Analysis 2

Delio Mugnolo

delio.mugnolo@uni-ulm.de

(Version von 13. Februar 2013)

Dies ist das Skript zur Vorlesung *Analysis 2*, welche ich im Wintersemester 2012 an der Universität Ulm gehalten habe.

Es ist durchaus möglich, dass ich im Text Fehler vergessen habe. Ich werde jedem/jeder dankbar sein, der/die mich darauf hinweisen wird – am besten direkt an `delio.mugnolo@uni-ulm.de`.

Ulm, den 13. Februar 2013 Delio Mugnolo

Inhaltsverzeichnis

Kapitel 1. Konvergenz von Funktionenfolgen und -reihen 5
1.1. Funktionenfolgen 5
1.2. Funktionenreihen 13
1.3. Approximierbarkeit und Taylorentwicklung 17

Kapitel 2. Integralrechnung 29
2.1. Uneigentliche Integrale 50

Kapitel 3. Einige Anwendungen der Integralrechnung 53
3.1. Konvergenz von Reihen 53
3.2. Taylor-Entwicklung einer Stammfunktionenreihe 54
3.3. Der Satz von Weierstraß 54
3.4. Existenz von transzendenten Zahlen (optional) 58

Kapitel 4. Differenzialrechnung von Funktionen mehrerer Variablen 61
4.1. Beschränkte lineare Operatoren 61
4.2. Fréchet-Differenzierbarkeit 63
4.3. Richtungsdifferenzierbarkeit 66
4.4. Der Spezialfall von $Y = \mathbb{R}^n$ 67

Kapitel 5. Einige Anwendungen der Differenzialrechnung 85

Kapitel 6. Metrische Räume und Kompaktheit 97

Literaturverzeichnis 103

KAPITEL 1

Konvergenz von Funktionenfolgen und -reihen

In [**6**] haben wir den Begriff von konvergenten Folgen ganz allgemein eingeführt als Abbildungen von einer abzählbarer Menge (insbesondere, $\mathbb{N}$) nach einen normierten Raum. Insbesondere können wir als normierten Raum $B(A;X)$, den Vektorraum der beschränkten Funktionen von A nach X; oder $C_b(A;X)$, den Vektorraum der stetigen *und* beschränkten Funktionen von A nach X, betrachten [1] und Folgen von Funktionen auf Konvergenz zu untersuchen. Dabei wiederholen wir, dass wir $B(A;X)$ und $C_b(A;X)$ mit der kanonischen Norm

$$\|f\|_\infty := \sup_{x\in A} \|f(x)\|_X \tag{1.1}$$

versehen. In diesem Sinne leiten wir natürlich einen Konvergenzbegriff für Folgen stetiger Funktionen ab: Eine Folge $(f_n)_{n\in\mathbb{N}}$ beschränkter Funktionen gilt als konvergent (und man spricht oft von *Konvergenz in ∞-Norm*), falls sie im Sinne von [**6**, Definition 4.24] konvergiert – d.h., falls es ein $f \in B(A;X)$ bfzw. $f \in C_b(A;X)$ gibt, für das (1.1) gilt, d.h.,

$$\forall \varepsilon > 0\ \exists N_\varepsilon \in \mathbb{N} \text{ s.d. } \|f_n - f\|_\infty < \varepsilon \text{ für alle } n \geq N_\varepsilon.$$

1.1. Funktionenfolgen

Es stellt sich aber schnell heraus, dass neben diesem auch andere Konvergenzbegriffe sinnvoll erscheinen, zumal (1.1) nicht nur für beschränkte Funktionen sinnvoll ist – dazu reicht es lediglich, dass $f_n - f$ für alle $n \in \mathbb{N}$ beschränkt bleibt.

