu disjunkten Teilmengen $A_i \in \mathcal{A}$ mit $\mu(A_i) < \infty$ durch

$$\int_\Omega g(x)\, d\mu(x) := \sum_{i=1}^k g_i \mu(A_i)$$

definieren. Dies setzt im Falle des Lebesgue-Maßes das Integral von Treppenfunktionen zu Quadern fort. Ohne Probleme können wir hier zulassen, dass $g_i \in Y$ Elemente eines Banach-Raumes Y sind, auch wenn wir hauptsächlich an dem Fall $Y = \mathbb{R}$ interessiert sind. Um die Notation einheitlich zu halten, schreiben wir für die Norm auf Y wie auch für den reellen Betrag einfach $|\cdot|$. Das Integral ist dann linear und erfüllt die Ungleichung

$$|\int_\Omega g(x)\, d\mu(x)| \leq \int_\Omega |g(x)|\, d\mu(x) \quad .$$

Man bemerke, dass dieses Integral unabhängig von der Wahl der Zerlegung der Menge $A := \{x \,|\, g(x) \neq 0\}$ ist, denn ist g bei $A = \bigcup_{j=1}^l B_j$ auch auf jeder der disjunkten Mengen B_j konstant, dann ist g auch auf den Mengen der feineren Zerlegung $A = \bigcup\limits_{i=1}^k \bigcup\limits_{j=1}^l (A_i \cap B_j)$ konstant und wegen

$$\sum_{j=1}^l g_i \mu(A_i \cap B_j) = g_i \mu(A_i)$$

ergibt sich daher bzgl. der anderen Zerlegung derselbe Integralwert.

Nun wollen wir den Raum $\mathrm{Step}_\mu(\Omega, Y)$ der Treppenfunktionen zu Mengen endlichen Maßes mit Werten in Y bzgl. der Norm

$$\|g\|_1 := \int_\Omega |g(x)|\, d\mu(x) = \sum_{i=1}^k |g_i| \mu(A_i)$$

vervollständigen. Das folgende Lemma ist dabei fundamental, denn es besagt, dass man in jeder Cauchy-Folge von Treppenfunktionen bzgl. $\|\cdot\|_1$ eine Teilfolge finden kann, die außerhalb einer Nullmenge punktweise konvergiert.

Vor dessen Beweis führen wir aber noch eine weitere Schreibweise ein: Statt $\int_\Omega 1_A(x) g(x)\, d\mu(x)$ für eine meßbare Menge A (wobei man $g|_A := 1_A \cdot g$ als die Einschränkung von g auf A ansehen kann) schreiben wir $\int_A g(x)\, d\mu(x)$. Dann gilt für disjunkte A, B die Gleichung

$$\int_{A \cup B} g(x)\, d\mu(x) = \int_A g(x)\, d\mu(x) + \int_B g(x)\, d\mu(x) \quad .$$

Lemma 3.42 *Sei $f_k \in \mathrm{Step}_\mu(\Omega, Y)$ eine Cauchy-Folge von Treppenfunktionen zu Mengen endlichen Maßes bzgl. $\|\cdot\|_1$. Dann gibt es eine Teilfolge von f_k, die außerhalb einer μ-Nullmenge punktweise konvergiert.*

Beweis: Da f_k eine Cauchy-Folge bzgl. $\|\cdot\|_1$ ist, gibt es für jedes $m \in \mathbb{N}$ ein $K_m \in \mathbb{N}$ mit $\|f_k - f_l\|_1 \le \frac{1}{2^{2m}}$ bei $k, l \ge K_m$. Betrachte die Teilfolge $g_m := f_{\max(K_1,\ldots,K_m)}$, dann gilt $\|g_k - g_l\|_1 \le \frac{1}{2^{2l}}$ für $k \ge l$.

Wir zeigen, dass die Reihe $g_1(x) + \sum_{k=1}^\infty (g_{k+1}(x) - g_k(x))$ für alle Punkte außerhalb einer Nullmenge Z absolut konvergiert. Dann konvergiert auch $g_k(x)$ für alle $x \notin Z$, denn die k-te Partialsumme der Reihe ist gerade $g_k(x)$.

Sei dazu Y_l die Menge der Punkte $x \in \Omega$, für die $|g_{l+1}(x) - g_l(x)| \ge \frac{1}{2^l}$ gilt. Natürlich hat Y_l ein endliches Maß, denn die Funktionen g_l und g_{l+1} sind Treppenfunktionen zu Mengen endlichen Maßes. Nach Definition von Y_l gilt

$$\frac{1}{2^l}\mu(Y_n) = \int_{Y_l} \frac{1}{2^l}\, d\mu(x) \le \int_{Y_l} |g_{l+1} - g_l(x)|\, d\mu(x) \le \|g_{l+1} - g_l\|_1 \le \frac{1}{2^{2l}} \quad .$$

Also gilt $\mu(Y_l) \le \frac{1}{2^l}$. Bilde $Z_l := \bigcup_{j=l}^{\infty} Y_j$, dann gilt $\mu(Z_l) \le \frac{1}{2^{l-1}}$. Ist nun $x \notin Z_l$, dann gilt für $k \ge l$ die Ungleichung $|g_{k+1}(x) - g_k(x)| < \frac{1}{2^k}$, und somit konvergiert die Reihe $\sum_{k=l}^\infty g_{k+1}(x) - g_k(x)$ absolut und gleichmäßig auf Z_l^c. Ist also $Z := \bigcap_{l=1}^\infty Z_l$, dann ist Z eine Nullmenge wegen $\mu(Z) = \lim_{l\to\infty} \mu(Z_l) = 0$, und die Reihe konvergiert für alle $x \in Z^c$. □

Aus dem Beweis sieht man, dass es sogar zu jedem $\epsilon > 0$ eine Menge Z vom Maß $\mu(Z) < \epsilon$ gibt, so dass f_k außerhalb von Z gleichmäßig konvergiert.

Das vorige Lemma erlaubt es also, jedem Element der Vervollständigung des Raumes $\mathrm{Step}_\mu(\Omega, Y)$ bzgl. $\|\cdot\|_1$, das ja nach dem Beweis von Satz 1.31 gerade eine Äquivalenzklasse von Cauchy-Folgen ist, eine Funktion auf Ω mit Werten in Y zuzuordnen, nämlich den außerhalb einer Nullmenge existierenden punktweisen Grenzwert einer Teilfolge der Cauchy-Folge. Wir werden in Korollar 3.54 beweisen, dass die punktweisen Grenzwerte solcher Teilfolgen bis auf Nullmengen übereinstimmen, so dass man dadurch wirklich der gesamten Cauchy-Folge eine fast überall definierte Funktion zuordnen kann.

Definition 3.43 *Eine Funktion f, für die eine Cauchy-Folge f_k von Treppenfunktionen zu Mengen endlichen Maßes bzgl. $\|\cdot\|_1$ existiert, die außerhalb einer Nullmenge punktweise gegen f konvergiert, nennt man μ-integrierbar (oder integrierbar bzgl. μ, oder einfach integrierbar, wenn klar ist, auf welches Maß man sich bezieht).*

Tatsächlich kann man für μ-integrierbare f aufgrund des folgenden Lemmas das Integral bzgl. μ definieren.

Lemma 3.44 *Ist f integrierbar bzgl. μ und f_k eine Cauchy-Folge von Treppenfunktionen zu Mengen endlichen Maßes bzgl. $\|\cdot\|_1$, die außerhalb einer Nullmenge punktweise gegen f konvergiert, dann existiert der Grenzwert $\lim\limits_{k\to\infty} \int_\Omega f_k(x)\, d\mu(x)$ und ist unabhängig von der gewählten Folge f_k.*

Beweis: Die Existenz von $\lim_{k\to\infty} \int_\Omega f_k(x)\, d\mu(x)$ folgt direkt aus der Ungleichung

$$\left| \int_\Omega f_k(x)\, d\mu(x) - \int_\Omega f_l(x)\, d\mu(x) \right| \leq \int_\Omega |f_k - f_l|\, d\mu(x) = \|f_k - f_l\|_1 \quad ,$$

denn da f_k eine Cauchy-Folge bzgl. $\|\cdot\|_1$ ist, ist $\int_\Omega f_k(x)\, d\mu(x)$ eine Cauchy-Folge in Y und konvergiert somit aufgrund der Vollständigkeit von Y.

Sei nun $h_k := f_k - g_k$ die Differenz zweier Cauchy-Folgen, die punktweise außerhalb einer Nullmenge gegen f konvergieren. Dann ist h_k nicht nur eine Cauchy-Folge bzgl. $\|\cdot\|_1$, die punktweise außerhalb einer Nullmenge gegen die Nullfunktion konvergiert, sondern sogar eine Nullfolge bzgl. $\|\cdot\|_1$. Tatsächlich, zu vorgegebenem $\epsilon > 0$ existiert ein $K \in \mathbb{N}$ mit $\|h_k - h_l\|_1 \leq \epsilon$ für $k, l \geq K$, da h_k eine Cauchy-Folge bzgl. $\|\cdot\|_1$ ist. Verschwindet die Treppenfunktion h_K außerhalb der Menge A endlichen Maßes, dann gilt für $k \geq K$ die Ungleichung

$$\int_{A^c} |h_k(x)|\, d\mu(x) = \int_{A^c} |h_k(x) - h_K(x)|\, d\mu(x) \leq \int_\Omega |h_k(x) - h_K(x)|\, d\mu(x) \leq \epsilon$$

Nach dem Beweis von Lemma 3.42 gibt es eine meßbare Menge Z mit $\mu(Z) \leq \frac{\epsilon}{1+\|h_K\|_1}$ und eine Teilfolge von h_k, die außerhalb von Z gleichmäßig gegen die Nullfunktion konvergiert. Für genügend große zu dieser Teilfolge gehörige Indizes k gilt dann $\int_{A\setminus Z} |h_k| \leq \epsilon$ und auch

$$\int_Z |h_k(x)|\, d\mu(x) \leq \int_Z |h_k(x) - h_K(x)|\, d\mu(x) + \int_Z |h_K(x)|\, d\mu(x) \leq \|h_k - h_K\|_1 + \mu(Z)\|h_K(x)\|_1 \leq 2\epsilon$$

Also gilt

$$\int_\Omega |h_k(x)|\, d\mu(x) = \int_{A^c} |h_k(x)|\, d\mu(x) + \int_{A\setminus Z} |h_k(x)|\, d\mu(x) + \int_Z |h_k(x)|\, d\mu(x) \leq 4\epsilon \quad ,$$

d.h. h_k ist eine Nullfolge bzgl. $\|\cdot\|_1$. □

Definition 3.45 *Für eine μ-integrierbare Funktionen $f : \Omega \to Y$ heißt*

$$\int_\Omega f(x)\, d\mu(x) := \lim_{k\to\infty} \int_\Omega f_k(x)\, d\mu(x)$$

das Integral von f bzgl. des Maßes μ, wobei f_k irgendeine Cauchy-Folge von Treppenfunktionen zu Mengen endlichen Maßes bzgl. $\|\cdot\|_1$ ist, die fast überall gegen f konvergiert.

Speziell für das Lebesgue-Maß nennt man dieses Integral das Lebesgue-Integral und spricht von Lebesgue-integrierbaren Funktionen.

Insbesondere ist für eine integrierbare Funktion f auch $\|f\|_1 := \int_\Omega |f(x)|\, d\mu(x)$ wohldefiniert, denn ist f_k eine Cauchy-Folge bzgl. $\|\cdot\|_1$, die fast überall punktweise gegen f konvergiert, dann ist auch $|f_k|$ eine Cauchy-Folge, die fast überall punktweise gegen $|f|$ konvergiert.

Beispiel 3.46 *Die Funktion* $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = 1_{\mathbb{Q}^n \cap [0,1]^n}(x) = \begin{cases} 1 & x \in \mathbb{Q}^n \cap [0,1]^n \\ 0 & \textit{sonst} \end{cases}$, *ist Lebesgue-integrierbar mit* $\int_{\mathbb{R}^n} f(x)\, d\mu(x) = 0$, *obwohl* $1_{\mathbb{Q}\cap[0,1]}$ *nicht Riemann-integrierbar ist.*

Elementare Eigenschaften des Integrals

Lemma 3.47 *Das Integral* $I : f \mapsto \int_\Omega f(x)\, d\mu(x)$ *bzgl. des Maßes* μ *ist linear, mit* f *ist auch* $f|_A$ *für jede meßbare Menge* A *integrierbar bzgl.* μ, *und es gilt* $\left|\int_A f(x)\, d\mu(x)\right| \leq \int_A |f(x)|\, d\mu(x)$ *(insbesondere ist das Integral für* $\mathbb{R}$*-wertige Funktionen monoton).*

Beweis: Sind f_k bzw. g_k Cauchy-Folgen von Treppenfunktionen zu Mengen endlichen Maßes bzgl. $\|\cdot\|_1$, die fast überall punktweise gegen f bzw. g konvergieren, dann ist auch $f_k + g_k$ eine Cauchy-Folge bzgl. $\|\cdot\|_1$, die fast überall punktweise gegen $f + g$ konvergiert, und für festes $\lambda \in \mathbb{R}$ ist auch λf_k eine Cauchy-Folge bzgl. $\|\cdot\|_1$, die fast überall gegen λf konvergiert. Somit ist das Integral linear aufgrund seiner Definition als Grenzwert der Integrale von Treppenfunktionen.

Die beiden anderen Eigenschaften übertragen sich ebenso sofort aus der Gültigkeit für Treppenfunktionen. □

Eine Folgerung aus der Linearität und der Integrierbarkeit von $f|_A$ und $f|_B$ für disjunkte meßbare Mengen A, B ist die Gleichung

$$\int_{A\cup B} f(x)\, d\mu(x) = \int_A f(x)\, d\mu(x) + \int_B f(x)\, d\mu(x) \quad .$$

Nun wenden wir uns der Vollständigkeit des Raumes der integrierbaren Funktionen zu. Offensichtlich definiert $\|\cdot\|_1$ eine Halbnorm auf dem Raum aller integrierbaren Funktionen.

Satz 3.48 *Der Raum der integrierbaren Funktionen ist bzgl. der Halbnorm* $\|\cdot\|_1$ *vollständig.*

Beweis: Ist f_k eine Cauchy-Folge integrierbarer Funktionen bzgl. $\|\cdot\|_1$, dann gibt es aufgrund der Definition des Integrals zu jedem k eine Treppenfunktion g_k zu Mengen endlichen Maßes mit $\|f_k - g_k\|_1 \leq \frac{1}{k}$. Die Folge g_k ist dann eine Cauchy-Folge von Treppenfunktionen, denn die Terme auf der rechten Seite von

$$\|g_k - g_l\|_1 \leq \|g_k - f_k\|_1 + \|f_k - f_l\|_1 + \|f_l - g_l\|_1$$

werden für genügend große k, l beliebig klein. Nach Lemma 3.42 hat g_k eine Teilfolge g_{k_i}, die fast überall punktweise gegen eine integrierbare Funktion f konvergiert, und wegen $\|f_{k_i} - f\|_1 \leq \|f_{k_i} - g_{k_i}\|_1 + \|g_{k_i} - f\|_1$ mit dieser Teilfolge g_{k_i} konvergiert f_{k_i} bzgl. $\|\cdot\|_1$ gegen f, also auch f_k (denn eine Cauchy-Folge konvergiert bereits dann, wenn eine Teilfolge konvergiert). □

Man beachte dabei, dass $\|\cdot\|_1$ nur eine Halbnorm auf dem Raum der integrierbaren Funktionen ist. Identifiziert man daher integrierbare Funktionen f und g, wenn $\|f - g\|_1 = 0$

ist (was gleichbedeutend damit ist, dass sich f und g nur auf Nullmengen unterscheiden, wie wir in Korollar 3.53 beweisen werden), so erhält man den Raum $L^1_\mu(\Omega, Y)$ der Äquivalenzklassen von bis auf Nullmengen definierten integrierbaren Funktionen. Dieser ist ein Banach-Raum bzgl. der Norm $\|\cdot\|_1$ und nichts anderes als die Vervollständigung des Raumes der Treppenfunktionen zu Quadern (oder auch Mengen endlichen Maßes). Die vorigen Sätze haben dabei gerade gezeigt, dass sich die Elemente der Vervollständigung als Funktionen interpretieren lassen, wenn man diese bei Übereinstimmung außerhalb von Nullmengen identifiziert.

Integrierbarkeit stetiger Funktionen Wir wissen aufgrund der Approximierbarkeit von stetigen Funktionen mit kompaktem Träger durch Treppenfunktionen zu Quadern bzgl. $\|\cdot\|_1$ schon, dass das iterierte Riemann-Integral solcher stetigen Funktionen mit kompaktem Träger mit ihrem Lebesgue-Integral übereinstimmt. Umgekehrt wissen wir auch schon, dass man auch Treppenfunktionen zu Quadern mittels stetiger Funktionen mit kompaktem Träger bzgl. $\|\cdot\|_1$ approximieren kann. Dies halten wir in folgendem Satz fest.

Satz 3.49 *Die Menge $C_c(\mathbb{R}^n, Y)$ der stetigen Funktionen mit kompaktem Träger liegt bzgl. $\|\cdot\|_1$ dicht im Raum der Lebesgue-integrierbaren Funktionen mit Werten im Banach-Raum Y.*

Man kann sogar jede stetige Funktion über kompakte Teilmengen integrieren.

Satz 3.50 *Jede stetige Funktion $f : K \to Y$ auf einer kompakten Menge $K \subset \mathbb{R}^n$ ist Lebesgue-integrierbar.*

Darüberhinaus kann man das Integral von f über K durch Riemann-Summen $\sum\limits_{k=1}^{m} f(\xi_k)\mu(A_k)$ beliebig genau annähern, wobei $\bigcup\limits_{k=1}^{m} A_k$ eine Zerlegung von K in disjunkte Teilmengen ist und $\xi_k \in A_k$ Stützstellen sind.

Beweis: Ohne Einschränkung sei K keine Lebesgue-Nullmenge.

Da f als stetige Funktion auf der kompakten Teilmenge K sogar gleichmäßig stetig ist, gibt es zu jedem $\epsilon > 0$ ein $\delta > 0$ mit $\|f(x) - f(x')\|_Y \leq \frac{\epsilon}{\mu(K)}$ für $x, x' \in K$ mit $\|x - x'\|_{\mathbb{R}^n} \leq \delta$.

Nun sei $A_1, \ldots, A_m$ eine Zerlegung von K in endlich viele disjunkte Lebesgue-meßbare Mengen von endlichem Maß mit maximalem Durchmesser δ bzgl. $\|\cdot\|_{\mathbb{R}^n}$. Solch eine existiert, man überdecke K einfach mit Kugeln vom Radius $\delta/2$, dann reichen aufgrund der Kompaktheit endlich viele zum Überdecken von K aus, und danach betrachte man die Schnitte dieser Kugeln.

Wählt man beliebige Stützstellen $\xi_k \in A_k$, dann gilt $\|f(x) - f(\xi_k)\|_Y \leq \frac{\epsilon}{\mu(K)}$ für alle $x \in A_k$ und daher für die Treppenfunktion $g := \sum_{k=1}^m f(\xi_k) 1_{A_k}$ die Ungleichung

$$\|f(x) - g(x)\|_Y \leq \frac{\epsilon}{\mu(K)}$$

für alle $x \in K$.

Somit gibt es einerseits eine Folge von Treppenfunktionen g_k, die auf K gleichmäßig gegen f konvergiert. Setzt man f durch Null auf ganz $\mathbb{R}^n$ fort, dann ist g_k sogar eine Cauchy-Folge bzgl. $\|\cdot\|_1$, und da g_k insbesondere punktweise gegen f konvergiert, ist f Lebesgue-integrierbar.

Andererseits gilt wegen

$$\| \left(\int_{A_k} f(x)\, d\mu(x) \right) - f(\xi_k)\mu(A_k)\|_Y \leq \int_{A_k} \|f(x) - f(\xi_k)\|_Y \, d\mu(x) \leq \frac{\epsilon}{\mu(K)} \mu(A_k)$$

sogar die Ungleichung

$$\| \left(\int_{A} f(x)\, d\mu(x) \right) - \sum_{k=1}^{m} f(\xi_k)\mu(A_k)\|_Y \leq \frac{\epsilon}{\mu(K)} \sum_{k=1}^{m} \mu(A_k) = \epsilon \quad ,$$

d.h. man kann das Lebesgue-Integral beliebig genau durch Riemann-Summen annähern.