Definition 1. *Es seien X, Y Normierte Räume, A eine Teilmenge von Y und $(f_n)_{n\in\mathbb{N}}$ eine Folge von Funktionen von A nach X.*

- *Die Folge $(f_n)_{n\in\mathbb{N}}$ heißt* gleichmäßig konvergent*, falls es ein $f : A \to X$ gibt, für das*
$$\forall \varepsilon > 0\ \exists N_\varepsilon \in \mathbb{N} \textit{ s.d. } \|f_n(x) - f(x)\|_X < \varepsilon \textit{ für alle } n \geq N_\varepsilon \textit{ und alle } x \in A.$$
gilt.
- *Die Folge $(f_n)_{n\in\mathbb{N}}$ heißt* punktweise konvergent*, falls jede Punktauswertung $(f_n(x))_{n\in\mathbb{N}} \subset X$ konvergiert – d.h., falls es ein $g : A \to X$ gibt, für das*
$$\forall \varepsilon > 0\ \forall x \in A\ \exists N_\varepsilon \in \mathbb{N} \textit{ s.d. } \|f_n(x) - g(x)\|_X < \varepsilon \textit{ für alle } n \geq N_\varepsilon$$
gilt.

Definitionsgemäß stimmen die Begriffe der Konvergenz in ∞-Norm bzw. der gleichmäßigen Konvergenz für Folgen stetiger Funktionen überein. Man kann diese Beobachtung leicht verallgemeinern.

[1] Einen Spezialfall haben wir in [**6**, Korollar 7.87 und Anmerkung 7.89] betrachtet: Ist $A \subset \mathbb{R}^n$, dann hat jede Funktion von A nach $\mathbb{R}$ ein globales Maximum und somit ist $C_b(A;X) = C(A;X)$ ein abgeschlossener Unterraum des Banachraumes $B(A;X)$ und somit selber ein Banachraum

Satz 1.1. *Es seien X, Y Normierte Räume, A eine Teilmenge von Y und $(f_n)_{n\in\mathbb{N}}$ eine Folge von Funktionen von A nach X, von denen nur endlich viele an einer gegebenen Stelle $x_0 \in A$* nicht *stetig sind. Konvergiert die Folge $(f_n)_{n\in\mathbb{N}}$ gleichmäßig, so ist ihr Grenzwert ebenfalls eine an der Stelle x_0 stetige Funktion.*

Beweis. Es sei $\varepsilon > 0$ und $N_\varepsilon \in \mathbb{N}$ so, dass $\|f_n - f\|_\infty < \varepsilon$. Es sei $\tilde{N} \in \mathbb{N}$ der Index, für den $f_n \in C(A;X)$ für alle $n \geq \tilde{N}$ und betrachte $\tilde{N}_\varepsilon := \max\{N_\varepsilon, \tilde{N}\}$. Wegen der Stetigkeit von $f_{\tilde{N}_\varepsilon}$ gibt es eine Umgebung U von x, in der gilt

$$\|f_{\tilde{N}_\varepsilon}(x_0) - f_{\tilde{N}_\varepsilon}(x)\|_X \varepsilon \qquad \text{für alle } x \in U.$$

Nun folgt die Aussage mit einem 3ε-Argument: Für alle $x \in U$ gilt nämlich

$$\begin{aligned}\|f(x) - f(x_0)\|_X &\leq \|f(x) - f_{\tilde{N}_\varepsilon}(x)\|_X + \|f_{\tilde{N}_\varepsilon}(x) - f_{\tilde{N}_\varepsilon}(x_0)\|_X + \|f_{\tilde{N}_\varepsilon}(x_0) - f(x_0)\|_X \\ &\leq 2\|f - f_{\tilde{N}_\varepsilon}\|_\infty + \|f_{\tilde{N}_\varepsilon}(x) - f_{\tilde{N}_\varepsilon}(x_0)\|_X \leq 3\varepsilon,\end{aligned}$$

was den Beweis vervollständigt. □

Übungsaufgabe 2. Es seien X, Y Normierte Räume, A eine Teilmenge von Y und $(f_n)_{n\in\mathbb{N}}$ eine Folge von Funktionen von A nach X. Eine Folge $(f_n)_{n\in\mathbb{N}}$ heißt *lokal gleichmäßig konvergent*, falls es ein $f : A \to X$ gibt so, dass für alle $x \in A$ eine Umgebung U_x existiert mit

$$\forall \varepsilon > 0 \; \exists N_\varepsilon \in \mathbb{N} \text{ s.d. } \|f_n(y) - f(y)\|_X < \varepsilon \text{ für alle } n \geq N_\varepsilon \text{ und alle } y \in U_x$$

gilt.

Zeige: Konvergiert eine Folge $(f_n)_{n\in\mathbb{N}}$ lokal gleichmäßig, und sind nur endlich viele Folgenglieder *nicht* stetig, so ist ihr Grenzwert ebenfalls eine stetige Funktion.

Korollar 1.2. *Es seien X, Y Normierte Räume, A eine Teilmenge von Y. Ist X ein Banachraum, so sind auch $B(A;X)$ und $C_b(A;X)$ Banachräume.*

Beweis. 1) Wir wissen schon, dass $B(A;X)$ ein normierter Raum bzgl. der ∞-Norm ist. Wir wollen nun seine Vollständigkeit zeigen.

Es sei $(f_n)_{n\in\mathbb{N}} \subset B(A;X)$ eine Cauchy-Folge (bzgl. der $\|\cdot\|_\infty$-Norm! Insbesondere ist laut [**6**, Lemma 4.33] die Folge beschränkt, d.h., $\|f_n\|_\infty \leq M$ für ein $M > 0$ und alle $n \in \mathbb{N}$), d.h.,

$$\forall \varepsilon > 0 \; \exists N_\varepsilon \in \mathbb{N} \text{ s.d. } \|f_n(x) - f_m(x)\|_X \text{ für alle } n, m \geq N_\varepsilon \text{ und alle } x \in A. \tag{1.2}$$

Dann ist auch jede Folge von Punktauswertungen $(f_n(x))_{n\in\mathbb{N}} \subset X$ eine Cauchy-Folge, die wegen der Vollständigkeit von X konvergiert. Dadurch definiert man eine Grenzfunktion $f : A \to X$, gegen die $(f_n)_{n\in\mathbb{N}}$ punktweise konvergiert.

Wir wollen nun zeigen, dass die Konvergenz sogar gleichmäßig ist. Dazu zeigen wir zuerst, dass f beschränkt ist: Denn aus (1.2) folgt (im Grenzwert $m \to \infty$), dass

$$\forall \varepsilon > 0 \; \exists N_\varepsilon \in \mathbb{N} \text{ s.d. } \|f_n(x) - f(x)\|_X < \varepsilon \text{ für alle } n \geq N_\varepsilon \text{ und alle } x \in A, \tag{1.3}$$

und Dank der Dreiecksungleichung

$$\forall \varepsilon > 0 \; \exists N_\varepsilon \in \mathbb{N} \text{ s.d. } \|f(x)\|_X \leq \|f_N\|_\infty + \varepsilon \leq M + \varepsilon \text{ für alle } x \in A,$$

und somit ist $f \in B(A;X)$. Aus (1.3) folgt schließlich, dass f auch gleichmäßiger Grenzwert von $(f_n)_{n\in\mathbb{N}}$ ist – d.h., für jede Cauchy-Folge $(f_n)_{n\in\mathbb{N}}$ gibt es $f \in B(A;X)$ so, dass $(f_n)_{n\in\mathbb{N}}$ gegen f in der Norm von $B(A;X)$ konvergiert.

2) Um nun die Vollständigkeit von $C_b(A;X)$ zu beweisen, reicht es nach [**6**, üA 7.28] zu zeigen, dass $C_b(A;X)$ ein abgeschlossener Unterraum von $B(A;X)$ ist. Betrachte dazu eine bzgl. der Norm von $C_b(A;X)$ (d.h., bzgl. der ∞-Norm) konvergente Folge $(f_n)_{n\in\mathbb{N}} \subset C_b(A;X)$. Wir wissen schon aus 1), dass der Grenzwert sicherlich auch beschränkt ist. Dass er auch stetig ist, folgt aus dem Satz 1.1. $\square$

Beispiel 3. Es ist klar, dass gleichmäßige Konvergenz auch punktweise Konvergenz impliziert. Die Umkehrung gilt nicht: Um ein Gegenbeispiel zu finden, brauchen wir angesichts des Satzes 1.1 einfach eine Folge stetiger Funktionen zu finden, welche punktweise gegen eine nichtstetige Funktion konvergiert.