□

Meßbare Abbildungen

Definition 3.51 *Eine Abbildung $f : \Omega \to \Omega'$ zwischen mit σ-Algebren $\mathcal{A}$ bzw. $\mathcal{A}'$ versehenen Mengen Ω bzw. Ω' heißt meßbar, falls $f^{-1}(A') \in \mathcal{A}$ für alle $A' \in \mathcal{A}'$ gilt.*

Beispiele für Lebesgue-meßbare Abbildungen sind alle Treppenfunktionen, aber auch alle punktweisen Grenzwerte von Folgen meßbarer Funktionen (was wir hier aber nicht beweisen werden).

Sind daher wie beim Lebesgue-Maß alle Teilmengen von Nullmengen meßbar, dann sind insbesondere alle integrierbaren Funktionen meßbar, denn sie sind bis auf eine Nullmenge solche punktweisen Grenzwerte. Deswegen spielt bei unserem Zugang zur Integrierbarkeit die Meßbarkeit nur eine untergeordnete Rolle.

3.4 Konvergenzsätze

In diesem Abschnitt wollen wir uns fragen, wann man für eine Folge f_k integrierbarer Funktionen Limesbildung und Integration vertauschen darf, und daraus Konsequenzen ableiten.

Zunächst wollen wir das fundamentale Lemma 3.42 auf Folgen von integrierbaren Funktionen ausdehnen.

Satz 3.52 *Sei f_k eine Folge μ-integrierbarer Funktionen, die bzgl. $\|\cdot\|_1$ gegen eine μ-integrierbare Funktion f konvergiert. Dann gibt es eine Teilfolge von f_k, die fast überall punktweise gegen f konvergiert.*

Beweis: Da f_k bzgl. $\|\cdot\|_1$ gegen f genau dann konvergiert, wenn $f_k - f$ bzgl. $\|\cdot\|_1$ gegen 0 konvergiert, reicht es, die Aussage für $f = 0$ zu zeigen.

Aufgrund der Konvergenz von f_k bzgl. $\|\cdot\|_1$ gegen die Nullfunktion gibt es eine Teilfolge (die wir wieder mit f_k bezeichnen), so dass $\|f_k\|_1 \leq \frac{1}{2^{2k}}$ gilt. Sei nun Y_k die Menge aller x, für die $|f_k(x)| > \frac{1}{2^k}$ gilt. Dann ist die Menge Y_k meßbar mit Maß

$$\frac{1}{2^k}\mu(Y_k) \leq \int_{Y_k} |f_k(x)|\, d\mu(x) \leq \|f_k\|_1 \leq \frac{1}{2^{2k}}$$

und daher gilt $\mu(Y_k) \leq \frac{1}{2^k}$. Somit hat $Z_k := \bigcup_{l=k}^{\infty} \mu(Y_l)$ ein Maß $\mu(Z_k) \leq \frac{1}{2^{k-1}}$, und außerhalb von Z_k konvergiert f_k wegen $|f_k(x)| \leq \frac{1}{2^k}$ bei $x \notin Z_k$ gleichmäßig gegen die Nullfunktion. Auf dem Schnitt aller Z_k, der eine μ-Nullmenge ist, konvergiert f_k daher zumindest noch punktweise gegen die Nullfunktion. □

Wir wollen zwei Folgerungen aus diesem Satz festhalten. Die erste haben wir schon bei der Definition von $L^1_\mu(\Omega, Y)$ genutzt.

Korollar 3.53 *Jede integrierbare Funktion f mit $\|f\|_1 = 0$ verschwindet fast überall.*

Beweis: Die konstante Folge $0, 0, \ldots$ konvergiert bzgl. μ gegen f und daher nach Satz 3.52 auch fast überall punktweise gegen f. Also ist f fast überall die Nullfunktion. □

Korollar 3.54 *Für eine Cauchy-Folge μ-integrierbarer Funktionen f_k, die fast überall punktweise gegen eine Funktion f konvergiert, ist f sowohl integrierbar als auch der Grenzwert von f_k bzgl. $\|\cdot\|_1$ ist.*

Beweis: Aufgrund der Vollständigkeit konvergiert f_k gegen eine integrierbare Funktion g, und daher konvergiert nach Satz 3.52 eine Teilfolge auch schon punktweise fast überall gegen g. Nun konvergiert f_k aber schon punktweise fast überall gegen f, also muss fast überall $f = g$ gelten. □

Insbesondere beweist dies, dass die punktweisen Grenzfunktionen verschiedener punktweise fast überall konvergenter Teilfolgen einer L^1-Cauchy-Folge außerhalb einer Nullmenge übereinstimmen, denn solche Teilfolgen konvergieren bzgl. der L^1-Norm gegen dieselbe integrierbare Funktion, nämlich den Grenzwert der gesamten Cauchy-Folge im vollständigen Raum der integrierbaren Funktionen. Dies zeigt endgültig, dass man die fast überall definierten punktweisen Grenzfunktionen von Teilfolgen einer Cauchy-Folgen von Treppenfunktionen mit den Elementen der Vervollständigung identifizieren kann.

Im folgenden beweisen wir zwei wichtige Konvergenzsätze, den Satz von Beppo-Levi über montone Konvergenz und den Satz von Lebesgue über majorisierte Konvergenz, und zeigen exemplarisch, wie man diese anwenden kann.

Der Satz von Beppo-Levi über monotone Konvergenz

Satz 3.55 *Sei f_k eine monotone Folge reellwertiger μ-integrierbarer Funktionen, für die die Folge der Integrale $\int_\Omega f_k(x)\,d\mu(x)$ beschränkt ist. Dann ist f_k eine Cauchy-Folge bzgl. $\|\cdot\|_1$, die sowohl bzgl. $\|\cdot\|_1$ als auch μ-fast überall punktweise gegen eine μ-integrierbare Funktion f konvergiert, und es gilt*

$$\int_\Omega f(x)\,d\mu(x) = \lim_{k\to\infty}\int_\Omega f_k(x)\,d\mu(x) \quad .$$

Beweis: Sei f_k monoton wachsend und $S := \sup_k \int_\Omega f_k(x)\,d\mu(x)$, dann gilt für $n \geq m$ die Ungleichung

$$\|f_n - f_m\|_1 = \int_\Omega f_n(x) - f_m(x)\,d\mu(x) \leq S - \int_\Omega f_m(x)\,d\mu(x) \quad .$$

Da die rechte Seite für $m \to \infty$ gegen Null strebt, ist f_k eine Cauchy-Folge, und diese konvergiert aufgrund der Vollständigkeit gegen eine Funktion f bzgl. $\|\cdot\|_1$. Nach Satz 3.52 hat f_k eine fast überall punktweise konvergente Teilfolge, aber aufgrund der Monotonie trifft dies dann auch schon auf f_k selbst zu. Somit konvergiert f_k sowohl bzgl. $\|\cdot\|_1$ als auch punktweise fast überall gegen f. Die Konvergenz der Integrale von f_k gegen das Integral von f folgt dann nach Korollar 3.54.

Der Beweis für monoton fallende Folgen f_k verläuft analog. □

Die zwei folgenden Korollare zeigen, wie man den Satz von Beppo-Levi über monotone Konvergenz anwenden kann.

Korollar 3.56 *Ist $A_1 \subset A_2 \subset \ldots$ eine Ausschöpfung der Menge $A = \bigcup_{k=1}^{\infty} A_k$ durch Lebesgue-meßbare Mengen, dann ist $f : A \to \mathbb{R}$ genau dann μ-integrierbar über A, wenn die Folge der Integrale $\int_{A_k} |f(x)|\,d\mu(x)$ beschränkt ist.*

Beweis: Einerseits folgt aus Integrierbarkeit von f über A die Ungleichung

$$\int_{A_k} |f(x)|\,d\mu(x) \leq \int_A |f(x)|\,d\mu(x) < \infty \quad ,$$

also ist die Beschränktheit der Integrale $\int_{A_k} |f(x)|\,d\mu(x)$ notwendig.

Andererseits ist die Folge des Betrags von $f|_{A_k}$ monoton und konvergiert punktweise gegen den Betrag von f auf A. Daher folgt aus der Beschränktheit der Integrale $\int_{A_k} |f(x)|\,d\mu(x)$ nach dem Satz über monotone Konvergenz die Integrierbarkeit des Betrages von f über A, also insbesondere die Integrierbarkeit von f über A. □

Korollar 3.57 *Sei $f : \Omega \to Y$ eine μ-integrierbare Funktion mit Werten in Banach-Raum Y, dann gibt es zu jedem $\epsilon > 0$ eine Menge A endlichen Maßes mit*
$\left|\int_\Omega f(x)\,d\mu(x) - \int_A f(x)\,d\mu(x)\right| \leq \epsilon.$

Beweis: Sei $\Omega = \bigcup_{k=1}^{\infty} A_k$ eine abzählbare Zerlegung in Mengen A_k endlichen Maßes (μ ist nach unseren generellen Voraussetzungen σ-endlich auf Ω).

Dann gilt mit $B_n := \bigcup_{k=1}^{n} A_k$ die Ungleichung

$$\left| \int_\Omega f(x)\, d\mu(x) - \int_{B_n} f(x)\, d\mu(x) \right| \leq \int_{\Omega \setminus B_n} |f(x)|\, d\mu(x) = \int_\Omega |f(x)|(1 - 1_{B_n})\, d\mu(x) \quad .$$

Da $|f(x)|(1 - 1_{B_n})$ für $n \to \infty$ monoton gegen die Nullfunktion fällt und die Folge der Integrale nach unten durch Null beschränkt ist, werden die Integrale $\int_\Omega |f(x)|(1 - 1_{B_n})\, d\mu(x)$ nach dem Satz über monotone Konvergenz beliebig klein, also sind für großes n auch kleiner als ϵ. Nun wähle $A := B_n$ mit einem solch großen n. □

Der Satz von Lebesgue über majorisierte Konvergenz Zur Vorbereitung des Satzes von Lebesgue über majorisierte Konvergenz halten wir eine Konsequenz aus dem Satz von Beppo-Levi über monotone Konvergenz fest.

Korollar 3.58 *Ist f_k eine Folge reellwertiger μ-integrierbarer Funktionen, und gibt es eine reellwertige μ-integrierbare Funktion g mit $|f_k(x)| \leq g(x)$ für alle $x \in \Omega$ und $k \in \mathbb{N}$, dann sind das Supremum $\sup_k f_k$ und das Infimum $\inf_k f_k$ integrierbar und es gilt*

$$\sup_k \int_\Omega f_k(x)\, d\mu(x) \leq \int_\Omega (\sup_k f_k(x))\, d\mu(x) \quad , \quad \int_\Omega (\inf_k f_k(x))\, d\mu(x) \leq \inf_k \int_\Omega f_k(x)\, d\mu(x)$$

Beweis: Die endlichen Suprema $\sup\limits_{k=1,\ldots,n} f_k(x)$ sind integrierbar, da man das Supremum von zwei integrierbaren Funktionen f, h durch $\sup(f, h) = \frac{1}{2}(f + h + |f - h|)$ als Linearkombination integrierbarer Funktionen schreiben kann.

Die Funktionenfolge $\left(\sup\limits_{k=1,\ldots,n} f_k(x) \right)_n$ ist monoton wachsend und nach Voraussetzung durch g beschränkt. Also ist der punktweise Grenzwert $f := \sup_k f_k(x)$ dieser Folge nach Satz 3.55 integrierbar, und wegen

$$\int_\Omega f_l(x)\, d\mu(x) \leq \int_\Omega \left(\sup_{k=1,\ldots,n} f_k(x) \right) d\mu(x)$$

für $l \leq n$ und $\int_\Omega \left(\sup\limits_{k=1,\ldots,n} f_k(x) \right) d\mu(x) \to \int_\Omega f(x)\, d\mu(x)$ für $n \to \infty$ gilt die behauptete Ungleichung für das Supremum.

Ganz analog verläuft der Beweis für das Infimum. □

Satz 3.59 *Sei f_k eine Folge μ-integrierbarer Funktionen mit Werten im Banach-Raum Y, die μ-fast überall punktweise gegen eine Funktion f konvergiert und für die es eine μ-integrierbare reellwertige Funktion g gibt, so dass $|f_k(x)| \leq g(x)$ für alle $x \in \Omega$ gilt. Dann ist f integrierbar bzgl. μ und f_k konvergiert bzgl. $\|\cdot\|_1$ gegen f, so dass insbesondere gilt*

$$\int_\Omega f(x)\, d\mu(x) = \lim_{k \to \infty} \int_\Omega f_k(x)\, d\mu(x) \quad .$$

Beweis: Sei $g_k := \sup_{m,n\geq k} |f_n - f_m|$, dann ist g_k als Supremum wegen $|f_n - f_m| \leq 2g$ nach dem vorigen Korollar integrierbar.

Desweiteren ist g_k nach Definition eine monoton fallende Folge, deren Integrale durch $2\|g\|_1$ beschränkt sind, und da f_k fast überall punktweise konvergiert, konvergiert g_k fast überall punktweise gegen die Nullfunktion. Also ist g_k nach dem Satz 3.55 über monotone Konvergenz eine Cauchy-Folge und konvergiert bzgl. $\|\cdot\|_1$ gegen die Nullfunktion.

Dies bedeutet aber wegen $\|f_n - f_m\|_1 \leq \|g_k\|_1$ bei $m, n \geq k$, dass f_k eine Cauchy-Folge ist. Da f_k fast überall punktweise gegen f konvergiert, folgt nach Korollar 3.54 auch die Konvergenz gegen f bzgl. $\|\cdot\|_1$, und insbesondere ist das Integral von f der Grenzwert der Integrale von f_k. □

Eine einfache Anwendung des Satzes von Lebesgue über majorisierte Konvergenz ist das folgende Majorantenkriterium.

Korollar 3.60 *Sei $A_1 \subset A_2 \subset \dots$ eine abzählbare Ausschöpfung der Menge $A = \bigcup_{k=1}^{\infty} A_k \subset \mathbb{R}^n$ durch kompakte Mengen A_k und $f : A \to \mathbb{R}$ eine Funktion, die über jedes A_k Lebesgue-integrierbar ist (solche f werden lokal Lebesgue-integrierbar über A genannt). Gibt es eine über A Lebesgue-integrierbare Funktion $g : A \to \mathbb{R}$ mit $|f(x)| \leq g(x)$ für alle $x \in A$, dann ist auch f über ganz A Lebesgue-integrierbar.*

Beweis: Die Folge $f|_{A_k}$ konvergiert punktweise gegen $f|_A$. Jede der Funktionen $f|_{A_k}$ ist Lebesgue-integrierbar und es gilt $|f|_{A_k}| \leq g$, also ist nach dem Satz von Lebesgue auch $f|_A$ Lebesgue-integrierbar. □

Das nächste Korollar ist eine nützliche Folgerung für fast überall stetige Funktionen.

Korollar 3.61 *Sei $f : U \to \mathbb{R}$ fast überall stetig auf der offenen Menge $U \subset \mathbb{R}^n$ und es gebe eine über U Lebesgue-integrierbare Funktion $g : U \to \mathbb{R}$ mit $|f(x)| \leq g(x)$ für alle $x \in U$. Dann ist auch f über U Lebesgue-integrierbar.*

Beweis: Wir müssen nur zeigen, dass f über kompakte Teilmengen K von U integrierbar ist, da U durch abzählbar viele kompakte Teilmengen ausgeschöpft werden kann. Betrachte dazu die Folge $f_k := \min(f|_K, k)$. Dann ist f_k beschränkt und ihre Einschränkung auf K fast überall stetig, also ist f_k über K Lebesgue-integrierbar. Da $f|_K$ der punktweise Grenzwert der f_k ist, die wegen $|f_k| \leq g$ durch g majorisiert werden, ist auch f über K integrierbar. □

Eine schöne Anwendung des Satzes von Lebesgue über majorisierte Konvergenz ist auch das folgende Korollar.

Korollar 3.62 *Ist $f : [a, b] \to \mathbb{R}$ differenzierbar mit beschränkter Ableitung, dann ist f' Lebesgue-integrierbar über $[a, b]$ und es gilt $f(b) - f(a) = \int_{[a,b]} f'(x)\, d\mu(x)$*

Beweis: Die durch $f_k(x) := k(f(x + \frac{1}{k}) - f(x))$ für $x \in [a, b - \frac{1}{k}]$ und $f_k(x) := 0$ sonst definierte Folge von Funktionen konvergiert auf $[a, b)$ punktweise gegen $f'(x)$ und ist gleichmäßig beschränkt durch $\sup_{x\in[a,b]} |f'(x)|$. Als stückweise stetige Funktion ist f_k sowohl Riemann-integrierbar als auch Lebesgue-integrierbar, und es gilt für das Riemann-Integral

$$\int_a^b f_k(x)\,dx = k\left(\int_{b-\frac{1}{k}}^b f(x)\,dx - \int_a^{a+\frac{1}{k}} f(x)\,dx\right) \to f(b) - f(a)$$

bei $k \to \infty$ nach dem Mittelwertsatz der Integralrechnung für stetige Funktionen.

Als gleichmäßig beschränkte Folge von Lebesgue-integrierbaren Funktionen auf der Menge $[a,b]$ endlichen Maßes erfüllt f_k aber auch die Voraussetzungen des Satzes über majorisierte Konvergenz, und somit ist die punktweise Grenzfunktion f' von f_k Lebesgue-integrierbar über $[a,b]$ mit

$$\int_{[a,b]} f'(x)\,d\mu(x) = \lim_{k\to\infty}\int_a^b f_k(x)\,dx = f(b) - f(a) \quad .$$

□

Parameterabhängige Lebesgue-Integrale Der Satz über die majorisierte Konvergenz erlaubt darüberhinaus eine Erweiterung der Aussagen über parameterabhängige Integrale.

Satz 3.63 *Sei μ ein Maß auf einer σ-Algebra von Teilmengen von Ω, (X,d) ein metrischer Raum und Y ein Banach-Raum. Sei $f : X \times \Omega \to Y$ eine Funktion, für die $t \mapsto f(x,t)$ für jedes $x \in X$ integrierbar bzgl. μ und $x \mapsto f(x,t)$ für fast alle $t \in \Omega$ stetig ist, und $x^* \in X$ ein Punkt, für den es eine Umgebung U und eine μ-integrierbare reellwertige Funktion g gibt mit $|f(x,t)| \le g(x)$ für alle $x \in U$ und $t \in \Omega$.*

Dann ist die durch $F(x) := \int_\Omega f(x,t)\,d\mu(t)$ definierte Funktion $F : X \to Y$ stetig im Punkt x^*, und auch die durch $x \mapsto f(x,\cdot)$ definierte Abbildung von X nach $L^1_\mu(\Omega, Y)$ ist stetig in x^*.

Beweis: Ist x_n eine gegen x^* konvergente Folge in X, dann wende man einfach den Satz über majorisierte Konvergenz auf die fast überall punktweise gegen $f(x^*,\cdot)$ konvergente Folge $f(x_n,\cdot)$ an. □

Ein Analogon gilt auch für die Differentiation unter dem Lebesgue-Integral.

Satz 3.64 *Sei $U \subset \mathbb{R}^m$ eine offene Teilmenge und Ω eine Lebesgue-meßbare Teilmenge des $\mathbb{R}^n$. Sei $f : U \times \Omega \to Y$ eine Funktion, für die $t \mapsto f(x,t)$ für jedes $x \in U$ Lebesgue-integrierbar und $x \mapsto f(x,t)$ für jedes $t \in \Omega$ stetig differenzierbar ist, und für die es eine Lebesgue-integrierbare reellwertige Funktion g gibt mit $|\frac{\partial f}{\partial x_i}(x,t)| \le g(t)$ für alle $x \in U$, $t \in \Omega$ und $i = 1,\ldots,n$.*

Dann ist die durch $F(x) := \int_\Omega f(x,t)\,d\mu(t)$ definierte Funktion F stetig differenzierbar auf U mit partiellen Ableitungen $\frac{\partial F}{\partial x_i}(x) = \int_\Omega \frac{\partial f}{\partial x_i}(x,t)\,d\mu(t)$.