Betrachte dazu die Funktionen $f_n : [0,1] \to \mathbb{R}$, welche durch

$$f_n(x) := x^n, \qquad x \in [0,1]$$

definiert sind. Dann ist $f_n(1) = 1$ für alle $n \in \mathbb{N}$ und somit $\lim_{n\to\infty} f_n(1) = 1$, während $\lim_{n\to\infty} x^n = 0$ für alle $x \in [0,1)$ (s. [**6**, üA 4.48]). Somit konvergiert $(f_n)_{n\in\mathbb{N}}$ punktweise gegen die Funktion f, welche durch

$$f_n(x) := \begin{cases} 1, & \text{falls } x = 1, \\ 0, & \text{sonst,} \end{cases}$$

definiert ist und offensichtlich nicht stetig ist. $\square$

Übungsaufgabe 4. Finde eine Folge beschränkter Funktionen, welche punktweise gegen eine nichtbeschränkte Funktion konvergiert.

Übungsaufgabe 5. Es sei Y ein normierter Raum, A eine Teilmenge von Y, und X ein Banachraum. Es sei darüber hinaus $(f_n)_{n\in\mathbb{N}}$ eine Folge von Funktionen von A nach X. Beweise die folgenden Aussagen:

- Die Funktionenfolge ist genau dann punktweise konvergent, wenn für alle $x \in A$ und alle $\varepsilon > 0$ es ein $N \in \mathbb{N}$ gibt so, dass
$$\|f_n(x) - f_m(x)\|_X < \varepsilon \qquad \forall m,n \geq N.$$
- Die Funktionenfolge ist genau dann gleichmäßig konvergent, wenn für alle $\varepsilon > 0$ es ein $N \in \mathbb{N}$ gibt so, dass
$$\|f_n(x) - f_m(x)\|_X < \varepsilon \qquad \forall m,n \geq N,\ \forall x \in A,$$
oder äquivalent, dass
$$\sup_{x\in A} \|f_n(x) - f_m(x)\|_X < \varepsilon \qquad \forall m,n \geq N.$$

Wir haben gesehen, dass der gleichmäßige Grenzwert eine Folge stetiger Funktionen selber stetig ist: das liegt im Wesentlichen daran, dass die gleichmäßige Konvergenz gerade der Norm von C_b entspricht. ähnlich kann man sich fragen, ob weitere Eigenschaften unter Grentzwertbildung invariant bleiben, etwa die stetige Differenzierbarkeit. Da für jedes Intervall $A \subset \mathbb{R}$ $C^1_b(A;X) := \{f \in C_b(A;X) \cap C^1(A;X) : f' \in C_b(A;X)\}$ ein Banachraum ist bzgl.

$$\|f\|_{C^1_b} := \|f\|_\infty + \|f'\|_\infty,$$

liegt die Vermutung nah, dass eine Folge $(f_n)_{n\in\mathbb{N}}$ stetig differenzierbarer Funktionen gegen eine stetig differenzierbare Funktion konvergiert, wenn sowohl $(f_n)_{n\in\mathbb{N}}$ als auch $(f'_n)_{n\in\mathbb{N}}$ gleichmäßig konvergieren.

Übungsaufgabe 6. Es seien X, Y Normierte Räume, A eine Teilmenge von Y und $(f_n)_{n\in\mathbb{N}}$ eine Folge differenzierbarer Funktionen von A nach X, von denen nur endlich viele an einer gegebenen Stelle $x_0 \in A$ *weder* stetig sind, *noch* mit stetiger Ableitung. Konvergiert die Folgen $(f_n)_{n\in\mathbb{N}}$ und $(f'_n)_{n\in\mathbb{N}}$ gleichmäßig, so ist ihr Grenzwert ebenfalls eine an der Stelle x_0 stetig differenzierbare Funktion.

Tatsächlich gilt es sogar eine stärkere Aussage (warum ist sie stärker als die Aussage von der obigen übungsaufgabe?), wenn wir den Spezialfall einer Folgen von Funktionen einer reellen Variablen betrachten. Zum Beweise von dieser Aussage (Satz 1.6) benötigen wir erst zwei Lemmata.