Beweis: Für $x \in U$ und eine Nullfolge $0 \neq h_k \to 0$ betrachte man die Folge von Funktionen $\phi_k(t) := \frac{f(x+h_k e_i,t)-f(x,t)}{h_k}$. Da mit $f(x,\cdot)$ auch $\phi_k(\cdot)$ für genügend großes k integrierbar über Ω ist, für jedes $t \in \Omega$ die punktweise Konvergenz

$$\lim_{k\to\infty} \phi_k(t) = \frac{\partial f}{\partial x_i}(x,t)$$

erfüllt ist und nach dem Schrankensatz auch $|\phi_k| \leq g$ mit der integrierbaren Funktion g gilt, kann man den Satz über majorisierte Konvergenz anwenden und erhält, dass $\frac{\partial f}{\partial x_i}(x,t)$ integrierbar ist und

$$\lim_{k\to\infty} \frac{F(x+h_k)-F(x)}{h_k} = \lim_{k\to\infty} \int_\Omega \phi_k(t)\,d\mu(t) = \int_\Omega \frac{\partial f}{\partial x_i}(x,t)\,d\mu(t)$$

gilt. Also ist F partiell differenzierbar und es gilt die im Satz angegebene Formel. Nun ist aber $\frac{\partial f}{\partial x_i}(x,t)$ auch noch stetig in x, also ist die partielle Ableitung von F nach Satz 3.63 auch stetig. □

Beispiel 3.65 *Ist $\rho : K \to \mathbb{R}$ eine integrierbare Funktion auf einer kompakten Menge $K \subset \mathbb{R}^3$, die die Masse-Dichte eines Körpers modelliert, so ist das Potential u der Anziehungskraft* grad *u dieses Körpers im Punkt $x \in \mathbb{R}^3 \setminus K$ bis auf einen Faktor durch*

$$u(x) := \int_K \frac{\rho(y)}{\|x-y\|_{Euklid}}\,d\mu(y)$$

gegeben, und u ist in $\mathbb{R}^3 \setminus K$ harmonisch, d.h. $\Delta u = 0$ gilt mit dem Laplace-Operator $\Delta u := \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \frac{\partial^2 u}{\partial x_3^2}$ auf dem $\mathbb{R}^3$.

Tatsächlich, u ist eine C^2-Funktion auf jeder offenen Teilmenge $U \subset \mathbb{R}^n \setminus K$ vom Mindestabstand $d := \operatorname{dist}(\overline{U},K) > 0$, denn mit dem Euklidischen Abstand $r := \|x-y\|_{Euklid}$ der Punkte x und y gilt für $f(x,y) := \frac{\rho(y)}{\|x-y\|_{Euklid}}$

$$\frac{\partial f}{\partial x_i}(x,y) = -\frac{\rho(y)}{\|x-y\|^3_{Euklid}}(x_i-y_i)$$
$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x,y) = -\frac{\rho(y)}{\|x-y\|^3_{Euklid}}\delta_{ij} + 3\frac{\rho(y)}{\|x-y\|^5_{Euklid}}(x_i-y_i)(x_j-y_j)$$

mit dem durch $\delta_{ii} := 1$ und $\delta_{ij} := 0$ bei $i \neq j$ definierten Kronecker-Delta. Also folgt aus

$$\left|\frac{\partial f}{\partial x_i}(x,y)\right| \leq \frac{1}{d^2}|\rho(y)|$$
$$\left|\frac{\partial^2 f}{\partial x_j \partial x_i}(x,y)\right| \leq \frac{4}{d^3}|\rho(y)|$$

und der Integrierbarkeit von ρ mit Satz 3.64, dass sowohl u als auch $\frac{\partial u}{\partial x_i}$ stetig differenzierbar auf U sind, und Δu verschwindet wegen

$$\Delta u = \int_K \left(-3\frac{\rho(y)}{\|x-y\|^3_{Euklid}} + 3\frac{\rho(y)}{\|x-y\|^5_{Euklid}}\sum_{i=1}^3 (x_i-y_i)^2\right) d\mu(y) = 0 \quad .$$

Riemann-Integral vs. Lebesgue-Integral im Eindimensionalen In diesem Abschnitt wollen wir diskutieren, inwieweit das Integral bzgl. des Lebesgue-Maßes auf $\mathbb{R}$, d.h. das Lebesgue-Integral im Eindimensionalen, mit dem eigentlichen und uneigentlichen Riemann-Integral übereinstimmt.

Zunächst einmal sei daran erinnert, dass das eigentliche Riemann-Integral nur für beschränkte Funktionen existieren kann. Denn nur für beschränkte Funktionen kann man überhaupt Ober- und Unterintegral bilden, und das eigentliche Riemann-Integral existiert genau dann, wenn Ober- und Unterintegral zusätzlich noch den gleichen Wert haben.

Satz 3.66 *Eine beschränkte Funktion $f : [a, b] \to \mathbb{R}$ ist genau dann Riemann-integrierbar, wenn die Menge ihrer Unstetigkeitsstellen eine Lebesgue-Nullmenge ist, und dann stimmen Riemann- und Lebesgue-Integral überein.*

Beweis: Die Ober- bzw. Untersummen von f kann man als das Lebesgue-Integral von Treppenfunktionen ansehen. Hat man nun eine Folge Z_n von immer feiner werdenden Zerlegungen von $[a, b]$, deren Feinheit gegen Null konvergiert, so gehört zu den Untersummen U_n eine monoton wachsende Folge g_n und zu den Obersummen O_n eine monoton fallende Folge h_n von Treppenfunktionen mit $g_n(x) \leq f(x) \leq h_n(x)$ für alle $x \in [a, b]$. Die punktweisen Grenzfunktionen $g := \lim_{n\to\infty} g_n$ und $h := \lim_{n\to\infty} h_n$ existieren aufgrund der Monotonie und Beschränktheit von f und sind selbst beschränkt.

Also liefert für Riemann-integrierbares f der Satz von der majorisierten Konvergenz

$$\int_{\mathbb{R}^n} g(x)\, d\mu(x) = \lim_{n\to\infty} U_n = \int_a^b f(x)\, dx = \lim_{n\to\infty} O_n = \int_{\mathbb{R}^n} h(x)\, d\mu(x) \quad .$$

Somit gilt wegen $g \leq h$ auch $\|h - g\|_1 = \int_{\mathbb{R}^n} h(x) - g(x)\, d\mu(x) = 0$. Also ist $h = g$ fast überall und somit $f = g = h$ fast überall wegen $g \leq f \leq h$. Daher ist f Lebesgue-integrierbar, es gilt

$$\int_{\mathbb{R}^n} f(x)\, d\mu(x) = \int_{\mathbb{R}^n} g(x)\, d\mu(x) = \int_a^b f(x)\, dx \quad ,$$

und die Menge der Unstetigkeitsstellen von f ist in der Nullmenge enthalten, die aus den Punkten der Zerlegungen Z_n und den Punkten x mit $g(x) < h(x)$ besteht.

Ist umgekehrt die Menge der Unstetigkeitsstellen der Funktion f eine Nullmenge, so gilt $f = g = h$ fast überall, und der Satz von der majorisierten Konvergenz liefert

$$\lim_{n\to\infty} U_n = \int_{\mathbb{R}^n} g(x)\, d\mu(x) = \int_{\mathbb{R}^n} h(x)\, d\mu(x) = \lim_{n\to\infty} O_n \quad .$$

Also haben Unter- und Oberintegral denselben Wert, d.h. f ist Riemann-integrierbar.
$\square$

Beispiel 3.67 *Die Funktion $1_{\mathbb{Q}}$ ist in jedem Punkt unstetig, also insbesondere nicht Riemann-integrierbar über $[0, 1]$, wohl aber Lebesgue-integrierbar mit $\int_{[0,1]} 1_{\mathbb{Q}}(x)\, d\mu(x) = 0$, wie wir auch schon in Beispiel 3.46 gesehen hatten.*

Dagegen ist die durch $f(x) := \frac{1}{q}$ für rationales $x \in \mathbb{Q}$, falls $x = \frac{p}{q}$ die gekürzte Bruchdarstellung von $x \in \mathbb{Q}$ ist, und $f(x) := 0$ für $x \in \mathbb{R} \setminus \mathbb{Q}$ definierte Funktion $f : [0,1] \to \mathbb{R}$ sowohl Lebesgue- als auch Riemann-integrierbar mit $\int_{[0,1]} f(x)\, d\mu(x) = 0 = \int_0^1 f(x)\, dx$, denn f ist nur in den rationalen Punkten aus $[0,1]$ unstetig, die eine Lebesgue-Nullmenge sind.

Satz 3.68 *Eine über alle kompakten Teilmengen eines Intervalls (a,b) Riemann-integrierbare Funktion $f : (a,b) \to \mathbb{R}$ ist genau dann Lebesgue-integrierbar, wenn ihr Betrag $|f|$ uneigentlich Riemann-integrierbar über I ist, und dann stimmt das uneigentliche Riemann-Integral mit dem Lebesgue-Integral überein.*

Beweis: Gelte $a \neq a_n \searrow a$, $b \neq b_n \nearrow b$. Nach Satz 3.66 ist f Lebesgue-integrierbar über $[a_n, b_n]$ und es gilt

$$\int_{a_n}^{b_n} |f(x)|\, dx = \int_{[a_n,b_n]} |f(x)|\, d\mu(x)\,.$$

Ist nun $|f|$ uneigentlich Riemann-integrierbar (a,b), so ist der Grenzwert der linken Seite für $n \to \infty$ endlich, und somit ist die Funktion f nach Korollar 3.56 Lebesgue-integrierbar über (a,b) und ihr Lebesgue-Integral ist das uneigentliche Riemann-Integral

$$\int_{[a,b]} f(x)\, d\mu(x) = \lim_{n\to\infty} \int_{[a_n,b_n]} f(x)\, d\mu(x) \lim_{n\to\infty} \int_{a_n}^{b_n} f(x)\, dx\,.$$

Ist umgekehrt f Lebesgue-integrierbar über (a,b), dann hat $\int_{[a_n,b_n]} |f(x)|\, d\mu(x)$ für $n \to \infty$ einen endlichen Grenzwert und somit auch $\int_{a_n}^{b_n} |f(x)|\, dx$, d.h. $|f|$ ist uneigentlich Riemann-integrierbar über (a,b). □

Beispiel 3.69 *Die Funktion $f(x) := \frac{\sin(x)}{x}$ ist uneigentlich Riemann-integrierbar über $(0,\infty)$, aber $|f|$ ist nicht uneigentlich Riemann-integrierbar über $(0,\infty)$, so dass f über $(0,\infty)$ nicht Lebesgue-integrierbar ist.*

3.5 Der Satz von Fubini

Zwar haben wir das Lebesgue-Integral für eine große Klasse von Funktionen definiert und nützliche Eigenschaften herausgefunden, können es aber außer im Eindimensionalen mit Hilfe des eigentlichen und uneigentlichen Riemann-Integrals oder im Mehrdimensionalen für stetige Funktionen mit kompaktem Träger mittels des iterierten Riemann-Integrals praktisch noch nicht gut ausrechnen.

Abhilfe schafft hier der Satz von Fubini, der die Berechnung des Integrals bzgl. eines Produktmaßes auf die Integration über die einzelnen Faktoren zurückführt, also im Fall des n-dimensionalen Lebsgue-Maßes die Integration auf die Berechnung n eindimensionaler Integrale reduziert.

Produkt-Maße und Produkt-σ-Algebren Zunächst beschäftigen wir und mit σ-Algebren auf Produkten.

Definition 3.70 *Seien $\mathcal{H}$ bzw. $\mathcal{H}'$ Halbringe auf den Mengen Ω bzw. Ω', dann bezeichnet man die vom Halbring*

$$\mathcal{H} \times \mathcal{H}' := \{A \times B \mid A \in \mathcal{H}\,,\, B \in \mathcal{H}'\}$$

auf $\Omega \times \Omega'$ erzeugte σ-Algebra $\sigma(\mathcal{H} \times \mathcal{H}')$ als Produkt-σ-Algbra.

Man bemerke, dass $\sigma(\mathcal{H}) \times \sigma(\mathcal{H}')$ nicht selbst wieder eine σ-Algebra ist, denn Vereinigungen von Produkten sind nicht wieder Produkte. Stattdessen gilt das folgende Lemma.

Lemma 3.71 *Es gilt $\sigma(\sigma(\mathcal{H}) \times \sigma(\mathcal{H}')) = \sigma(\mathcal{H} \times \mathcal{H}')$.*

Beweis: Offensichtlich ist $\mathcal{H}\times\mathcal{H}' \subset \sigma(\mathcal{H})\times\sigma(\mathcal{H}')$ und somit nur noch $\sigma(\sigma(\mathcal{H})\times\sigma(\mathcal{H}')) \subset \sigma(\mathcal{H} \times \mathcal{H}')$ zu zeigen.

Für festes $B \in \mathcal{H}'$ betrachte dazu die von den Mengen $A \times B$ ($A \in \mathcal{H}$) erzeugte σ-Algebra $\sigma(\mathcal{H}) \times B$ auf $\Omega \times B$. Diese ist in $\sigma(\mathcal{H} \times \mathcal{H}')$ enthalten. Analog umgekehrt, d.h. es gelten die Inklusionen

$$\sigma(\mathcal{H}) \times B, A \times \sigma(\mathcal{H}') \subset \sigma(\mathcal{H} \times \mathcal{H}')$$

für alle $A \in \mathcal{H}$, $B \in \mathcal{H}'$. Somit gilt aber auch $\sigma(\mathcal{H}) \times \sigma(\mathcal{H}') \subset \sigma(\mathcal{H} \times \mathcal{H}')$, und daher ist auch die von der linken Seite erzeugte σ-Algebra in der rechten Seite enthalten. $\square$

Beispiel 3.72 *Für die Borel-σ-Algebren auf dem $\mathbb{R}^n$ gilt $\sigma(\mathcal{B}(\mathbb{R}^m)\times\mathcal{B}(\mathbb{R}^n)) = \mathcal{B}(\mathbb{R}^{m+n})$, denn die halboffenen Quader im $\mathbb{R}^{m+n}$ sind gerade Produkte von halboffenen Quadern im $\mathbb{R}^m$ und $\mathbb{R}^n$.*

Dagegen gilt für die Lebesgue-σ-Algebren $\sigma(\mathcal{L}(\mathbb{R}^m) \times \mathcal{L}(\mathbb{R}^n)) \subsetneq \mathcal{L}(\mathbb{R}^{m+n})$, denn nicht alle Nullmengen des $(n+m)$-dimensionalen Lebesgue-Maßes sind auf der linken Seite enthalten. Fügt man diese aber hinzu, so ergibt sich ganz $\mathcal{L}(\mathbb{R}^{m+n})$.

Seien nun μ bzw. ν Maße auf den σ-Algebren $\sigma(\mathcal{H})$ bzw. $\sigma(\mathcal{H}')$ von Teilmengen von Ω bzw. Ω'.

Um ein Maß $\mu \times \nu$ auf dem Produkt $\Omega \times \Omega'$ mit der Eigenschaft $(\mu \times \nu)(A \times B) := \mu(A)\nu(B)$ für alle $A \times B \in \mathcal{H} \times \mathcal{H}'$ zu gewinnen (ein Maß mit dieser Eigenschaft nennt man Produkt-Maß), beobachte man, dass für eine Treppenfunktion g auf $\Omega \times \Omega'$ zu Mengen aus $\mathcal{H} \times \mathcal{H}$ (oder dem davon erzeugten Ring) und ein festes $x \in \Omega$ die Funktion $y \mapsto g(x,y)$ eine Treppenfunktion auf Ω' zu Mengen aus $\mathcal{H}'$ ist (oder dem davon erzeugten Ring). Tatsächlich, für $g := \sum_{i=1}^{k} g_i 1_{A_i \times B_i}$ gilt die Gleichung

$$g(x,\cdot) = \sum_{i=1}^{k} (g_i 1_{A_i}(x))\, 1_{B_i}(\cdot) \quad .$$

Daher können wir für festes $x \in \Omega$ das Integral $\int_{\Omega'} g(x,y)\,d\nu(y)$ bilden und erhalten

$$\int_{\Omega'} g(x,y)\,d\nu(y) = \sum_{i=1}^{k} (g_i 1_{A_i}(x))\,\nu(B_i) = \sum_{i=1}^{k} (g_i \nu(B_i))\,1_{A_i}(x) \quad .$$

Also ist die Abbildung $x \mapsto \int_{\Omega'} g(x,y)\,d\nu(y)$ wiederum eine Treppenfunktion auf Ω zu Mengen aus $\mathcal{H}$, und somit können wir ihr μ-Integral

$$\int_{\Omega} \left(\int_{\Omega'} g(x,y)\,d\nu(y) \right) d\mu(x) = \sum_{i=1}^{k} g_i \mu(A_i)\nu(B_i)$$

bilden. Hätten wir erst nach x und dann nach y integriert, so hätte sich natürlich derselbe Wert ergeben, daher gilt für Treppenfunktionen auf $\Omega \times \Omega'$ zu Mengen aus $\mathcal{H} \times \mathcal{H}$ (oder dem davon erzeugten Ring) die Gleichung

$$\int_{\Omega} \left(\int_{\Omega'} g(x,y)\,d\nu(y) \right) d\mu(x) = \int_{\Omega'} \left(\int_{\Omega} g(x,y)\,d\mu(x) \right) d\nu(y) \quad .$$

Schränkt man das so definierte iterierte Integral auf charakteristische Funktionen zu Mengen aus dem von $\mathcal{H} \times \mathcal{H}$ erzeugten Ring ein, so erhält man wegen der Linearität des iterierten Integrals einen eindeutigen Inhalt $\mu \times \nu$ auf dem von $\mathcal{H} \times \mathcal{H}$ erzeugten Ring mit der gewünschten Produkt-Eigenschaft

$$(\mu \times \nu)(A \times B) = \mu(A)\nu(B) \quad .$$

Es stellt sich die Frage, ob man diesen Inhalt $\mu \times \nu$ eindeutig zu einem Maß fortsetzen kann.

Satz 3.73 *Seien μ bzw. ν σ-endliche Maße auf den σ-Algebren $\mathcal{A}$ bzw. $\mathcal{B}$ von Teilmengen von Ω bzw. Ω'. Dann gibt es genau ein Maß $\mu \times \nu$ auf $\sigma(\mathcal{A} \times \mathcal{B})$ mit*

$$(\mu \times \nu)(A \times B) = \mu(A)\nu(B)$$

für alle $A \in \mathcal{A}$ und $B \in \mathcal{B}$.

Beweis: Nach dem Fortsetzungssatz von Carathéodory 3.33 müssen wir nur zeigen, dass der Inhalt $\mu \times \nu$ auf dem Ring, der von den Produkten $A \times B$ von Mengen A, B endlichen Maßes erzeugt wird, ein Prämaß oder äquivalenterweise σ-additiv ist. Dann nämlich kann man den Inhalt $\mu \times \nu$, der ja schon die Produkt-Eigenschaft besitzt, eindeutig zu einem Maß fortsetzen.

Sei also $Q_1 \subset Q_2 \subset \ldots$ eine aufsteigende Folge von Mengen endlichen Maßes aus diesem Ring, für die $Q := \bigcup_{k=1}^{\infty} Q_k$ wieder ein Element des Ringes ist, d.h. insbesondere $(\mu \times \nu)(Q)$ endlich ist. Dann ist die Folge der charakteristischen Funktionen 1_{Q_k} und insbesondere für festes $x \in \Omega$ auch die Folge der Funktionen $1_{Q_k}(x, \cdot)$ monoton wachsend. Aufgrund der Beschränktheit der Integrale durch $(\mu \times \nu)(Q)$ konvergiert nach dem

Satz 3.55 von Beppo-Levi die Folge der Integrale $\int_{\Omega'} 1_{Q_k}(x,y)\,d\nu(y)$ für jedes $x \in \Omega$ gegen $\int_{\Omega'} 1_Q(x,y)\,d\nu(y)$. Wendet man den Satz nun nocheinmal für die Integration bzgl. μ an, so erhält man schließlich die Gleichung

$$\lim_{k\to\infty} (\mu \times \nu)(Q_k) = \lim_{k\to\infty} \int_\Omega \left(\int_{\Omega'} 1_{Q_k}(x,y)\,d\nu(y) \right) d\mu(x) =$$
$$\int_\Omega \left(\int_{\Omega'} 1_Q(x,y)\,d\nu(y) \right) d\mu(x) = (\mu \times \nu)(Q)\,,$$

d.h. die Stetigkeit von $\mu \times \nu$ von unten, die äquivalent zur σ-Additivität ist. □

Da der Beweis den Fortsetzungssatz von Carathéodory 3.33 verwendet, erhält man sogar die Existenz eines eindeutigen Maßes mit der Produkteigenschaft auf der σ-Algebra der $(\mu \times \nu)^*$-meßbaren Mengen. Speziell für das Lebesgue-Maß ergibt sich das folgende Korollar.

Korollar 3.74 *Das Lebesgue-Maß auf dem $\mathbb{R}^{m+n}$ ist das Produkt der Lebesgue-Maße auf dem $\mathbb{R}^m$ und dem $\mathbb{R}^n$.*

Der Satz von Fubini Da das Lebesgue-Maß auf dem $\mathbb{R}^{m+n}$ das Produkt der Lebesgue-Maße auf dem $\mathbb{R}^m$ und dem $\mathbb{R}^n$ ist, gilt für Treppenfunktionen g auf dem $\mathbb{R}^{n+m}$ zu Mengen endlichen Lebesgue-Maßes insbesondere

$$\int_{\mathbb{R}^{m+n}} g(x,y)\,d\mu_{n+m}(x,y) = \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} g(x,y)\,d\mu_n(y) \right) d\mu_m(x) = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} g(x,y)\,d\mu_m(y) \right) d\mu_n(x)$$

wobei der Index die Dimension des betrachteten Lebesgue-Maßes angibt. Der Inhalt des Satzes von Fubini ist, dass diese Gleichung auch für beliebige Lebesgue-integrierbare Funktionen auf dem Produktraum gilt. Wir formulieren und beweisen den Satz von Fubini gleich für allgemeine Produktmaße, brauchen dazu aber ein vorbereitendes Lemma.

Lemma 3.75 *Ist $Z \subset \Omega \times \Omega'$ eine $(\mu \times \nu)$-Nullmenge, dann ist für μ-fast alle $x \in \Omega$ die Menge $Z_x := \{y \in \Omega' \,|\, (x,y) \in Z\}$ eine ν-Nullmenge.*

Beweis: Sei S_n die Menge der $x \in \Omega$ mit $\nu(Z_x) \geq \frac{1}{n}$ und $S := \bigcup_{n=1}^{\infty} S_n$. Dann brauchen wir nur zu zeigen, dass S in einer μ-Nullmenge enthalten ist. Da Z eine Nullmenge ist, gibt es zu $\epsilon > 0$ eine Folge Q_k halboffener Quader mit $Z \subset \bigcup_{k=1}^{\infty} Q_k$ und $\sum_{k=1}^{\infty} (\mu \times \nu)(Q_k) \leq \frac{1}{n2^n}\epsilon$.

Sei T_n die Menge aller $x \in \Omega$ mit $\frac{1}{n} \leq \sum_{k=1}^{\infty} \nu((Q_k)_x)$, dann ist T_n Lebesgue-meßbar mit $S_n \subset T_n$ wegen $Z_x \subset \bigcup_{k=1}^{\infty} (Q_k)_x$. Desweiteren ist die Funktion $x \mapsto \sum_{k=1}^{\infty} \nu((Q_k)_x)$ integrierbar bzgl. μ und liefert die Ungleichung

$$\frac{1}{n}\mu(T_n) \leq \sum_{k=1}^{\infty} \int_\Omega \nu((Q_k)_x)\,d\mu(x) = \sum_{k=1}^{\infty} (\mu \times \nu)(Q_k) < \frac{1}{n2^n}\epsilon\,.$$

Also gilt $\mu(T_n) \leq 2^{-n}\epsilon$, d.h.

$$\mu(S) \leq \sum_{n=1}^{\infty} \mu(S_n) \leq \sum_{n=1}^{\infty} \mu(T_n) \leq \epsilon\,.$$

Da diese Ungleichung für jedes $\epsilon > 0$ gilt, ist S eine μ-Nullmenge. □

Satz 3.76 *Ist die Funktion f auf $\Omega\times\Omega'$ integrierbar bzgl. des σ-endlichen Produktmaßes $\mu \times \nu$, dann ist für μ-fast alle $x \in \Omega$ die Funktion $y \mapsto f(x,y)$ bzgl. ν integrierbar, und deren ν-Integrale sind als Funktion $x \mapsto \int_{\Omega'} f(x,y)\,d\nu(y)$ auch μ-integrierbar mit Integralwert*

$$\int_{\Omega\times\Omega'} f(x,y)\,d(\mu\times\nu)(x,y) = \int_{\Omega}\left(\int_{\Omega'} f(x,y)\,d\nu(y)\right)\,d\mu(x) = \int_{\Omega'}\left(\int_{\Omega} f(x,y)\,d\mu(x)\right)\,d\nu(y) \quad .$$

Beweis: Nach Definition gibt es für jede $(\mu \times \nu)$-integrierbare Funktion f auf $\Omega \times \Omega'$ eine Folge f_k von Treppenfunktionen zu Produktmengen endlichen Maßes, die sowohl bzgl. der $L^1_{\mu\times\nu}$-Halbnorm auf dem Produkt $\Omega \times \Omega'$ als auch punktweise außerhalb einer $(\mu \times \nu)$-Nullmenge $Z \subset \Omega \times \Omega'$ fast überall gegen f konvergiert. Nach Lemma 3.75 gibt es zu Z eine μ-Nullmenge $S \subset \Omega$ mit $\nu(Z_x) = 0$ für alle $x \in S^c$.

Betrachten wir $f_k(\cdot,\cdot)$ als Funktion $f_k(\cdot)(\cdot)$ von Ω mit Werten im Raum der Treppenfunktionen $\mathrm{Step}_\nu(\Omega')$ (die ihrerseits wieder Werte in einem Banach-Raum Y haben können), desssen Vervollständigung bzgl. der L^1-Halbnorm gerade der Banach-Raum der ν-integrierbaren Funktionen ist, so ist $f_k : \Omega \to \mathrm{Step}_\nu(\Omega')$ wegen

$$\|f_k(\cdot) - f_l(\cdot)\|_1 = \int_{\Omega} |f_k(x) - f_l(x)|_{\mathrm{Step}_\nu(\Omega')}\,d\mu(x) =$$
$$\int_{\Omega}\int_{\Omega'} |f_k(x,y) - f_l(x,y)|_Y\,d\mu(x) = \|f_k(\cdot,\cdot) - f_l(\cdot,\cdot)\|_1$$

auch als solch eine $\mathrm{Step}_\nu(\Omega')$-wertige Funktion eine Cauchy-Folge.

Nach dem ersten Fundamentallemma 3.42 gibt es einerseits eine μ-Nullmenge T und eine Teilfolge f_{k_i}, so dass $f_{k_i}(x)$ für jedes $x \in T^c$ im Banach-Raum der ν-integrierbaren Funktionen konvergiert, insbesondere also eine Cauchy-Folge bzgl. der L^1_ν-Halbnorm ist. Für $x \in S^c$ konvergiert andererseits die Folge $f_{k_i}(x)(y)$ punktweise gegen $f(x)(y)$ für ν-fast alle $y \in \Omega'$. Also ist $f(x)$ für $x \in (S \cup T)^c$ nach Korollar 3.54 eine ν-integrierbare Funktion, gegen die $f_{k_i}(x)$ auch bzgl. der L^1_ν-Halbnorm konvergiert, und die Nachbemerkungen zu diesem Korollar zeigen, dass

$$\lim_{k\to\infty} \int_{\Omega'} f_k(x,y)\,d\nu(y) = \int_{\Omega'} f(x,y)\,d\nu(y)$$

sogar für die gesamte Folge f_k gilt.

Nun betrachte die Folge $x \mapsto \int_{\Omega'} f_k(x,y)\, d\nu(y)$ von Treppenfunktionen auf Ω zu Mengen endlichen μ-Maßes. Diese ist eine Cauchy-Folge bzgl. der L^1_μ-Halbnorm, denn es gilt

$$\| \int_{\Omega'} f_k(\cdot,y)\, d\nu(y) - \int_{\Omega'} f_l(\cdot,y)\, d\nu(y) \|_1 =$$

$$\int_\Omega \left| \int_{\Omega'} f_k(\cdot,y)\, d\nu(y) - \int_{\Omega'} f_l(\cdot,y)\, d\nu(y) \right|_Y d\mu(x) \leq \|f_k(\cdot,\cdot) - f_l(\cdot,\cdot)\|_1 .$$

und nach dem zuvor Bewiesenen konvergiert $x \mapsto \int_{\Omega'} f_k(x,y)\, d\nu(y)$ punktweise außerhalb der μ-Nullmenge $S \cup T$ gegen $x \mapsto \int_{\Omega'} f(x,y)\, d\nu(y)$. Daher konvergiert die Funktion $x \mapsto \int_{\Omega'} f_k(x,y)\, d\nu(y)$ nach Korollar 3.54 auch bzgl. der L^1_μ-Halbnorm gegen die Funktion $x \mapsto \int_{\Omega'} f(x,y)\, d\nu(y)$, so dass insbesondere

$$\lim_{k\to\infty} \int_\Omega \left(\int_{\Omega'} f_k(x,y)\, d\nu(y) \right) d\mu(x) = \int_\Omega \left(\int_{\Omega'} f(x,y)\, d\nu(y) \right) d\mu(x)$$

gilt. Aber für die Treppenfunktionen f_k gilt

$$\int_\Omega \left(\int_{\Omega'} f_k(x,y)\, d\nu(y) \right) d\mu(x) = \int_{\Omega\times\Omega'} f_k(x,y)\, d(\mu\times\nu)(x,y)$$

nach Definition des Produktmaßes, und da f_k bzgl. der $L^1_{\mu\times\nu}$-Halbnorm gegen f konvergiert, auch

$$\lim_{k\to\infty} \int_{\Omega\times\Omega'} f_k(x,y)\, d(\mu\times\nu)(x,y) \int_{\Omega\times\Omega'} f(x,y)\, d(\mu\times\nu)(x,y) .$$

Die letzten drei abgesetzten Gleichungen zeigen daher

$$\int_{\Omega\times\Omega'} f(x,y)\, d(\mu\times\nu)(x,y) = \int_\Omega \left(\int_{\Omega'} f(x,y)\, d\nu(y) \right) d\mu(x) \quad .$$

Die Vertauschbarkeit der Integration für Treppenfunktion liefert schließlich auch die Gleichheit der vertauschten Integrale. □

Beispiel 3.77 *Bezeichne μ das eindimensionale Lebesgue-Maß. Wir wollen das Integral der Funktion $x + y^2$ über das Dreieck $D := \{(x,y) \,|\, 0 \leq x \leq 1\,,\, x \leq y \leq 1\}$ auf zwei verschiedene Weisen berechnen.*

Integration erst über y und dann über x ergibt

$$\int_{[0,1]} \left(\int_{[x,1]} x + y^2\, d\mu(y) \right) d\mu(x) = \int_{[0,1]} \left(\int_x^1 x + y^2\, dy \right) d\mu(x) = \int_{[0,1]} \left(xy + \frac{1}{3}y^3 \right) |_{y=x}^1\, d\mu(x) =$$

$$\int_{[0,1]} \left(x + \frac{1}{3} - x^2 - \frac{1}{3}x^3 \right) d\mu(x) = \int_0^1 \left(x + \frac{1}{3} - x^2 - \frac{1}{3}x^3 \right) dx = \frac{1}{2} + \frac{1}{3} - \frac{1}{3} - \frac{1}{12} = \frac{5}{12} \quad .$$

Andererseits kann man D auch als $D = \{(x,y) \,|\, 0 \leq y \leq 1\,,\, 0 \leq x \leq y\}$ darstellen, und die entsprechende Integration erst über x und dann über y liefert

$$\int_{[0,1]} \left(\int_{[0,y]} x + y^2\, d\mu(x) \right) d\mu(y) = \int_{[0,1]} \left(\int_0^y x + y^2\, dx \right) d\mu(y) = \int_{[0,1]} \left(\frac{1}{2}x^2 + xy^2 \right) |_{x=0}^y\, d\mu(y) =$$

$$\int_{[0,1]} \left(\frac{1}{2}y^2 + y^3 \right) d\mu(y) = \int_0^1 \left(\frac{1}{2}y^2 + y^3 \right) dy = \frac{1}{6} + \frac{1}{4} = \frac{5}{12} \quad ,$$

d.h. die beiden Integrale stimmen überein, wie es der Satz von Fubini auch voraussagt.

Als Folgerung ergibt sich aus dem Satz von Fubini durch Anwendung auf die charakteristische Funktion 1_A einer $(\mu \times \nu)$-meßbaren Teilmenge $A \subset \Omega \times \Omega'$ das Prinzip von Cavalieri zur Berechnung von Volumina.

Korollar 3.78 *Für eine $(\mu \times \nu)$-meßbare Teilmenge $A \subset \Omega \times \Omega'$ gilt*

$$(\mu \times \nu)(A) = \int_\Omega \nu(A_x)\, d\mu(x)$$

Insbesondere kann man das Prinzip von Cavalieri anwenden, falls μ das eindimensionale Lebesgue-Maß und ν das $(n-1)$-dimensionale Lebesgue-Maß ist, wobei man $\nu(A_x)$ dann iterativ mit Hilfe des $(n-2)$-dimensionalen Lebesgue-Maßes berechnen kann.

Beispiel 3.79

- *Sei $B \subset \mathbb{R}^{n-1}$ kompakt, dann heißt $Z := B \times [0,h] \subset \mathbb{R}^n$ Zylinder der Höhe h zur Basis B. Nach dem Prinzip von Cavalieri ist das n-dimensionale Maß von Z gerade*
$$\mu_n(Z) = \int_{[0,h]} \mu_{n-1}(Z_x)\, d\mu(x) = h\mu_{n-1}(B)$$
wegen $Z_x = B$ für jedes $x \in [0,h]$.

- *Sei $\Delta_n := \{x \in \mathbb{R}^n \,|\, x_1, \dots, x_n \geq 0\,,\; x_1 + \dots + x_n \leq 1\}$ das Standardsimplex im $\mathbb{R}^n$. Der Schnitt von Δ_n mit der durch $x_n = h$ bei $h \in [0,1]$ gegebenen Hyperebene ist die Menge $(1-h)\Delta_{n-1} \times \{h\}$. Da nach dem Transformationssatz für lineare Abbildungen*
$$\mu_{n-1}((1-h)\Delta_{n-1}) = (1-h)^{n-1}\mu_{n-1}(\Delta_{n-1})$$
gilt, ergibt sich also iterativ
$$\mu_n(\Delta_n) = \int_{[0,1]} (1-h)^{n-1}\mu_{n-1}(\Delta_{n-1})\, d\mu_1(h) = \mu_{n-1}(\Delta_{n-1}) \int_0^1 (1-h)^{n-1}\, dh =$$
$$\mu_{n-1}(\Delta_{n-1})\frac{1}{n} = \mu_{n-2}(\Delta_{n-2})\frac{1}{n(n-1)} = \dots = \frac{1}{n!} \quad .$$

Der Satz von Tonelli Die Voraussetzung $f \in L^1_{\mu\times\nu}(\Omega \times \Omega')$ des Satzes von Fubini ist oft nicht so leicht zu prüfen, wenn man nur die iterierten Integrale ausrechnet. Hier schafft der Satz von Tonelli Abhilfe, mit dem man die Anwendung des Satzes von Fubini in vielen Fällen rechtfertigen kann.

Satz 3.80 *Sei die Funktion f bzgl. der Produkt-σ-Algebra des σ-endlichen Produktmaßes $(\mu\times\nu)$ auf $\Omega\times\Omega'$ lokal integrierbar (oder auch nur meßbar). Ist für μ-fast alle $x \in \Omega$ die Funktion $f(x,\cdot)$ integrierbar bzgl. ν und ist auch $x \mapsto \int_{\Omega'} |f(x,y)|\, d\nu(y)$ integrierbar bzgl. μ, dann ist f schon integrierbar bzgl. des Produktmaßes $\mu \times \nu$.*

Beweis: Es reicht zu zeigen, dass $|f|$ integrierbar bzgl. $\mu \times \nu$ ist, also können wir ohne Einschränkung $f \geq 0$ annehmen. Sei dazu f_n eine monoton wachsende Folge von nichtnegativen Treppenfunktionen zu Mengen endlichen Maßes, die punktweise gegen f konvergiert. Dann ist $f_n(x, \cdot)$ für jedes $x \in \Omega$ monoton wachsend und konvergiert punktweise gegen $f(x, \cdot)$. Da außerhalb einer μ-Nullmenge von Punkten x sowohl $f(x, \cdot)$ integrierbar bzgl. ν ist als auch $f_n(x, \cdot)$ eine Treppenfunktion zu Mengen endlichen ν-Maßes ist, konvergiert für solche x nach dem Satz über monotone Konvergenz $f_n(x, \cdot)$ sogar bzgl. der L^1_ν-Halbnorm gegen f. Insbesondere gilt außerhalb eine μ-Nullmenge von Punkten x die Konvergenz

$$\lim_{n\to\infty} \int_{\Omega'} f_n(x,y)\, d\nu(y) = \int_{\Omega'} f(x,y)\, d\nu(y) \quad .$$

Da auch $\int_{\Omega'} f_n(x,y)\, d\nu(y)$ wiederum monoton wachsend ist und punktweise gegen die Funktion $\int_{\Omega'} f(x,y)\, d\nu(y)$ konvergiert, kann man bei der Integration bzgl. μ wiederum den Satz über monotone Konvergenz anwenden und erhält

$$\lim_{n\to\infty} \int_{\Omega} \left(\int_{\Omega'} f_n(x,y)\, d\nu(y) \right) d\mu(x) = \int_{\Omega} \left(\int_{\Omega'} f(x,y)\, d\nu(y) \right) d\mu(x) \quad .$$

Nun gilt für die Treppenfunktionen f_n aber

$$\int_{\Omega} \left(\int_{\Omega'} f_n(x,y)\, d\nu(y) \right) d\mu(x) = \int_{\Omega\times\Omega'} f_n(x,y)\, d(\mu\times\nu)(x,y)\,,$$

also ist auch die Folge der Integrale der f_n über das Produkt $\Omega \times \Omega'$ beschränkt, und daher ist f als Grenzwert der f_n nach dem Satz über monotone Konvergenz bzgl. $\mu \times \nu$ über das Produkt $\Omega \times \Omega'$ integrierbar. □

Also muß man zur Benutzung des Satzes von Fubini nur sicherstellen, dass die Funktionen, deren Integrale man bei der Berechnung des iterierten Integrals ausrechnet, auch wirklich integrierbar sind, dann ist die Anwendung des Satzes von Fubini auch schon erlaubt.

Das folgende Beispiel zeigt, wie nützlich die dadurch nach dem Satz von Fubini erlaubte Vertauschung der Integrationsreihenfolge ist.