Lemma 1.3. *Es seien $a, b \in \mathbb{R}$, $a < b$, X ein Banachraum und $(f_n)_{n\in\mathbb{N}}$ einer Folge differenzierbarer Funktionen von $[a, b]$ nach X. Konvergiert die Folge $(f'_n)_{n\in\mathbb{N}}$ gleichmäßig, und gibt es darüber hinaus ein $y_0 \in [a, b]$ so, dass $(f_n(y_0))_{n\in\mathbb{N}}$ konvergiert, so konvergiert auch $(f_n)_{n\in\mathbb{N}}$ gleichmäßig.*

Dabei sollte beachtet werden, dass wir (noch) keine Aussage über die Differenzierbarkeit des Grenzwerts der Funktionenfolge $(f_n)_{n\in\mathbb{N}}$ treffen.

Beweis. Wir wollen die Cauchy-Bedingung aus der übungsaufgabe 5 überprüfen. Es sei dazu $\varepsilon > 0$. Wir wollen eine Abschätzung für

$$\|f_n(x) - f_m(x)\|_X,$$

finden, gleichmäßig in $x \in [a, b]$, für m, n groß genug. Wir fangen damit an, dass wir diesen Term mittels Dreiecksungleichung umschreiben:

$$\begin{aligned}\|f_n(x) - f_m(x)\|_X &\leq \|f_n(x_0) - f_m(x_0)\|_X + \|(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0))\|_X \\ &= \|f_n(x_0) - f_m(x_0)\|_X + \|(f_n - f_m)(x) - (f_n - f_m)(x_0)\|_X.\end{aligned}$$

Nun erfüllt die Funktion $f_n - f_m$ die Voraussetzungen des vektorwertigen Mittelwertsatzes ([**6**, Satz 8.39]) und somit

$$\|(f_n - f_m)(x) - (f_n - f_m)(x_0)\|_X \leq \sup_{\xi\in(\min\{x_0,x\},\max\{x_0,x\})} \|(f_n - f_m)'(\xi)\|_X |x - x_0|.$$

und somit bekommen wir die Abschätzung

$$\begin{aligned}\|f_n(x) - f_n(x_0)\|_X &\leq \|f_n(x_0) - f_m(x_0)\|_X + \sup_{\xi\in(\min\{x_0,x\},\max\{x_0,x\})} \|(f_n - f_m)'(\xi)\|_X |x - x_0| \\ &\leq \|f_n(x_0) - f_m(x_0)\|_X + b \sup_{\xi\in(a,b)} \|f'_n(\xi) - f'_m(\xi)\|_X\end{aligned}$$

Nun der erste Summand kann durch Anwendung der Cauchy-Bedingung für $(f_n(x_0))_{n\in\mathbb{N}}$, während der zweite durch die gleichmäßige Cauchy-Bedingung für $(f'_n)_{n\in\mathbb{N}}$, abgeschätzt werden. Dies vervollständigt den Beweis. □

Lemma 1.4. *Es seien $a, b \in \mathbb{R}$, $a < b$, X ein Banachraum und $(f_n)_{n\in\mathbb{N}}$ einer Folge differenzierbarer Funktionen von $[a, b]$ nach X. Es sei $x_0 \in [a, b]$ und definiere eine Folge von Funktionen von $[a, b] \setminus \{x_0\}$ nach X durch*

$$g_n(x) := \frac{f_n(x) - f_n(x_0)}{x - x_0}, \qquad x \in [a, b] \setminus \{x_0\}.$$

Konvergiert die Folge $(f'_n)_{n\in\mathbb{N}}$ gleichmäßig, so konvergiert auch $(g_n)_{n\in\mathbb{N}}$ gleichmäßig.

Beweis. Wir wollen wieder den vektorwertigen Mittelwertsatz anwenden. Dazu betrachten wir für beliebige $n, m \in \mathbb{N}$ den Ausdruck

$$\frac{(f_n - f_m)(x) - (f_n - f_m)(x_0)}{x - x_0} = (g_n - g_m)(x), \qquad x \in [a,b] \setminus \{x_0\}, \qquad n, m \in \mathbb{N},$$

für den dann die Abschätzung

$$\|(g_n - g_m)(x)\|_X \le \sup_{\xi \in (\min\{x_0,x\},\max\{x_0,x\})} \|(f_n - f_m)'(\xi)\|_X \le \sup_{\xi \in (a,b)} \|(f_n - f_m)'(\xi)\|_X, \qquad x \in [a,b] \setminus \{x_0\}$$

gilt. Da der letzte Term mittels der Cauchy-Bedingung abgeschätzt werden kann, gilt eine entsprechende Abschätzung auch für $\sup_{x \in [a,b]\setminus\{x_0\}} \|(g_n - g_m)(x)\|_X$. □