Beispiel 3.81 *Zur Berechnung des uneigentlichen Riemann-Integrals $\int_{-\infty}^{\infty} \sin(t^2)\, dt = \int_0^\infty \frac{\sin(x)}{\sqrt{x}}\, dx$ bemerke man, dass aus der in der Übung bewiesenen Gleichung $\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$ durch Substitution auch $\int_{-\infty}^{\infty} e^{-y^2 x}\, dy = \frac{\sqrt{\pi}}{\sqrt{x}}$ folgt. Bezeichne μ das eindimensionale Lebesgue-Maß, dann erhält man*

$$\int_0^\infty \frac{\sin(x)}{\sqrt{x}}\, dx = \lim_{a\to\infty} \frac{1}{\sqrt{\pi}} \int_{(0,a)} \left(\int_{\mathbb{R}} \sin(x) e^{-y^2 x}\, d\mu(y) \right) d\mu(x)\,.$$

Wegen $|\sin(x)e^{-y^2x}| \leq e^{-y^2x}$ und der Integrierbarkeit der Majorante e^{-y^2x} darf man die Integrationsreihenfolge vertauschen und erhält bei $a = 2\pi k$ ($k \in \mathbb{Z}$)

$$\int_{(0,2\pi k)} \left(\int_{\mathbb{R}} \sin(x) e^{-y^2 x}\, d\mu(y) \right) d\mu(x) = \int_{\mathbb{R}} \left(\int_{(0,2\pi k)} \sin(x) e^{-y^2 x}\, d\mu(x) \right) d\mu(y) =$$
$$\int_{\mathbb{R}} \frac{1 - e^{-2\pi k y^2}}{1 + y^4}\, d\mu(y)\,.$$

Nun konvergiert der Integrand monoton wachsend gegen $\frac{1}{1+y^4}$, also gilt

$$\int_0^\infty \frac{\sin(x)}{\sqrt{x}}\,dx = \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} \frac{1}{1+y^4}\,d\mu(y) = \frac{\sqrt{\pi}}{\sqrt{2}}\,.$$

Der Satz von Gauß in der Ebene Eine schöne Anwendung des Satzes von Fubini ist der Satz von Gauss in der Ebene. Dazu sei $\Omega \subset \mathbb{R}^2$ eine zusammenhängende offene Menge (solche Teilmengen werden Gebiet genannt), die von einer stückweise stetig differenzierbaren parametrisierten Kurve γ berandet wird, d.h. $\partial\Omega$ ist die Spur von γ.

Aus technischen Gründen nehmen wir zusätzlich noch an, dass man sowohl

$$\overline{\Omega} = \{(x,y)\,|\,a \le x \le b\,,\ \phi(x) \le y \le \psi(x)\}$$

mit stetigen Funktionen $\phi, \psi : [a,b] \to \mathbb{R}$ als auch

$$\overline{\Omega} = \{(x,y)\,|\,\tilde{a} \le y \le \tilde{b}\,,\ \tilde{\phi}(y) \le x \le \tilde{\psi}(y)\}$$

mit stetigen Funktionen $\tilde{\phi}, \tilde{\psi} : [a,b] \to \mathbb{R}$ schreiben kann (man sagt, $\overline{\Omega}$ ist ein Normalbereich in Bezug auf beide Achsen).

Betrachtet man $\overline{\Omega}$ als Normalbereich in Bezug auf die x-Achse, dann entsteht die Kurve $\gamma : I \to \mathbb{R}^2$, $\gamma(t) = (x(t), y(t))$, durch Aneinanderhängen

- der Kurve $[a,b] \ni t \mapsto (t, \phi(t))$,
- der Verbindungsstrecke von $(b, \phi(b))$ nach $(b, \psi(b))$,
- der Kurve $[0, b-a] \ni t \mapsto (b-t, \psi(b-t))$ und
- der Verbindungsstrecke von $(a, \psi(a))$ nach $(a, \phi(a))$.

Man beachte, dass beim ersten bzw. dritten Punkt $x'(t) = \pm 1$ ist, während beim zweiten und vierten Punkt $x'(t) = 0$ gilt. Nach dem Satz von Fubini und Korollar 3.62 ergibt sich für eine stetige Funktion u auf $\overline{\Omega}$ mit beschränkter partieller Ableitung $\frac{\partial u}{\partial y}$ auf Ω die Gleichung

$$-\int_\Omega \frac{\partial u}{\partial y}(x,y)\,d\mu_2(x,y) = -\int_{[a,b]} \left(\int_{[\phi(x),\psi(x)]} \frac{\partial u}{\partial y}(x,y)\,d\mu_1(y)\right) d\mu_1(x) =$$
$$-\int_{[a,b]} u(x,\psi(x)) - u(x,\phi(x))\,d\mu_1(x) = \int_I u(\gamma(t))x'(t)\,dt\,.$$

Betrachtet man dagegen $\overline{\Omega}$ als Normalbereich in Bezug auf die y-Achse und eine stetige Funktion v auf $\overline{\Omega}$ mit beschränkter partieller Ableitung $\frac{\partial v}{\partial x}$ auf Ω, dann erhält man ganz analog

$$\int_\Omega \frac{\partial v}{\partial x}(x,y)\,d\mu_2(x,y) = \int_I v(\gamma(t))y'(t)\,dt\,.$$

Definiert man also das Integral des Vektorfeldes $(u,v) : \overline{\Omega} \to \mathbb{R}^2$ entlang der Kurve $\gamma = (x,y)$ durch

$$\int_\gamma (u(x,y)\,dx + v(x,y)\,dy) := \int_I u(\gamma(t))x'(t) + v(\gamma(t))y'(t)\,dt\,.$$

(wobei man $u(x,y)\,dx + v(x,y)\,dy$ nur als symbolische Abkürzung für das Vektorfeld (u,v) auffasse[3]), dann ergibt sich der Satz von Gauß in der Ebene.

Satz 3.82 *Sei $\Omega \subset \mathbb{R}^2$ ein beschränktes Gebiet mit einem durch eine stückweise stetig differenzierbare Kurve γ parametrisierbarem Rand $\partial\Omega$ und $(u,v) : \overline{\Omega} \to \mathbb{R}^2$ ein stetiges Vektorfeld, für das die partiellen Ableitungen $\frac{\partial u}{\partial x}$ und $\frac{\partial v}{\partial y}$ innerhalb von Ω existieren und beschränkt sind. Dann gilt*

$$\int_\Omega (\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y})\,d\mu_2(x,y) = \int_\gamma (u(x,y)\,dx + v(x,y)\,dy)$$

Benennt man die Funktionen u, v um, so ergibt sich

$$\int_\Omega (\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y})\,d\mu_2(x,y) = \int_\gamma (-v(x,y)\,dx + u(x,y)\,dy)$$

Modelliert das Vektorfeld (u,v) das Geschwindigkeitsfeld einer Strömung, so kann man die rechte Seite dieser Gleichung physikalisch als den Gesamtfluß der Strömung durch den Rand von Ω interpretieren, während die linke Seite die Ergiebigkeit der in Ω enthaltenen Quellen und Senken angibt. Der Integrand $\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$ wird dabei die Divergenz des Vektorfeldes (u,v) genannt.

Allgemeiner gilt für ein Vektorfeld $F = (u_1, \ldots, u_n)$ auf dem $\mathbb{R}^n$ bei $n \geq 3$ mit der Divergenz $\mathrm{div}(F) = \sum\limits_{i=1}^{n} \frac{\partial u_i}{\partial x_i}$ der Satz von Gauß bei gleichbleibender physikalischer Interpretation in der Form

$$\int_\Omega \mathrm{div}(F)(x)\,d\mu_n(x) = \int_{\partial\Omega} F d\vec{S}\,,$$

wobei auf der rechten Seite das Vektorfeld F entlang der $(n-1)$-dimensionalen Untermannigfaltigkeit $\partial\Omega$ integriert wird und man diese Integration natürlich erst noch mathematisch definieren müßte. Dies würde hier jedoch zu weit führen.

3.6 Der Transformationssatz

In diesem Abschnitt konzentrieren wir uns auf das Lebesgue-Maß μ auf der σ-Algebra der Lebesgue-meßbaren Teilmengen des $\mathbb{R}^n$ und leiten in Verallgemeinerung der Substitutionsregel für das eindimensionale Riemann-Integral den folgenden Transformationssatz für das n-dimensionale Lebesgue-Integral her.

[3] Später wird man den Ausdruck $u(x,y)\,dx + v(x,y)\,dy$ die vom Vektorfeld (u,v) induzierte 1-Form nennen und ihm mathematisch Sinn verleihen.

Satz 3.83 *Seien U und V offene Teilmengen des $\mathbb{R}^n$, und sei $\Phi : U \to V$ ein C^1-Diffeomorphismus. Dann ist die Funktion f über V genau dann Lebesgue-integrierbar, wenn $(f \circ \Phi)\,|\det(d\Phi)|$ über U Lebesgue-integrierbar ist, und es gilt in diesem Fall*

$$\int_U f(\Phi(x))\,|\det(d\Phi(x))|\,d\mu(x) = \int_V f(y)\,d\mu(y)$$

Für affin-linear Transformationen $\Phi(x) = Ax + b$ (wobei $U = \mathbb{R}^n = V$) und charakteristische Funktionen $f = 1_B$ zu Teilmengen $B \subset \mathbb{R}^n$ von endlichem Lebesgue-Maß haben wir die Gültigkeit der Transformationsformel schon in Satz 3.41 gezeigt, denn die linke Seite ist dann $\mu(\Phi^{-1}(B))|\det(A)|$ und die rechte Seite ist $\mu(B)$, also wird die Transformationsformel mit zu $\mu(\Phi^{-1}(B)) = |\det(A^{-1})|\mu(B)$.

Der allgemeine Transformationssatz 3.83 erweitert dieses Resultat in zweifacher Hinsicht, zum einen auf allgemeine Lebesgue-integrierbare Funktionen f statt nur charakteristischer Funktionen, und zum anderen auf durch nichtlineare Diffeomorphismen gegebene Transformationen statt nur affin-lineare Transformationen.

Den Beweis des Transformationssatzes 3.83 erbringen wir in mehreren Schritten. Zunächst machen wir uns klar, dass Nullmengne durch Diffeomorphismen auf Nullmengen abgebildet werden.

Lemma 3.84 *Ist $N \subset \mathbb{R}^n$ eine Lebesgue-Nullmenge und $\Phi : N \to \mathbb{R}^n$ Lipschitz-stetig, dann ist auch $\Phi(N)$ eine Lebesgue-Nullmenge.*

Beweis: Hat Φ die Lipschitzkonstante L bzgl. der Maximumsnorm auf dem $\mathbb{R}^n$, dann liegt für jeden halboffenen Quader $Q \subset \mathbb{R}^n$ das Bild $\Phi(N \cap Q)$ in einem Quader mit Volumen $(2L)^n \operatorname{Vol}(Q)$.

Da N eine Lebesgue-Nullmenge ist, gibt es zu $\epsilon > 0$ eine Folge Q_k von halboffenen Quadern mit $N \subset \bigcup_{k=1}^{\infty} Q_k$ und $\sum_{k=1}^{\infty} \operatorname{Vol}(Q_k) \leq \epsilon$. Das Bild $\Phi(N) = \bigcup_{k=1}^{\infty} \Phi(N \cap Q_k)$ wird daher durch abzählbar viele Quader mit Volumen $(2L)^n \operatorname{Vol}(Q_k)$ überdeckt, deren Gesamtvolumen kleiner oder gleich $\sum_{k=1}^{\infty} (2L)^n \operatorname{Vol}(Q_k) \leq (2L)^n \epsilon$ ist. Da dies für jedes $\epsilon > 0$ gilt, ist auch $\Phi(N)$ eine Lebesgue-Nullmenge. □

Ist nun $\Phi : U \to V$ eine stetig differenzierbare Abbildung von der offenen Teilmengen $U \subset R^n$ nach $V \subset R^n$, so ist Φ nach Korollar 2.37 lokal-Lipschitzstetig, und daher ist für jede Lebesgue-Nullmenge $N \subset U$ auch $\Phi(N) \subset V$ eine Lebesgue-Nullmenge. Denn die Einschränkung von Φ auf jede kompakte Teilmenge ist Lipschitz-stetig, es gibt eine abzählbare Ausschöpfung von U durch kompakte Mengen, und abzählbare Vereinigungen von Lebesgue-Nullmengen sind wiederum Lebesgue-Nullmengen.

Bemerkung 3.85 *Insbesondere ist der Transformationssatz auch für stetig differenzierbare Abbildungen $\Phi : U \to V$ wahr, für die es Lebesgue-Nullmengen N, N' gibt, so dass Φ ein Diffeomorphismus von $U \setminus N$ auf $V \setminus N'$ ist.*

Im nächsten Schritt beweisen wir den Transformationssatz 3.83 für den Fall, dass $f = 1_Q$ die charakteristische Funktion eines kompakten Quaders Q ist.

Lemma 3.86 *Seien U und V offene Teilmengen des $\mathbb{R}^n$, sei $\Phi : U \to V$ ein C^1-Diffeomorphismus und sei $Q \subset V$ ein kompakter Quader. Dann gilt*

$$\mu(\Phi(Q)) = \int_Q |\det(d\Phi(x))| \, d\mu(x) \, .$$

Beweis: Ohne Eischränkung kann man annehmen, dass Q ein Würfel ist, denn auf diesen Fall kann man den allgemeinen Fall mittels linearer Streckungen und des linearen Transformationssatzes 3.41 zurückführen.

Unterteile nun $Q = \bigcup_k Q_k$ in endlich viele kleine Würfel Q_k, die sich höchstens in ihren Randflächen und somit höchstens in Nullmengen schneiden. Dann gilt $\mu(\Phi(Q)) = \bigcup_k \mu(\Phi(Q_k))$, da sich auch die Mengen $\Phi(Q_k)$ nach Lemma 3.84 höchstens in Nullmengen schneiden.

Zu $\epsilon > 0$ kann man sicherlich so viele kleine Würfel Q_k wählen, dass mit dem Mittelpunkt a_k von Q_k für $x \in Q_k$ die Gleichung

$$\Phi(x) = \Phi(a_k) + A_k(x - a_k) + R(x - a_k)$$

mit $A_k := d\Phi(a_k)$ und einem Rest R gilt, der $\|R(h)\|_\infty \leq \|h\|_\infty \epsilon$ erfüllt – denn $d\Phi$ ist schließlich als stetige Funktion auf dem Kompaktem Würfel Q auch gleichmäßig stetig.

Um das Lebesgue-Maß von $\Phi(Q_k)$ zu bestimmen, müssen wir also nach Translation von a_k und $\Phi(a_k)$ in den Ursprung nur das Volumen von $\Phi(\tilde{Q}_k)$ für einen Würfel $\tilde{Q}_k$ bestimmen, dessen Mittelpunkt im Ursprung liegt und auf dem Φ die Form $\Phi(x) = A_k x + R(x)$ hat. Insbesondere liegt dann auf $\tilde{Q}_k$ die Abbildung $A_k^{-1} \circ \Phi$ nahe an der Identität wegen $(A_k^{-1} \circ \Phi)(x) = x + (A_k^{-1} \circ R)(x)$, wobei es eine (von k unabhängige) Konstante C mit $\|(A_k^{-1} \circ R)(x)\|_\infty \leq C\epsilon$ gibt. Als nahe an der Identität liegende Abbildung enthält $(A_k^{-1} \circ \Phi)(\tilde{Q}_k)$ aber einen Würfel vom Radius $(1 - C\epsilon)$ mal den Radius von $\tilde{Q}_k$, und ist andererseits in einem Würfel vom Radius $(1 + C\epsilon)$ mal den Radius von $\tilde{Q}_k$ enthalten. Wendet man A_k auf diese Würfel an, berechnet ihr Volumen und macht die Translationen rückgängig, so erhält man

$$|\det(A_k)| \operatorname{Vol}(Q_k) - C'\epsilon \operatorname{Vol}(Q_k) \leq \mu(\Phi(Q_k)) \leq |\det(A_k)| \operatorname{Vol}(Q_k) + C'\epsilon \operatorname{Vol}(Q_k)$$

mit einer neuen Konstanten C'. Summiert man nun über k, so erhält man

$$|\mu(\Phi(Q)) - \sum_k |\det(A_k)| \operatorname{Vol}(Q_k)| \leq C'\epsilon \operatorname{Vol}(Q)$$

Zu guter Letzt nähern die Riemannsche Summen $\sum_k |\det(A_k)| \operatorname{Vol}(Q_k)$ nach Satz 3.50 das Lebesgue-Integral $\int_Q |\det(d\Phi(x))| \, d\mu(x)$ der stetigen Funktion $|\det(d\Phi(x))|$ über den kompakten Würfel beliebig genau an, wenn man die Zerlegung Q_k von Q nur fein genug wählt. Also kann man die Differenz von $\mu(\Phi(Q_k))$ und $\int_Q |\det(d\Phi(x))| \, d\mu(x)$ durch ein festes Vielfaches von ϵ abschätzen, und da $\epsilon > 0$ beliebig war, erhält man

$$\mu(\Phi(Q)) = \int_Q |\det(d\Phi(x))| \, d\mu(x) \, .$$

$\square$

Das vorige Lemma 3.86 erlaubt uns nun den Beweis des allgemeinen Transformationssatzes 3.83.

Beweis: Nach Lemma 3.86 gilt der Transformationssatz für charakteristische Funktionen von Quadern und daher aufgrund der Linearität des Lebesgue-Integrals auch für Treppenfunktionen zu kompakten Quadern in V.

Nun gibt es aber zu jeder Lebesgue-integrierbaren Funktion f auf V eine Cauchy-Folge von Treppenfunktionen f_k zu kompakten Quadern in V bzgl. $\|\cdot\|_1$, die außerhalb einer Nullmenge $N \subset V$ punktweise gegen f konvergiert. Da der Transformationssatz für Treppenfunktionen zu kompakten Quadern in V gilt, ist wegen

$$\begin{gathered}\|(f_k \circ \Phi)|\det(d\Phi)| - (f_l \circ \Phi)|\det(d\Phi)|\|_1 = \\ \int_U ((|f_k - f_l|) \circ \Phi)(x)|\det(d\Phi(x))|\, d\mu(x) = \\ \int_V (|f_k - f_l|)(y)\, d\mu(y) = \|f_k - f_l\|_1\end{gathered}$$

die Folge $(f_k \circ \Phi)|\det(d\Phi)|$ eine L^1-Cauchy-Folge integrierbarer Funktionen auf U, und diese konvergiert außerhalb der Nullmenge $\Phi^{-1}(N)$ punktweise gegen $(f \circ \Phi)|\det(d\Phi)|$. Also ist die Funktion $(f \circ \Phi)|\det(d\Phi)|$ über U Lebesgue-integrierbar mit Lebesgue-Integral

$$\begin{gathered}\int_U (f \circ \Phi)(x)|\det(d\Phi(x))|\, d\mu(x) = \lim_{k\to\infty} \int_U (f_k \circ \Phi)(x)|\det(d\Phi(x))|\, d\mu(x) = \\ \lim_{k\to\infty} \int_V f_k(y)\, d\mu(y) = \int_V f(y)\, d\mu(y)\,,\end{gathered}$$

wobei wiederum die Gültigkeit des Transformationssatz für Treppenfunktionen zu kompakten Quadern in V ausgenutzt wurde. $\square$

Im folgenden Abschnitt wird gezeigt, wie man den Transformationssatz in dem Spezialfall anwenden kann, wo $\Phi = P_n$ die n-dimensionale Polarkoordinatenabbildung ist.

Integration mittels Polarkoordinaten Aus Beispiel 2.59 ist uns schon bekannt, dass die rekursiv durch

$$P_{n+1}(r, \phi_1, \ldots, \phi_n) := \begin{pmatrix} P_n(r, \phi_1, \ldots, \phi_{n-1}) \cos(\phi_n) \\ r \sin(\phi_n) \end{pmatrix}$$

und $P_2(r, \phi) := \begin{pmatrix} r\cos(\phi) \\ r\sin(\phi) \end{pmatrix}$ definierte n-dimensionale Polarkoordinatenabbildung P_n ein C^∞-Diffeomorphismus von $\mathbb{R}^+ \times (-\pi, \pi) \times (-\pi/2, \pi/2)^{n-2}$ auf $\mathbb{R}^n \setminus \{y \in \mathbb{R}^n \,|\, y_1 \leq 0, y_2 = 0\}$ ist, für die $\det(dP_n) = r^{n-1}\cos(\phi_2)\ldots\cos^{n-2}(\phi_{n-1})$ gilt. Wendet man mittels Bemerkung 3.85 den Transformationssatz auf $\Phi := P_n$ an und benutzt danach noch den Satz von Fubini, so ergibt sich das folgende Korollar.

Korollar 3.87 *Eine Funktion f auf der Kugelschale*

$$K(R_1, R_2) := \{y \in \mathbb{R}^n \mid R_1 \leq \|y\|_2 \leq R_2\}$$

ist genau dann Lebesgue-integrierbar über $K(R_1, R_2)$, wenn die Funktion

$$(r, \phi_1, \dots, \phi_{n-1}) \mapsto f(P_n(r, \phi_1, \dots, \phi_{n-1}))r^{n-1} \cos(\phi_2) \dots \cos^{n-2}(\phi_{n-1})$$

Lebesgue-integrierbar über $[R_1, R_2] \times (-\pi, \pi) \times (-\pi/2, \pi/2)^{n-2}$ ist, und dann gilt

$$\int_{K(R_1,R_2)} f(y)\, d\mu(y) = \int_{[R_1,R_2]} \int_{(-\pi,\pi)} \int_{(-\pi/2,\pi/2)^{n-2}} f(P_n(r, \phi_1, \dots, \phi_{n-1}))\cdot$$
$$r^{n-1} \cos(\phi_2) \dots \cos^{n-2}(\phi_{n-1})\, d\mu_{n-2}(\phi_2, \dots, \phi_{n-1})\, d\mu_1(\phi_1)\, d\mu_1(r)\,.$$

Wir wollen noch den Spezialfall von rotationssymmetrischen Funktionen $f(\|y\|_2)$ festhalten.

Korollar 3.88 *Ist f eine Funktion auf dem Intervall $[R_1, R_2] \subset [0, \infty)$, so ist $y \mapsto f(\|y\|_2)$ genau dann über die Kugelschale $K(R_1, R_2)$ Lebesgue-integrierbar, wenn die Funktion $f(r)r^{n-1}$ über $[R_1, R_2]$ Lebesgue-integrierbar ist, und es gilt*

$$\int_{K(R_1,R_2)} f(y)\, d\mu(y) = n\mu_n(B_1(0)) \int_{[R_1,R_2]} f(r)r^{n-1}\, d\mu_1(r)\,.$$

mit dem Lebesgue-Maß $\mu_n(B_1(0))$ der n-dimensionalen Euklidischen Einheitskugel $B_1(0) = \{x \in \mathbb{R}^n \mid \|x\|_2 \leq 1\}$.

Zum Beweis dieses Korollars ist nur noch

$$n\mu_n(B_1(0)) = \int_{(-\pi,\pi)} \int_{(-\pi/2,\pi/2)^{n-2}} \cos(\phi_2) \dots \cos^{n-2}(\phi_{n-1})\, d\mu_{n-2}(\phi_2, \dots, \phi_{n-1})\, d\mu_1(\phi_1)$$

zu zeigen, dies ergibt sich aber sofort, indem man Korollar 3.87 auf $f = 1_{B_1(0)}$ bei $[R_1, R_2] = [0, 1]$ anwendet und $\int_0^1 r^{n-1} = \frac{1}{n}$ ausnutzt.

Beispiel 3.89 *Die Formel $\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$ kann man mittels Polarkoordinaten leicht durch die Rechnung*

$$\left(\int_{\mathbb{R}} e^{-x^2}\, d\mu_1(x, y)\right)^2 = \int_{\mathbb{R}^2} e^{-x^2-y^2}\, d\mu_2(x, y) =$$
$$\int_{(0,\infty)\times(-\pi,\pi)} e^{-r^2} r\, d\mu_2(r, \phi) = 2\pi \int_0^{\infty} re^{-r^2}\, dr = \pi \int_0^{\infty} e^{-t}\, dt = \pi$$

einsehen, bei der in der ersten Gleichung der Satz von Fubini und in der zweiten Gleichung der Transformationssatz für die zweidimensionale Polarkoordinaten-Abbildung verwendet wurde.

Beispiel 3.90 *Das Volumen der Kugel im $\mathbb{R}^3$ mit Radius R ist*

$$\int_{K(0,R)} 1\, d\mu(y) = \int_{[0,R]} \int_{(-\pi,\pi)} \int_{(-\pi/2,\pi/2)} r^2 \cos(\phi_2)\, d\mu_1(\phi_2)\, d\mu_1(\phi_1)\, d\mu_1(r) =$$
$$\left(\int_0^R r^2\, dr\right) \cdot \left(\int_{-\pi}^{\pi} 1\, d\phi_1\right) \cdot \left(\int_{-\pi/2}^{\pi/2} \cos(\phi_2)\, d\phi_2\right) = \frac{1}{3} R^3 \cdot 2\pi \cdot 2 = \frac{4}{3}\pi R^3 \,.$$

3.7 Fourier-Theorie

In diesem abschließenden Kapitel diskutieren wir als eine Anwendung des Lebesgue-Integrals die diskrete und die kontinuierliche Fourier-Transformation im Mehrdimensionalen. Beide haben eine große Bedeutung sowohl innerhalb der Mathematik als auch in den die Mathematik anwendenden Wissenschaften.

Beispielsweise spielt die zweidimensionale diskrete Fourier-Transformation eine herausragende Rolle in der Bildverarbeitung und bei der Komprimierung von Bildern. So basiert z.B. das JPEG-Format auf der diskreten Kosinus-Transformation, einer reellen Variante der diskreten Fourier-Transformation.

Die kontinuierliche Fourier-Transformation wird insbesondere beim Studium von (partiellen) Differentialgleichungen auf dem $\mathbb{R}^n$ eingesetzt, da die die Differentiation unter der Fourier-Transformation in eine Multiplikation übergeht.

Der Hilbert-Raum L^2 Aus der Analysis I ist uns die diskrete Fourier-Transformation für 2π-periodische Funktionen $f : \mathbb{R} \to \mathbb{C}$ einer reellen Variablen schon bekannt. Dort wurde gezeigt, dass für über $[-\pi, \pi]$ (quadrat-)Riemann-integrierbare 2π-periodische Funktionen f die Fourier-Reihe $\sum\limits_{k=-\infty}^{\infty} c_k e^{ikx}$ bzgl. der Halbnorm $\|f\|_2 := \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2\, dx\right)^2$ gegen f konvergiert.

Die Umkehrung, dass bei Konvergenz von $\sum\limits_{k=-\infty}^{\infty} |c_k|^2$ auch die Reihe $\sum\limits_{k=-\infty}^{\infty} c_k e^{ikx}$ gegen eine quadrat-integrierbare Funktion konvergiert, konnte dort aber noch nicht gezeigt werden. Dies gelingt erst mit Hilfe des Lebesgue-Integrals aufgrund der Vollständigkeit der zugehörigen Funktionenräume.

Tatsächlich ist nicht nur $L^1(\mathbb{R}^n, \mathbb{C})$ vollständig.

Definition 3.91 *Die Menge der Äquivalenzklassen von 2π-periodischen Funktionen $f : \mathbb{R}^n \to \mathbb{C}$, die in $(-\pi, \pi)^n$ lokal-integrierbar sind und für die $|f|^2$ über $(-\pi, \pi)^n$ Lebesgue-integrierbar ist, wobei man zwei Funktionen bei Übereinstimmung außerhalb einer Nullmenge identifiziert, bezeichnet man mit $L^2_{per}(\mathbb{R}^n, \mathbb{C})$.*

Für jedes Element f von $L^2_{per}(\mathbb{R}^n, \mathbb{C})$ kann man aufgrund der Integrierbarkeit von $|f|^2$ den Wert

$$\|f\|_{L^2_{per}} := \left(\frac{1}{(2\pi)^n} \int_{(-\pi,\pi)^n} |f(x)|^2\, d\mu_n(x)\right)^{1/2}$$

definieren. Tatsächlich ist dies eine Norm, die vom (komplexen) Skalarprodukt

$$\langle f, g\rangle_{L^2_{per}} := \frac{1}{(2\pi)^n} \int_{(-\pi,\pi)^n} \overline{f(x)} g(x)\, d\mu_n(x)$$

induziert wird.

Lemma 3.92 $L^2_{per}(\mathbb{R}^n, \mathbb{C})$ *ist ein (komplexer) Vektorraum, und* $\|\cdot\|_{L^2_{per}}$ *ist die vom (komplexen) Skalarprodukt* $\langle \cdot, \cdot\rangle_{L^2_{per}}$ *auf* $L^2_{per}(\mathbb{R}^n, \mathbb{C})$ *induzierte Norm.*

Beweis: Offensichtlich folgt aus $\|f\|_{L^2_{per}} = 0$ auch $|f|^2 = 0$ fast überall und somit auch $f = 0$ fast überall, und auch $\|\lambda f\|_{L^2_{per}} = |\lambda|\|f\|_{L^2_{per}}$ gilt für $\lambda \in \mathbb{C}$.

Mit $f, g \in L^2_{per}(\mathbb{R}^n, \mathbb{C})$ ist auch $f + g$ quadrat-integrierbar, denn da f und g in $(-\pi, \pi)^n$ lokal-integrierbar sind, ist auch $f + g$ in $(-\pi, \pi)^n$ lokal-integrierbar, und aus der Ungleichung $|f + g|^2 \leq 2(|f|^2 + |g|^2)$ folgt nach Korollar 3.60 auch die Integrierbarkeit von $|f + g|^2$. Also ist insbesondere $L^2_{per}(\mathbb{R}^n, \mathbb{C})$ ein Vektorraum.

Außerdem sind mit $f \in L^2_{per}(\mathbb{R}^n, \mathbb{C})$ auch die reell-wertigen Funktionen $\Re(f)$ und $\Im(f)$ quadrat-integrierbar, da sie lokal-integrierbar sind und ihre L^2-Normen durch $\|f\|_{L^2_{per}}$ beschränkt sind. Umgekehrt folgt aus der Quadrat-Integrierbarkeit von $\Re(f)$ und $\Im(f)$ auch die von f.

Will man nun die Integrierbarkeit von $\bar{f}g$ oder äquivalenterweis fg für $f, g \in L^2_{per}(\mathbb{R}^n, \mathbb{C})$ zeigen, so reicht es daher, reellwertige Funktionen zu betrachten. Für diese gilt aber

$$fg = \frac{1}{2}\left((f + g)^2 - f^2 - g^2\right) ,$$

d.h. für quadrat-integrierbare f, g ist fg zumindest integrierbar. Insbesondere ist $\langle \cdot, \cdot\rangle_{L^2_{per}}$ wohldefiniert, und da $\langle \cdot, \cdot\rangle_{L^2_{per}}$ offensichtlich alle Eigenschaften hat, die man von einem komplexen Skalarprodukt verlangt – Definitheit, $\mathbb{C}$-Linearität in der zweiten Komponente und $\langle f, g\rangle_{L^2_{per}} = \overline{\langle g, f\rangle_{L^2_{per}}}$ – ist $\|\cdot\|_{L^2_{per}}$ wegen $\|f\|_{L^2_{per}} = \sqrt{\langle f, f\rangle_{L^2_{per}}}$ nach dem komplexen Analogon von Korollar 1.6 die zu $\langle \cdot, \cdot\rangle_{L^2_{per}}$ gehörige Norm. □

Festhalten wollen wir, dass analog zu Satz 1.5 auch für komplexe Skalarprodukte die Cauchy-Schwarzsche Ungleichung gilt, hier also

$$\langle f, g\rangle_{L^2_{per}} \leq \|f\|_{L^2_{per}}\|g\|_{L^2_{per}}$$

und sogar $\|fg\|_{L^1((-\pi,\pi)^n)} \leq \|f\|_{L^2_{per}}\|g\|_{L^2_{per}}$ gilt.

Nun beweisen wir, dass $L^2_{per}(\mathbb{R}^n, \mathbb{C})$ auch vollständig ist.

Satz 3.93 *Der mit der vom Skalarprodukt* $\langle \cdot, \cdot\rangle_{L^2_{per}}$ *induzierten Norm* $\|\cdot\|_{L^2_{per}}$ *versehene Vektorraum* $L^2_{per}(\mathbb{R}^n, \mathbb{C})$ *ist vollständig, d.h. ein Hilbert-Raum.*

Beweis: Neben der schon im vorigen Beweis erwähnten Beobachtung, dass f genau dann quadrat-integrierbar ist, wenn die reell-wertigen Funktionen $\Re(f)$ und $\Im(f)$ quadrat-integrierbar sind, bemerke man, dass ein rell-wertiges f genau dann quadrat-integrierbar ist, wenn die nicht-negativen Funktionen $f^+ := \max(f,0)$ und $f^- := -\min(f,0)$ quadrat-integrierbar sind. Daher reicht es zu zeigen, dass jede Cauchy-Folge nicht-negativer Funktionen $f_k \in L^2_{per}(\mathbb{R}^n,\mathbb{C})$ konvergiert.

Ist nun f_k solch eine Cauchy-Folge bzgl. $\|\cdot\|_{L^2_{per}}$, dann ist f_k^2 eine L^1-Cauchy-Folge auf $(-\pi,\pi)^n$, denn aus $|f_k^2 - f_l^2| \leq |f_k - f_l|^2 + 2|f_l|\,|f_k - f_l|$ folgt

$$\|f_k^2 - f_l^2\|_{L^1((-\pi,\pi)^n)} \leq \|f_k - f_l\|^2_{L^2_{per}(R^n)} + 2\|f_l\|_{L^2_{per}(R^n)}\,\|f_k - f_l\|_{L^2_{per}(R^n)}\,,$$

so dass wegen der Beschränktheit von $\|f_l\|_{L^2_{per}(R^n)}$ bei genügend großen k,l auch $\|f_k^2 - f_l^2\|_{L^1((-\pi,\pi)^n)}$ beliebig klein wird.

Aus der Vollständigkeit von $L^1((-\pi,\pi)^n)$ folgt nun, dass f_k^2 gegen eine Funktion $F \in L^1((-\pi,\pi)^n)$ konvergiert, Zu zeigen bleibt noch, dass f_k bzgl. $\|\cdot\|_{L^2_{per}}$ gegen die 2π-periodische Fortsetzung f von $\sqrt{F}$ konvergiert. Dies folgt jedoch wegen $f_k, f \geq 0$ aus

$$\|f_k - f\|^2_{L^2_{per}(R^n)} = \|(f_k - f)^2\|_{L^1((-\pi,\pi)^n)} \leq \|(f_k - f)(f_k + f)\|_{L^1((-\pi,\pi)^n)} = \|f_k^2 - f^2\|_{L^1((-\pi,\pi)^n)}$$

Zu guter Letzt ist auch f quadrat-integrierbar, denn $\|f\|_{L^2_{per}(R^n)} = \sqrt{\|F\|_{L^1((-\pi,\pi)^n)}} < \infty$. Also konvergiert jede Cauchy-Folge, d.h. $L^2_{per}(\mathbb{R}^n,\mathbb{C})$ ist vollständig. □

Bemerkung 3.94 *Tatsächlich gibt es zu jeder Cauchy-Folge $f_k \in L^2_{per}(\mathbb{R}^n,\mathbb{C})$ sogar eine Teilfolge, die punktweise fast überall gegen den L^2-Grenzwert f von f_k konvergiert. Denn dies wurde in Satz 3.52 für L^1-Konvergenz gezeigt, der obige Beweis führt aber die L^2-Konvergenz auf die L^1-Konvergenz zurück.*

Ähnlich kann man auch beweisen, dass für einen (komplexen) Hilbert-Raum Y und eine meßbare Menge $\Omega \subset \mathbb{R}^n$ der Raum $L^2(\Omega, Y)$ der Äquivalenzklassen von Funktionen $f : \Omega \to Y$, die in Ω lokal-integrierbar sind und für die $\|f\|_y^2$ über Ω integrierbar ist, ein Hilbert-Raum bzgl. des Skalarproduktes

$$\langle f, g\rangle_{L^2} := \int_\Omega \langle f(x), g(x)\rangle_Y \, d\mu_n(x)$$

ist. Man sagt dabei allgemein, dass eine Folge $f_k \in L^2(\Omega, Y)$, die bzgl. der zugehörigen Norm $\|\cdot\|_{L^2}$ gegen f konvergiert, im quadratischen Mittel gegen f konvergiert.

Diskrete Fourier-Transformation Die diskrete Fourier-Transformation weist jeder 2π-periodischen Funktion f, deren Einschränkung auf einen der Gitterbereiche von $\mathbb{R}^n/(2\pi\mathbb{Z}^n)$ Lebesgue-integrierbar ist (wir wählen hier den Gitterbereich $(-\pi,\pi)^n$), die durch

$$\hat{f}(k) := \frac{1}{(2\pi)^n} \int_{(-\pi,\pi)^n} f(x) e^{-i\langle k,x\rangle} \, d\mu_n(x)$$

definierte Funktion $\hat{f} : \mathbb{Z}^n \to \mathbb{C}$ zu. Dabei ist $k \in \mathbb{Z}^n$ ein n-Tupel ganzer Zahlen und $\langle k, x\rangle = \sum\limits_{i=1}^{n} k_i x_i$ das übliche Euklidische Skalarprodukt auf dem $\mathbb{R}^n$. Außerdem bemerke man, dass der Integrand $f(x)e^{-i\langle k,x\rangle}$ die Lebesgue-integrierbare Majorante $|f|$ hat und somit wirklich Lebesgue-integrierbar ist.

Obwohl die Fourier-Transformierte $\hat{f}$ also sogar schon für $f \in L^1_{per}(\mathbb{R}^n, \mathbb{C})$ wohldefiniert ist, betrachtet man üblicherweise nur die Fourier-Transformierten von quadrat-integrierbaren $f \in L^2_{per}(\mathbb{R}^n, \mathbb{C})$. Tatsächlich ist jede über $(-\pi, \pi)^n$ quadrat-integrierbare Funktion auch integrierbar, denn aus $\|fg\|_{L^1((-\pi,\pi)^n)} \leq \|f\|_{L^2_{per}} \|g\|_{L^2_{per}}$ folgt mit $g := 1$ schon $\|f\|_{L^1((-\pi,\pi)^n)} \leq \sqrt{\mu_n((-\pi,\pi)^n)} \|f\|_{L^2_{per}}$.

Die so durch $f \mapsto \hat{f}$ definierte diskrete Fourier-Transformation weist also quadrat-integrierbaren 2π-periodischen Funktionen f eine Funktion $\hat{f} : \mathbb{Z}^n \to \mathbb{C}$ zu, die man auch als eine (beidseitige n-)Folge bezeichnen könnte.

Mit dieser Folge kann man die zu f gehörige Fourier-Reihe $\sum\limits_{k \in \mathbb{Z}^n} \hat{f}(k) e^{i\langle k,x\rangle}$ betrachten. Dabei ist die in beliebiger Reihenfolge stattfindende Summation über die abzählbare Menge $\mathbb{Z}^n$ nur dann sinnvoll, wenn die Reihe absolut konvergiert. Dies werden wir mit Hilfe des folgenden, schon aus der Analysis I bekannten Lemmas beweisen.

Lemma 3.95 *Für $f \in L^2_{per}(\mathbb{R}^n, \mathbb{C})$ und jede endliche Indexmenge $I \subset \mathbb{Z}^n$ gilt*

$$\|f - \sum_{k \in I} \hat{f}(k) e^{i\langle k,x\rangle}\|_{L^2_{per}} = \|f\|_{L^2_{per}} - \sum_{k \in I} |\hat{f}(k)|^2 \quad .$$

Beweis: Es gilt

$$\langle f, \hat{f}(k) e^{i\langle k,x\rangle}\rangle_{L^2_{per}} = \hat{f}(k) \langle f, e^{i\langle k,x\rangle}\rangle_{L^2_{per}} = \hat{f}(k) \overline{\langle e^{i\langle k,x\rangle}, f\rangle_{L^2_{per}}} = \hat{f}(k)\overline{\hat{f}(k)} = |\hat{f}(k)|^2$$

und $\langle e^{i\langle k,x\rangle}, e^{i\langle l,x\rangle}\rangle_{L^2_{per}} = \delta_{kl}$, also

$$\|f - \sum_{k \in I} \hat{f}(k) e^{i\langle k,x\rangle}\|_{L^2_{per}} = \|f\|_{L^2_{per}} - 2\sum_{k \in I} \langle f, \hat{f}(k) e^{i\langle k,x\rangle}\rangle_{L^2_{per}} + \|\sum_{k \in I} \hat{f}(k) e^{i\langle k,x\rangle}\|^2_{L^2_{per}} =$$

$$\|f\|_{L^2_{per}} - \sum_{k \in I} |\hat{f}(k)|^2$$

□

Aus diesem Lemma folgt wegen $\|f - \sum\limits_{k \in I} \hat{f}(k) e^{i\langle k,x\rangle}\|_{L^2_{per}} \geq 0$ insbesondere

$$\sum_{k \in I} |\hat{f}(k)|^2 \leq \|f\|_{L^2_{per}} < \infty$$

und somit die absolute Konvergenz der Reihe $\sum\limits_{k \in \mathbb{Z}^n} |\hat{f}(k)|^2$ für quadrat-integrierbare f. Außerdem folgt auch, dass $\sum\limits_{k \in \mathbb{Z}^n} \hat{f}(k) e^{i\langle k,x\rangle}$ genau dann gegen f im quadratischen Mittel konvergiert, wenn $\sum\limits_{k \in \mathbb{Z}^n} |\hat{f}(k)|^2$ gegen $\|f\|_{L^2_{per}}$ konvergiert. Letzteres kann man aber wie in der Analysis I zunächst für Treppenfunktionen zu Quadern und dann für allgemeine quadrat-integrierbare Funktionen zeigen.