Lemma 1.5. *Es seien Y normierter Raum, A eine Teilmenge von Y, $x_0 \in A$, X ein Banachraum und $(f_n)_{n\in\mathbb{N}}$ eine Folge von Funktionen von A nach X. Konvergiert diese Folge gleichmäßig, etwa gegen f, und existiert darüber hinaus für alle $n \in \mathbb{N}$ der Grenzwert*

$$\lim_{x \to x_0} f_n(x) =: \ell_k,$$

so existieren beide Grenzwerte

$$\lim_{n\to\infty} \left(\lim_{x\to x_0} f_n(x) \right), \qquad \lim_{x\to x_0} \left(\lim_{n\to\infty} f_n(x) \right),$$

und sie stimmen überein.

Beweis. Die gleichmäßige Cauchy-Bedingung für $(f_n)_{n\in\mathbb{N}}$ besagt, dass für alle $\varepsilon > 0$ ein $N \in \mathbb{N}$ existiert so, dass

$$\|f_n(x) - f_m(x)\|_X \le \varepsilon \qquad \forall x \in A.$$

Im Grenzwert für $x \to x_0$ liefert die obige Abschätzung auch

$$\|\ell_n - \ell_m\|_X \le \varepsilon,$$

und aus der Cauchy-Eigenschaft folgt die Konvergenz der Folge $(\ell_n)_{n\in\mathbb{N}}$ – etwa gegen ein $\ell \in X$. Um den Beweis zu vollenden, reicht es zu zeigen, dass $\lim_{x\to x_0} f(x) = \ell$: Wir wollen somit

$$\|f(x) - \ell\|_X \le \|f(x) - f_N(x)\|_X + \|f_N(x) - \ell_N\|_X + \|\ell_N - \ell\|_X$$

betrachten und die drei rechten Terme gegen ein gegebenes ε abschätzen. Es sei dazu $N \in \mathbb{N}$ so, dass

$$\text{sowohl} \quad \|f_N(x) - f(x)\|_X < \varepsilon \qquad \text{als auch} \qquad \text{sowohl} \quad \|\ell_N - \ell\|_X < \varepsilon.$$

Da f_N stetig ist, können wir nun $\delta > 0$ finden, für das

$$\text{aus } x \in A \text{ und } \|x - x_0\|_X < \delta \qquad \|f_N(x) - \ell_N\|_X < \varepsilon \text{ folgt.}$$

Insgesamt haben wir zu jedem $\varepsilon > 0$ ein $\delta > 0$ gefunden, für das aus

$$\text{aus } x \in A \text{ und } \|x - x_0\|_X < \delta \qquad \|f(x) - \ell\|_X < 3\varepsilon \text{ folgt.}$$

Somit ist die Aussage vollständig bewiesen. □

Endlich können wir den folgenden fundamentalen Satz formulieren.

Satz 1.6. *Es seien $a, b \in \mathbb{R}$, $a < b$, X ein Banachraum und $(f_n)_{n\in\mathbb{N}}$ einer Folge differenzierbarer* [2] *Funktionen von $[a, b]$ nach X. Konvergiert die Folge $(f_n')_{n\in\mathbb{N}}$ gleichmäßig, und gibt es darüber hinaus ein $x_0 \in [a, b]$ so, dass $(f_n(x_0))_{n\in\mathbb{N}}$ konvergiert, so konvergiert auch die Folge $(f_n)_{n\in\mathbb{N}}$ gleichmäßig, und ihr Grenzwert f ist eine differenzierbare Funktion, deren Ableitung mit dem Grenzwert von $(f_n')_{n\in\mathbb{N}}$ übereinstimmt.*

Besteht unter den obigen Voraussetzungen die Folge $(f_n)_{n\in\mathbb{N}}$ aus *stetig* differenzierbaren Funktionen, so ist selbstverständlich auch f *stetig* differenzierbar, da der gleichmäßige Grenzwert der Folge stetiger Funktionen $(f_n')_{n\in\mathbb{N}}$ ebenfalls stetig ist.

Beweis. Es folgt aus dem Lemma 1.3, dass $(f_n)_{n\in\mathbb{N}}$ gleichmäßig konvergiert. Darüber hinaus konvergiert dank Lemma 1.4 die Folge $(g_n)_{n\in\mathbb{N}}$ von Funktionen von $[a, b] \setminus \{x_0\}$ nach X, welche durch

$$g_n(x) := \frac{f_n(x) - f_n(x_0)}{x - x_0}, \qquad x \in [a, b] \setminus \{x_0\},\ n \in \mathbb{N},$$

definiert ist, gleichmäßig gegen die Funktion $g : [a, b] \setminus \{x_0\} \to X$, welche durch

$$g(x) := \frac{f(x) - f(x_0)}{x - x_0}, \qquad x \in [a, b] \setminus \{x_0\},$$

definiert ist. Schließlich folgt aus dem Lemma 1.5, dass

$$\begin{aligned}
\lim_{n\to\infty} f_n'(x_0) &= \lim_{n\to\infty} \left(\lim_{x\to x_0} g_n(x) \right) \\
&= \lim_{x\to x_0} \left(\lim_{n\to\infty} g_n(x) \right) \\
&= \lim_{x\to x_0} g(x) \\
&= f'(x_0),
\end{aligned}$$

wo die letzte Identität gerade der Definition der Ableitung von f an der Stelle x_0 entspricht. □

Im allgemeinen stellt die überprüfung der gleichmäßigen Konvergenz von Funtkionenfolgen große Schwierigkeiten, während punktweise Konvergenz ist meistens leicht überprüft. Man sucht also Kriterien, welche erlauben, aus punktweise Konvergenz doch gleichmäßige Konvergenz zu schließen. Es stellt sich heraus, dass Monotonie dabei sehr hilfreich ist.

Definition 7. *Eine Folge von Funktionen $(f_n)_{n\in\mathbb{N}}$ von einer Menge A nach einer halbgeordneter Menge G heißt* monoton, *wenn sie bzgl. der Halbordnung in* [**6**, üA 3.42] *ist, d.h. wenn*

$$f_n(x) \le f_{n+1}(x) \quad \forall x \in A,\ \forall n \in \mathbb{N} \qquad \textit{(monoton wachsend)}$$

bzw.

$$f_n(x) \ge f_{n+1}(x) \quad \forall x \in A,\ \forall n \in \mathbb{N} \qquad \textit{(monoton fallend).}$$

(Dabei brauchen die einzelnen Funktionen f_n nicht selber monoton zu Sein!)

Der folgende Satz geht auf Ulisse Dini zurück.

Satz 1.7. *Es seien $A \subset \mathbb{R}^n$ kompakt und $(f_n)_{n\in\mathbb{N}}$ eine monotone Folge stetiger Funktionen von A nach $\mathbb{R}$, welche punktweise gegen eine Funktion f konvergiert. Ist f stetig, so ist die Konvergenz bereits gleichmäßig.*

[2] Aber nicht notwendigerweise *stetig* differenzierbar!

Beweis. Es sei o.B.d.A. $(f_n)_{n\in\mathbb{N}}$ monoton wachsend, und somit $f_n(x) \leq f(x)$ für alle $x \in A$ und alle $n \in \mathbb{N}$. Wäre die Konvergenz nicht gleichmäßig, so gäbe es $\varepsilon > 0$, für das zu jedem $N \in \mathbb{N}$ es $n_N \geq N_\varepsilon$ und $x_N \in A$ gibt mit

$$|f_{n_N}(x_N) - f(x_N)| = f(x_N) - f_{n_N}(x_N) \geq \varepsilon_0. \tag{1.4}$$

Somit bildet man zwei Folgen $(n_N)_{N\in\mathbb{N}} \subset \mathbb{N}$ und $(x_N)_{N\in\mathbb{N}} \subset A$: die erstere Folge konvergiert uneigentlich gegen $+\infty$, für die zweitere folgt aus der Kompaktheit von A die Existenz einer Teilfolge $(x_{N_h})_{h\in\mathbb{N}}$, die gegen einen Punkt – etwa $x_0 \in A$ – konvergiert. Insbesondere gilt aus (1.4), dass

$$f(x_N) - f_m(x_N) \geq \varepsilon, \qquad \forall N \in \mathbb{N},\ \forall m \leq n_N. \tag{1.5}$$

und insbesondere

$$f(x_{N_h}) - f_m(x_{N_h}) \geq \varepsilon, \qquad \forall N \in \mathbb{N},\ \forall m \leq n_{N_h}. \tag{1.6}$$