Satz 3.96 *Ist $f \in L^2_{per}(\mathbb{R}^n, \mathbb{C})$, so konvergiert die Fourier-Reihe $\sum\limits_{k \in \mathbb{Z}^n} \hat{f}(k) e^{i\langle k,x\rangle}$ im quadratischen Mittel gegen f.*

Beweis: Für die 2π-periodische Fortsetzung f der charakteristischen Funktion eines Quaders $Q := [a_1, b_1] \times \cdots \times [a_n, b_n] \subset (-\pi, \pi)^n$ ist $\hat{f}(k)$ das Produkt der Fourier-Transormierten $\hat{f}_j(k_j)$ der eindimensionalen Funktionen $f_j : x_j \mapsto 1_{[a_j,b_j]}(x_j)$.

Aus der Analysis I wissen wir schon $\sum\limits_{k_j=-\infty}^{\infty} |\hat{f}_j(k_j)|^2 = \frac{b_j - a_j}{2\pi}$. Daher gilt auch im Mehrdimensionalen für die charakteristische Funktion f des Quaders Q

$$\sum_{k \in \mathbb{Z}^n} |\hat{f}(k)|^2 = \prod_{j=1}^{n} \left(\sum_{k_j=-\infty}^{\infty} |\hat{f}_j(k_j)|^2 \right) =$$
$$\prod_{j=1}^{n} \frac{b_j - a_j}{2\pi} = \frac{1}{(2\pi)^n} \int_{(-\pi,\pi)^n} |1_Q(x)|^2 \, d\mu_n(x) = \|f\|^2_{L^2_{per}} .$$

Somit konvergiert die Fourier-Reihe zur periodischen Fortsetzung f der charakteristischen Funktion eines Quaders Q im quadratischen Mittel gegen f.

Aufgrund der Linearität der Fourier-Transformation gilt dies dann auch für 2π-periodische Funktionen f, deren Einschränkung auf $(-\pi, \pi)^n$ eine Treppenfunktion zu Quadern ist. Genauer gilt bei $f|_{(-\pi,\pi)^n} = \sum_j f_j 1_{Q_j}$ mit den Partialsummen S_I der Fourier-Reihe von f und $S_{j,I}$ der Fourier-Reihe der 2π-periodischen Fortsetzung von 1_{Q_j} wegen $S_I = \sum_j f_j S_{j,I}$ die Ungleichung

$$\|f - S_I\|_{L^2_{per}} = \| \sum_j f_j (1_{Q_j} - S_{j,I}) \|_{L^2((-\pi,\pi)^n)} \leq \sum_j |f_j| \|1_{Q_j} - S_{j,I}\|_{L^2((-\pi,\pi)^n)}$$

für jede Indexmenge $I \subset \mathbb{Z}^n$, also aufgrund der Konvergenz von $S_{j,I}$ gegen 1_{Q_j} im quadratischen Mittel auch die Konvergenz der Fourier-Reihe von f gegen f.

Der allgemeine Fall folgt nun daraus, dass man jede quadrat-integrierbare Funktion beliebig genau durch Treppenfunktionen zu Quadern bzgl. der L^2-Norm approximieren kann. Zum Beweis dieser Behauptung reicht es dabei wieder aus, reellwertige nicht-negative quadrat-integrierbare Funktionen f zu betrachten. Dann kann man f^2 als integrierbare Funktion auf $(-\pi, \pi)^n$ durch eine Folge $F_k \geq 0$ von Treppenfunktionen zu Quadern bzgl. der L^1-Norm annähern. Somit ist aber $f_k := \sqrt{F_k}$ eine Folge von Treppenfunktionen zu Quadern, die f bzgl. der L^2-Norm beliebig genau annähert, denn wegen $f, f_k \geq 0$ gilt

$$\|f_k - f\|^2_{L^2((-\pi,\pi)^n)} = \|(f_k - f)^2\|_{L^1((-\pi,\pi)^n)} \leq \|(f_k - f)(f_k + f)\|_{L^1((-\pi,\pi)^n)} = \|f_k^2 - f^2\|_{L^1((-\pi,\pi)^n)} .$$

Sei also $\epsilon > 0$ vorgegeben und ist g eine Treppenfunktion mit $\|f - g\|_{L^2} \leq \epsilon/2$, dann gilt mit der Partialsumme $S_{f-g,I}$ zu $f - g$ nach Lemma 3.95 insbesondere $\|(f - g) - S_{f-g,I}\|_{L^2} \leq \|f - g\|_{L^2} \leq \epsilon/2$, und andererseits gibt es aufgrund der Konvergenz der Partialsummen $S_{g,I}$ der Fourier-Reihe von g bzgl. der L^2-Norm gegen g eine Indexmenge, so dass für alle gröseren Indexmengen I die Ungleichung $\|g - S_{g,I}\|_{L^2} \leq \epsilon/2$ gilt. Aus der Linearität $S_{f,I} = S_{g,I} + S_{f-g,I}$ der Fourier-Transformation folgt dann aber auch

$$\|f - S_{f,I}\|_{L^2} \leq \|g - S_{g,I}\|_{L^2} + \|(f-g) - S_{f-g,I}\|_{L^2} \leq \epsilon/2 + \epsilon/2 = \epsilon$$

für genügend große Indexmengen I, d.h. die Konvergenz der Fourier-Reihe von f im quadratischen Mittel gegen f. $\square$

Die Umkehrung dieses Satzes für Koeffizienten c_k mit $\sum_{k\in\mathbb{Z}^n} |c_k|^2 < \infty$ konnten wir aber noch nicht zeigen. Dies ist aufgrund der Vollständigkeit von L^2 jetzt jedoch ganz einfach, denn zu ϵ gibt es wegen der Konvergenz von $\sum_{k\in\mathbb{Z}^n} |c_k|^2 < \infty$ eine Indexmenge I, so dass für jede Indexmenge $I' \supset I$ die Ungleichung

$$\|\sum_{k'\in I'} c_{k'} e^{i\langle k',x\rangle} - \sum_{k\in I} c_k e^{i\langle k,x\rangle}\|_{L^2} \leq \sum_{k\in I'\setminus I} |c_k|^2 \leq \sum_{k\in\mathbb{Z}^n\setminus I} |c_k|^2 \leq \epsilon$$

gilt, also ist die Folge $\sum\limits_{k\in I} c_k e^{i\langle k,x\rangle}$ der Partialsummen der Fourierreihe $\sum\limits_{k\in\mathbb{Z}^n} c_k e^{i\langle k,x\rangle}$ eine Cauchy-Folge bzgl. der L^2-Norm. Somit konvergiert sie aufgrin der Vollständigkeit von $L^2_{per}(\mathbb{R}^n, \mathbb{C})$ im quadratischen Mittel gegen eine Funktion aus $L^2_{per}(\mathbb{R}^n, \mathbb{C})$, die man sinnvollerweise wieder durch $\sum\limits_{k\in\mathbb{Z}^n} c_k e^{i\langle k,x\rangle}$ symbolisiert.

Daher stellt jede Fourier-Reihe $\sum\limits_{k\in\mathbb{Z}^n} c_k e^{i\langle k,x\rangle}$ mit Koeffizienten c_k, die $\sum_{k\in\mathbb{Z}^n} |c_k|^2 < \infty$ erfüllen, eine 2π-periodische quadrat-integrierbare Funktion dar, und die Fourier-Transformation ist eine Bijektion zwischen den entsprechenden Räumen. Dies halten wir im folgenden Satz fest, führen aber vorher noch die Bezeichnung $l^2(\mathbb{Z}^n)$ für den Vektorraum der Folgen c_k zu $k \in \mathbb{Z}^n$ ein, der mit dem Skalarprodukt

$$\langle c, d\rangle_{l^2} := \sum_{k\in\mathbb{Z}^n} \bar{c}_k d_k$$

ein Hilbert-Raum wird.

Satz 3.97 *Die diskrete Fourier-Transformation $f \mapsto \hat{f}$ ist eine Bijektion vom Hilbert-Raum $L^2_{per}(\mathbb{R}^n, \mathbb{C})$ in den Hilbert-Raum der quadrat-summierbaren (komplexen beidseitigen n-)Folgen $l^2(\mathbb{Z}^n, \mathbb{C})$, die $\|f\|^2_{L^2_{per}} = \sum_{k\in\mathbb{Z}^n} |\hat{f}(k)|^2 = \|\hat{f}\|^2_{l^2}$ erfüllt und daher eine Isometrie genannt wird.*

Über die diskrete Fourier-Transformation wissen wir durch diesen Satz also genau Bescheid. Im folgenden wollen wir in ähnlicher Weise die kontinuierliche Fourier-Transformation studieren, benötigen dazu aber die Faltung zweier Funktionen.

Faltung Für integrierbare reell- oder komplex-wertige Funktionen f, g auf dem $\mathbb{R}^n$ ist $(x, y) \mapsto f(x)g(y)$ eine integrierbare Funktion auf dem $\mathbb{R}^{2n}$, und unter der Transformation $(x, y) \mapsto (x - y, y)$ geht diese in die integrierbare Funktion $(x, y) \mapsto f(x - y)g(y)$ über. Nun existiert nach dem Satz von Fubini für fast alle $x \in \mathbb{R}^n$ das Integral

$$(f * g)(x) := \int_{\mathbb{R}^n} f(x - y)g(y)\, d\mu_n(y)\,, \tag{3.2}$$

und die so fast überall definierte Funktion $f * g$ nennt man die Faltung von f und g. Das folgende Lemma gibt einige Eigenschaften der Faltung an.

Lemma 3.98 *Sind f, g integrierbare Funktionen auf dem $\mathbb{R}^n$, dann ist auch $f * g$ integrierbar mit*

$$\int_{\mathbb{R}^n} (f * g)(x)\, d\mu_n(x) = \left(\int_{\mathbb{R}^n} f(x)\, d\mu_n(x)\right) \cdot \left(\int_{\mathbb{R}^n} g(y)\, d\mu_n(y)\right)$$

und es gilt $\|f * g\|_{L^1} \leq \|f\|_{L^1}\|g\|_{L^1}$. Außerdem ist die Faltung bilinear, assoziativ, kommutativ und erfüllt $\mathrm{Tr}(f * g) \subset \mathrm{Tr}(f) + \mathrm{Tr}(g)$.

Beweis:

- Wegen der Translationsinvarianz des Lebesgue-Maßes gilt nach dem Satz von Fubini

$$\int_{\mathbb{R}^n} (f * g)(x)\, d\mu_n(x) = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-y)g(y)\, d\mu_n(y)\right) d\mu_n(x) =$$
$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-y)g(y)\, d\mu_n(x)\right) d\mu_n(y) = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x)g(y)\, d\mu_n(x)\right) d\mu_n(y) =$$
$$\left(\int_{\mathbb{R}^n} f(x)\, d\mu_n(x)\right) \cdot \left(\int_{\mathbb{R}^n} g(y)\, d\mu_n(y)\right).$$

- Insbesondere gilt wegen $|f * g| \leq |f| * |g|$ auch $\|f * g\|_{L^1} \leq \|f\|_{L^1}\|g\|_{L^1}$.
- Bilinearität von $(f, g) \mapsto f * g$ ist offensichtlich.
- Wendet man die Transformation $y \mapsto x - y$ an, so ergibt sich

$$(f * g)(x) = \int_{\mathbb{R}^n} f(x-y)g(y)\, d\mu_n(y) = \int_{\mathbb{R}^n} f(y)g(x-y)|\det(-\mathrm{Id})|\, d\mu_n(y) =$$
$$\int_{\mathbb{R}^n} g(x-y)f(y)\, d\mu_n(y) = (g * f)(x),$$

  also die Kommutativität von $*$.

- Assoziativität von $*$ gilt wegen

$$(f * (g * h))(x) \int_{\mathbb{R}^n} f(x-y)(g * h)(y)\, d\mu_n(y) =$$
$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-y)g(y-z)h(z)\, d\mu_n(z)\right) d\mu_n(y) =$$
$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-y)g(y-z)h(z)\, d\mu_n(y)\right) d\mu_n(z) =$$
$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(y)g(x-y-z)h(z)\, d\mu_n(y)\right) d\mu_n(z) =$$
$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} h(z)g(x-z-y)f(y)\, d\mu_n(y)\right) d\mu_n(z) =$$
$$\int_{\mathbb{R}^n} h(z)(g * f)(x-z)\, d\mu_n(z) = (g * f) * h$$

und somit $f * (g * h) = (g * f) * h = (f * g) * h$ wegen $f * g = g * f$.

- Gilt $(f * g)(x) \neq 0$, so gibt es ein $y \in \mathrm{Tr}(g)$ mit $x - y \in \mathrm{Tr}(f)$. Also ist $x = (x - y) + y \in \mathrm{Tr}(f) + \mathrm{Tr}(g)$.

□

Man kann den Wert $(f * g)(x)$ für eine Funktion g mit $g \geq 0$, $\mathrm{Tr}(g) \subset B_r(0)$ und $\|g\|_{L^1} = 1$ als das mit g gewichtete Mittel von f in $B_r(x)$ ansehen, denn

$$(f * g)(x) = \int_{B_r(x)} f(y) g(x - y) \, d\mu_n(y)$$

gilt aufgrund der Kommutativität $f * g = g * f$ und wegen $\mathrm{Tr}(g) \subset B_r(0)$. Diese Beobachtung liefert somit eine anschauliche Interpretation der Faltung.

Eine sehr nützliche Eigenschaft der Faltung ist die folgende Differentiationsregel.

Lemma 3.99 *Ist f eine integrierbare Funktion auf $\mathbb{R}^n$ und $g \in C_c^\infty(\mathbb{R}^n)$, dann ist $f * g$ beliebig oft differenzierbar mit partiellen Ableitungen*

$$\partial^k (f * g) = f * (\partial^k g) \,,$$

wobei ∂^k für $k \in \mathbb{N}_0^n$ angibt, wie häufig nach welcher Variable abgeleitet wird.

Beweis: Nach Satz 3.64 ist das parameterabhängige Integral $x \mapsto \int_{\mathbb{R}^n} f(y) g(x-y) \, d\mu_n(y)$ dann k-mal stetig differenzierbar, wenn für jedes feste y die Funktionen $x \mapsto f(y) g(x-y)$ k-mal stetig differenzierbar sind (diese sind hier sogar C_c^∞-Funktionen) und wenn die Ableitungen bis zur Ordnung k durch eine integrierbare Funktion majorisiert werden (was hier der Fall ist, denn ist M eine Schranke aller Ableitungen von g bis zur Ordnung k, dann gilt $|f(y) \partial^k g(x - y)| \leq M |f(y)|$ mit der integrierbaren Funktion f). Also ist $(f * g)(x) = (g * f)(x)$ k-mal stetig differenzierbar und es gilt

$$\partial^k (f * g) = \int_{\mathbb{R}^n} f(y) (\partial^k g)(x - y) \, d\mu_n(y) = f * (\partial^k g) \,.$$

□

Man kann also selbst unstetige Funktionen f glätten, indem man sie mit einer glatten Funktion faltet. Tatsächlich kann man durch die Faltung sogar eine Folge glatter Funktionen gewinnen, die gegen f konvergiert. Dazu betrachtet man Folgen g_k von glatten und integrierbaren Funktionen mit der Eigenschaft $g_k \geq 0$ $\|g_k\|_{L^1} = 1$ und

$$\lim_{k \to \infty} \int_{\mathbb{R}^n \setminus B_r(0)} g_k(x) \, d\mu_n(x) = 0$$

für jedes $r > 0$, sogenannte Dirac-Folgen. Die vorige Eigenschaft besagt, dass der für den Integralwert Eins von g_k wesentliche Teil für große k in der Nähe des Ursprungs konzentriert ist.

Beispiel 3.100 *Mit $g(x) := \frac{1}{(2\pi)^{n/2}} e^{-\|x\|_2^2/2}$ ist $g_k(x) := k^n g(kx)$ eine Dirac-Folge.*

Satz 3.101 *Sei g_k eine Dirac-Folge glatter Funktionen. Dann konvergiert die Folge glatter Funktionen $f * g_k$ bzgl. der L^1-Norm gegen f.*

Beweis: Zu zeigen ist nur die L^1-Konvergenz. Diese beweisen wir zuerst für den Fall, dass $f = 1_Q$ die charakteristische Funktion eines Quaders $Q \subset \mathbb{R}^n$ ist.

Wegen $\int_{\mathbb{R}^n} g_k(y)\, d\mu_n(y) = 1$ gilt $1_Q(x) = \int_{\mathbb{R}^n} 1_Q(x) g_k(y)\, d\mu_n(y)$ und daher wegen $g_k \geq 0$

$$\|1_Q - (1_Q * g_k)\|_{L^1} \leq \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |1_Q(x) - 1_Q(x-y)|\, g_k(y)\, d\mu_n(y) \right) d\mu_n(x)\,.$$

Vertauschung der Integrationsreihenfolge liefert

$$\|1_Q - (1_Q * g_k)\|_{L^1} \leq \int_{\mathbb{R}^n} g_k(y) \left(\int_{\mathbb{R}^n} |1_Q(x) - 1_Q(x-y)|\, d\mu_n(x) \right) d\mu_n(y)\,,$$

wobei das innere Integral $\int_{\mathbb{R}^n} |1_Q(x) - 1_Q(x-y)|\, d\mu_n(x)$ das Volumen von $Q\Delta(y+Q)$ ist. Nun gibt es einerseits wegen der Stetigkeit des Volumens zu $\epsilon > 0$ ein $r > 0$, so dass für $y \in B_r(0)$ die Ungleichung

$$\int_{\mathbb{R}^n} |1_Q(x) - 1_Q(x-y)|\, d\mu_n(x) = \mathrm{Vol}_n(Q\Delta(y+Q)) \leq \epsilon$$

gilt. Andererseits läßt sich $\mathrm{Vol}_n(Q\Delta(y+Q))$ für beliebige $y \in \mathbb{R}^n$ durch $2\,\mathrm{Vol}_n(Q)$ abschätzen, und da g_k eine Dirac-Folge ist, gibt es zu obigem $\epsilon > 0$ ein $K \in \mathbb{N}$, so dass $\int_{\mathbb{R}^n \setminus B_r(0)} g_k(x)\, d\mu_n(x) \leq \epsilon$ für $k \geq K$ gilt. Daher gilt für $k \geq K$

$$\|1_Q - (1_Q * g_k)\|_{L^1} \leq \epsilon \int_{B_r(0)} g_k(y)\, d\mu_n(y) + 2\,\mathrm{Vol}_n(Q) \int_{\mathbb{R}^n \setminus B_r(0)} g_k(y)\, d\mu_n(y) \leq (1 + 2\,\mathrm{Vol}_n(Q))\epsilon\,,$$

und somit konvergiert $1_Q * g_k$ bzgl. der L^1-Norm gegen 1_Q.

Aufgrund der Linearität von $f \mapsto f * g_k$ konvergiert dann auch $f * g_k$ in $L^1(\mathbb{R}^n)$ für jede Treppenfunktion f zu Quadern gegen g_k.

Sei zu guter Letzt $f \in L^1(\mathbb{R}^n)$ beliebig und $\epsilon > 0$ vorgegeben. Dann gibt es eine Treppenfunktion $\tilde{f}$ mit $\|f - \tilde{f}\|_1 \leq \epsilon/3$ und einen Index $K \in \mathbb{N}$ mit $\|\tilde{f} - (\tilde{f} * g_k)\|_1 \leq \epsilon/3$ für $k \geq K$. Wegen $\|g_k\|_1 = 1$ und also $\|(\tilde{f} - f) * g_k\|_1 \leq \|f - \tilde{f}\|_1$ ergibt sich somit

$$\|f - (f * g_k)\|_1 \leq \|f - \tilde{f}\|_1 + \|\tilde{f} - (\tilde{f} * g_k)\|_1 + \|(\tilde{f} - f) * g_k\|_1 \leq \epsilon\,,$$

d.h. $f * g_k$ konvergiert bzgl. der L^1-Norm gegen f. $\square$

Bemerkung 3.102 *Eine Dirac-Folge besitzt keine L^1-Funktion als Grenzwert. In der Distributionentheorie erweitert man aber den Funktionen-Begriff so, dass Dirac-Folgen gegen eine Distribution konvergieren, die sogenannte Delta-Distribution.*

Kontinuierliche Fourier-Transformation Die kontinuierliche Fourier-Transformierte einer integrierbaren Funktion $f : \mathbb{R}^n \to \mathbb{C}$ ist die durch

$$\hat{f}(\xi) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{-i\langle \xi, x\rangle} \, d\mu_n(x)\,,$$

definierte stetige Funktion $\hat{f} : \mathbb{R}^n \to \mathbb{C}$. Tatsächlich ist $\hat{f}$ wohldefiniert und stetig nach Satz 3.63 über die stetige Abhängigkeit des Lebesgue-Integrals von Parametern, da der Integrand $f(x)e^{-i\langle \xi,x\rangle}$ die integrierbare Majorante $|f(x)|$ hat und $\xi \mapsto f(x)e^{-i\langle \xi,x\rangle}$ für festes x stetig ist.

Beispiel 3.103 *Die kontinuierliche Fourier-Transformierte von $g(x) := \frac{1}{(2\pi)^{n/2}} e^{-\|x\|_2^2/2}$ ist die Funktion $\hat{g}(\xi) = g(\xi)$ selbst. Denn mittels der quadratischen Ergänzung $x_j^2 + 2i\xi_j x_j - \xi_j^2 = (x_j + i\xi_j)^2$ erhält man aufgrund der Translationsinvarianz des Lebesgue-Maßes*

$$\hat{g}(\xi) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-\|x\|_2^2/2} e^{-i\langle \xi,x\rangle} \, d\mu_n(x) =$$

$$\frac{1}{(2\pi)^n} \prod_{j=1}^n \left(\int_{-\infty}^{\infty} e^{-(x_j + i\xi_j)^2/2} \, d\mu_n(x) \right) e^{-\xi_j^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-\|\xi\|_2^2/2}\,.$$

Beispiel 3.104 *Nicht jede kontinuierliche Fourier-Transformierte $\hat{f}(\xi)$ einer L^1-Funktion ist wieder eine L^1-Funktion. So hat beispielsweise die charakteristische Funktion $1_{[-1,1]}$ des Intervalls $[-1,1]$ die kontinuierliche Fourier-Transformierte*

$$\hat{1}_{[-1,1]}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-i\langle \xi,x\rangle} \, dx = \frac{2}{\sqrt{2\pi}} \frac{\sin(\xi)}{\xi}$$

aber $\xi \mapsto \frac{\sin(\xi)}{\xi}$ ist nach Beispiel 3.69 nicht Lebesgue-integrierbar, sondern nur uneigentlich Riemann-integrierbar.

Es gilt aber der folgende Satz.

Satz 3.105 *Ist $f : \mathbb{R}^n \to \mathbb{C}$ eine Lebesgue-integrierbare Funktion, deren kontinuierliche Fourier-Transformierte $\hat{f}$ wiederum Lebesgue-integrierbar ist, dann gilt fast überall*

$$f(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{i\langle x,\xi\rangle} \, d\mu_n(\xi)\,.$$

Beweis: Die Faltung $f * g_k$ von f mit der Dirac-Folge g_k, die durch $g_k(x) := k^n g(kx)$ bei $g(x) := \frac{1}{(2\pi)^{n/2}} e^{-\|x\|_2^2/2}$ definiert wird, konvergiert für $k \to \infty$ bzgl. der L^1-Norm gegen f.

Nun gilt aber für die kontinuierliche Fourier-Transformierte von g nach Beispiel 3.103 die Beziehung $\hat{g} = g$ und somit

$$g_k(x) = k^n \hat{g}(kx) = \frac{k^n}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-\|\xi\|_2^2/2} e^{-i\langle kx,\xi\rangle} \, d\mu_n(\xi) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-\|\xi\|_2^2/2k^2} e^{i\langle x,\xi\rangle} \, d\mu_n(\xi)\,,$$

wobei die Transformation $-k\xi \mapsto \xi$ benutzt wurde. Also ergibt sich durch Vertauschung der Integrationsreihenfolge

$$(f * g_k)(x) = \int_{\mathbb{R}^n} f(y) g_k(x-y)\, d\mu_n(y) =$$
$$\frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(y) e^{-\|\xi\|_2^2/2k^2} e^{i\langle x-y,\xi\rangle}\, d\mu_n(\xi) \right) d\mu_n(y) =$$
$$\frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-\|\xi\|_2^2/2k^2} e^{i\langle x,\xi\rangle} \left(\int_{\mathbb{R}^n} f(y) e^{-i\langle y,\xi\rangle}\, d\mu_n(y) \right) d\mu_n(\xi) =$$
$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-\|\xi\|_2^2/2k^2} e^{i\langle x,\xi\rangle} \hat{f}(\xi)\, d\mu_n(\xi)$$

Der Integrand auf der rechten Seite konvergiert für $k \to \infty$ punktweise gegen $\hat{f}(\xi) e^{i\langle x,\xi\rangle}$ und wird durch die als integrierbar vorausgesetzte Funktion $|\hat{f}|$ majorisiert, und die linke Seite konvergiert bzgl. der L^1-Norm für $k \to \infty$ gegen f, insbesondere konvergiert also eine Teilfolge auch punktweise fast überall gegen f. Daher gilt wie gewünscht

$$f(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{i\langle x,\xi\rangle}\, d\mu_n(\xi)\,.$$

□

Somit ist die Umkehrung der kontinuierlichen Fourier-Transformation zumindest für diejenigen Funktionen f, deren kontinuierliche Fourier-Transformierte integrierbar ist, durch

$$\hat{f} \mapsto \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{i\langle x,\xi\rangle}\, d\mu_n(\xi)$$

gegeben. Wir werden im letzten Abschnitt in Analogie zur diskreten Fourier-Transformation zeigen, dass die kontinuierliche Fourier-Transformation auf $L^1(\mathbb{R}^n)$ eine Bijektion von $L^2(\mathbb{R}^n)$ nach $L^2(\mathbb{R}^n)$ induziert.

Zuvor aber wenden wir uns noch zwei Sätzen zu, mit denen man leicht die kontinuierliche Fourier-Transformation einer Faltung bzw. einer Ableitung ausrechnen kann.

Satz 3.106 *Für $f, g \in L^1(\mathbb{R}^n)$ gilt $\widehat{f * g}(\xi) = (2\pi)^{n/2} \hat{f}(\xi) \cdot \hat{g}(\xi)$.*

Beweis: Der Faltungssatz folgt mittel des Transformationssatzes aus

$$\widehat{(f * g)}(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} (f * g)(x) e^{-i\langle \xi,x\rangle}\, d\mu_n(x) =$$
$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-y) g(y)\, d\mu_n(y) \right) e^{-i\langle \xi,x\rangle}\, d\mu_n(x) =$$
$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x) g(y) e^{-i\langle \xi,x+y\rangle} d\mu_n(x) \right) d\mu_n(y) =$$
$$(2\pi)^{n/2} \left(\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{-i\langle \xi,x\rangle}\, d\mu_n(x) \right) \cdot \left(\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} g(y) e^{-i\langle \xi,y\rangle}\, d\mu_n(y) \right) =$$
$$(2\pi)^{n/2} \hat{f}(\xi) \cdot \hat{g}(\xi)\,.$$

□

Satz 3.107 *Ist $f \in L^1(\mathbb{R}^n)$ eine stetig differenzierbare Funktion mit $\frac{\partial f}{\partial x_j} \in L^1(\mathbb{R}^n)$, dann gilt $\widehat{\frac{\partial f}{\partial x_j}}(\xi) = i\xi_j \cdot \hat{f}(\xi)$.*

Beweis: Mittels partieller Integration ergibt sich aufgrund des Verschwindens von $f(x)$ für $|x_j| \to \infty$ die Gleichung

$$\widehat{\frac{\partial f}{\partial x_j}}(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \frac{\partial f}{\partial x_j}(x) e^{-i\langle \xi, x\rangle}\, d\mu_n(x) =$$
$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^{n-1}} \left(\int_{-\infty}^{\infty} \frac{\partial f}{\partial x_j}(x) e^{-i\langle \xi, x\rangle}\, dx_j \right) d\mu_{n-1}(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) =$$
$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^{n-1}} \left(f(x) e^{-i\langle \xi, x\rangle}\Big|_{x_j=-\infty}^{\infty} + i\xi_j \int_{-\infty}^{\infty} f(x) e^{-i\langle \xi, x\rangle}\, dx_j \right) d\mu_{n-1}(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) =$$
$$\frac{1}{(2\pi)^{n/2}} i\xi_j \int_{\mathbb{R}^n} f(x) e^{-i\langle \xi, x\rangle}\, d\mu_n(x) = i\xi_j \hat{f}(\xi)\,.$$

□

Dieser Differentiationssatz für die kontinuierliche Fourier-Transformation läßt sich besonders gut anwenden, wenn man Differentialgleichungen lösen will. Aus einer Differentialgleichung entsteht nämlich durch Fourier-Transformation eine algebraische Gleichung für die Fourier-Transformierte $\hat{f}$. Hat man deren Lösung ermittelt, so muß man nur noch die Umkehrung der Fourier-Transformation anwenden und erhält dann die Lösung der Differentialgleichung.

Beispiel 3.108 *Die inhomogene lineare partielle Differentialgleichung*

$$-(\Delta u)(x,y) + \lambda u(x,y) := -\frac{\partial^2 u}{\partial x^2}(x,y) - \frac{\partial^2 u}{\partial y^2}(x,y) + \lambda u(x,y) = f(x,y)\,,$$

für eine L^1-Funktion $u : \mathbb{R}^2 \to \mathbb{R}$ bei vorgegebenem $f \in L^1(\mathbb{R}^2)$ mit $\hat{f} \in L^1(\mathbb{R}^2)$ geht unter der Fourier-Transformation in die algebraische Gleichung $(\xi_1^2 + \xi_2^2 + \lambda)\hat{u}(\xi) = \hat{f}(\xi)$ über. Für $\lambda > 0$ kann man ihre Lösung daher durch

$$u(x,y) = \frac{1}{2\pi} \int_{\mathbb{R}^2} \frac{\hat{f}(\xi)}{\xi_1^2 + \xi_2^2 + \lambda} e^{i\langle (x,y), (\xi_1, \xi_2)\rangle}\, d\mu_2(\xi_1, \xi_2)$$

angeben, da der Integrand die integrierbare Majorante $\frac{1}{\lambda}\hat{f}$ besitzt.

Bemerkung 3.109 *Ähnlich ergibt sich für die diskrete Fourier-Transformation mit der durch $(f*g)(x) := \frac{1}{(2\pi)^n} \int\limits_{(-\pi,\pi)^n} f(x-y) g(y)\, d\mu_n(y)$ definierten Faltung zweier Funktionen $f, g \in L^1_{per}(\mathbb{R}^n, \mathbb{C})$ die Gleichung $\widehat{f*g}(k) = \hat{f}(k) \cdot \hat{g}(k)$ und für stetig differenzierbare $f \in L^1_{per}(\mathbb{R}^n, \mathbb{C})$ die Gleichung $\widehat{\frac{\partial f}{\partial x_j}}(k) = ik_j \cdot \hat{f}(\xi)$.*

L^2**-Fourier-Theorie** In diesem letzten Abschnitt wollen wir zeigen, dass die kontinuierliche Fourier-Transformation eine Bijektion von $L^2(\mathbb{R}^n)$ nach $L^2(\mathbb{R}^n)$ induziert. Allerdings muß man zunächst einmal beachten, dass nicht jede L^2-Funktion auf dem $\mathbb{R}^n$ auch eine L^1-Funktion ist.

Beispiel 3.110 *Die durch $f(x) := \frac{1}{x}$ für $|x| \geq 1$ und $f(x) := 0$ sonst definierte Funktion $f : \mathbb{R} \to \mathbb{R}$ ist quadrat-integrierbar, aber nicht integrierbar über ganz $\mathbb{R}$.*

Somit kann man mit der obigen Definition nicht einfach naiv die kontinuierliche Fourier-Transformierte einer Funktion $f \in L^2(\mathbb{R}^n)$ bilden, sondern nur die Fourier-Transformierte von Funktionen $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$. Allerdings liegen solche Funktionen, zu denen Beispielsweise alle Funktionen aus $C_c(\mathbb{R}^n)$ gehören, dicht in $L^2(\mathbb{R}^n)$.

Mittels des folgenden Satzes von Plancherel können wir aber nachweisen, dass die Fourier-Transformation auf der dichten Teilmenge $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ von $L^2(\mathbb{R}^n)$ eine Isometrie ist.

Satz 3.111 *Für $f, g \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ gilt*

$$\int_{\mathbb{R}^n} \overline{\hat{f}(\xi)} \cdot \hat{g}(\xi) d\mu_n(\xi) = \int_{\mathbb{R}^n} \overline{f(x)} \cdot g(x) d\mu_n(x) \,.$$

Beweis: Wir brauchen den Satz von Plancherel nur für $f, g \in C_c(\mathbb{R}^n)$ zu beweisen, da man jede Funktion in $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ durch solche Funktionen approximieren kann und das L^2-Skalarprodukt stetig ist. Den Beweis für solche Funktionen führen wir in drei Schritten:

- Es gilt
$$\int_{\mathbb{R}^n} h(\xi)\hat{g}(\xi)e^{i\langle\xi,y\rangle}\, d\mu_n(\xi) = \int_{\mathbb{R}^n} \hat{h}(x)g(x+y)\, d\mu_n(x)$$
wegen
$$\begin{aligned}\int_{\mathbb{R}^n} h(\xi)\hat{g}(\xi)e^{i\langle\xi,y\rangle} d\mu_n(\xi) &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} h(\xi)g(x)e^{-i\langle\xi,x-y\rangle} d\mu_n(x) \right) d\mu_n(\xi) = \\ &\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} h(\xi)e^{-i\langle\xi,x\rangle} g(x+y) d\mu_n(x) \right) d\mu_n(\xi) = \\ &\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} h(\xi)e^{-i\langle\xi,x\rangle} d\mu_n(\xi) \right) g(x+y) d\mu_n(x) = \\ &\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \hat{h}(x)g(x+y)d\mu_n(x)\end{aligned}$$

- Setzen wir $y = 0$, so folgt
$$\int_{\mathbb{R}^n} h(\xi) \cdot \hat{g}(\xi)d\mu_n(\xi) = \int_{\mathbb{R}^n} \hat{h}(x) \cdot g(x)d\mu_n(x)$$

- Bezeichnen wir die inverse Fourier-Transformation aus Satz 3.105 mit $\check{\cdot}$, so folgt die Behauptung, indem wir in die vorige Gleichung

$$h(\xi) := \overline{\hat{f}}(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \overline{f(x)} e^{i\langle \xi, x\rangle} \, d\mu_n(x) = \check{\bar{f}}(\xi)$$

einsetzen, denn dann ergibt sich wegen $\hat{\check{\bar{f}}} = \bar{f}$

$$\int_{\mathbb{R}^n} \overline{\hat{f}}(\xi) \cdot \hat{g}(\xi) d\mu_n(\xi) = \int_{\mathbb{R}^n} \overline{f(x)} \cdot g(x) d\mu_n(x)$$

□

Wendet man den Satz von Plancherel auf $g := f$ an, so ergibt sich das folgende Korollar.

Korollar 3.112 *Für $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ gilt die Parsevalsche Gleichung*

$$\int_{\mathbb{R}^n} \left|\hat{f}(\xi)\right|^2 d\mu_n(\xi) = \int_{\mathbb{R}^n} |f(x)|^2 \, d\mu_n(x)$$

Die Parsevalsche Gleichung besagt, dass die kontinuierliche Fourier-Transformation eine Isometrie von $L^2(\mathbb{R}^n)$ nach $L^2(\mathbb{R}^n)$ ist. Nähert man also eine beliebige Funktion $f \in L^2(\mathbb{R}^n)$ durch Funktionen $f_k \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ bzgl. der L^2-Norm an (was aufgrund der Dichtheit von $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ in $L^2(\mathbb{R}^n)$ möglich ist), dann bildet $\hat{f}_k$ eine Cauchy-Folge in $L^2(\mathbb{R}^n)$ wegen $\|\hat{f}_k - \hat{f}_l\|_{L^2} = \|f_k - f_l\|_{L^2}$ und konvergiert somit gegen eine L^2-Funktion, die unabhängig von der Wahl der gegen f konvergenten Folge f_k ist und deshalb von uns durch $\hat{f}$ symbolisiert wird. Die so definierte Abbildung $f \mapsto \hat{f}$ auf ganz L^2 bezeichnet man wiederum als kontinuierliche Fourier-Transformation (auf $L^2(\mathbb{R}^n)$). Die so definierte Abbildung $\hat{\cdot} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ ist eine Bijektion, denn aufgrund der Erhaltung der L^2-Norm ist sie injektiv, und nach dem Inversionssatz für L^1-Funktionen, deren kontinuierliche Fourier-Transformation selbst eine L^1-Funktion ist, ist sie auf einer dichten Menge invertierbar, also wegen der Normerhaltung insgesamt invertierbar. Dies halten wir im folgenden abschließenden Satz fest.

Satz 3.113 *Die wie oben konstruierte kontinuierliche Fourier-Transformation ist eine Bijektion von $L^2(\mathbb{R}^n)$ auf $L^2(\mathbb{R}^n)$ und sogar eine Isometrie.*

Literaturverzeichnis

[Amann,Escher II] HERBERT AMANN, JOACHIM ESCHER, *Analysis II*, Birkhäuser, 1999.

[Amann,Escher III] HERBERT AMANN, JOACHIM ESCHER, *Analysis III*, Birkhäuser, 1999.

[Elstrodt] JÜRGEN ELSTRODT, *Maß- und Integrationstheorie*, Springer, 1999.

[Forster II] OTTO FORSTER, *Analysis II*, Vieweg, 1999.

[Forster III] OTTO FORSTER, *Analysis III*, Vieweg, 2008.

[Hatcher] ALLEN HATCHER, *Algebraic topology*, 2001.

[Heuser II] HARRO HEUSER, *Analysis II*, Teubner, 2004.

[Heuser III] HARRO HEUSER, *Funktionalanalysis*, Teubner, 1992.

[Jänich1] KLAUS JÄNICH, *Lineare Algebra*, Springer, 1996.

[Jänich2] KLAUS JÄNICH, *Topologie*, Springer, 2005.

[Königsberger I] KONRAD KÖNIGSBERGER, *Analysis I*, Springer, 1995.

[Königsberger II] KONRAD KÖNIGSBERGER, *Analysis II*, Springer, 1993.

[Lang] SERGE LANG, *Real and Functional Analysis*, Springer, 1993.

[Mathieu] MARTIN MATHIEU, *Funktionalanalysis*, Spektrum Akademischer Verlag, 1998.

[Rudin] W. RUDIN, *Analysis*, Oldenbourg, 2005.

[Walter II] W. WALTER, *Analysis II*, Springer, 2002.