Somit erhalten wir im Grenzwert für $h \to \infty$ (da f und alle f_n stetig sind)

$$f(x_0) - f_m(x_0) \geq \varepsilon, \qquad \forall N \in \mathbb{N},\ \forall m \in \mathbb{N}. \tag{1.7}$$

Schließlich liefert die Grenzwertbildung für $m \to \infty$

$$0 = f(x_0) - f(x_0) \geq \varepsilon, \qquad \forall N \in \mathbb{N},\ \forall m \in \mathbb{N}. \tag{1.8}$$

Diese unmögliche Ungleichung folgt aus der Annahme, dass die Funktionenfolge nicht gleichmäßig konvergiert. Dadurch ist die Aussage bewiesen.

(Im Fall einer monoton fallende Folge, wiederhole einfach den Beweis für die Folge $(-f_n)_{n\in\mathbb{N}}$). □

Beispiel 8. Die Folge stetiger Funktionen im Beispiel 3 ist monoton fallend. Dennoch ist ihr punktweise Grenzwert nichtstetig, und somit kann der Satz von Dini nicht daran angewendet werden – tatsächlich ist erfolgt die Konvergenz nicht gleichmäßig. □

Zum Schluss dieses Abschnitts zeigen wir ein Existenzresultat für gleichmäßige konvergente Teilfolgen.

Definition 9. *Es sei $K \subset \mathbb{R}^n$ kompakt, X ein normierter Raum. Eine Funktionenfolge $(f_n)_{n\in\mathbb{N}} \subset C(K;X)$ heißt*

- **punktweise beschränkt**, *falls es für alle $x \in K$ ein $M > 0$ existiert, so dass$\sup_{n\in\mathbb{N}} \|f_n(x)\|_X \leq M$ und*
- **gleichgradig stetig**, *falls für alle $\varepsilon > 0$ es ein $\delta > 0$ gibt, so dass $\|f_n(x) - f_n(y)\|_X < \varepsilon$ für alle $n \in \mathbb{N}$ und alle $x, y \in K$ mit $|x - y| < \delta$.*

Der folgende fundamentaler Satz wurde von Giulio Ascoli 1883 und dann (sauberer) von Cesare Arzelà 1895 bewiesen, und ist somit mit dem Namen von **Satz von Ascoli–Arzelà** bekannt.

Satz 1.8. *Es seien $a, b \in \mathbb{R}$, $a < b$. Ist eine Funktionenfolge $(f_n)_{n\in\mathbb{N}} \subset C([a,b];\mathbb{R}^m)$ punktweise beschränkt und gleichgradig stetig, so hat sie eine gleichmäßig konvergente Teilfolge.*

Beweis. 1) Zuerst betrachten wir $\mathbb{Q} \cap [a,b]$. Da diese eine abzählbare Menge ist, können wir sie mithilfe einer Folge darstellen, etwa $\{x_j : j \in \mathbb{N}\} := \mathbb{Q} \cap [a,b]$. Betrachte die Folge $(f_n(x_1))_{n\in\mathbb{N}} \subset \mathbb{R}^m$, welche wegen ihrer Beschränktheit (beachte die punktweise Beschränktheit der Funktionenfolge!) und Dank dem Satz von Bolzano–Weierstraß eine konvergenten Teilfolge hat, die wir mit $(f_{n_k^1}(x_1))_{k\in\mathbb{N}}$ bezeichnen wollen, um zu bekräftigen, dass die Extraktionsregel von x_1 abhängen kann. Entsprechend hat auch $(f_{n_k^1}(x_2))_{k\in\mathbb{N}}$ eine konvergente Teilfolge $(f_{n_k^2}(x_2))_{k\in\mathbb{N}}$, und i.A. hat für jedes $j \in \mathbb{N}$ $(f_{n_k^{j-1}}(x_j))_{n\in\mathbb{N}}$

eine konvergente Teilfolge $(f_{n_k^j}(x_j))_{n\in\mathbb{N}}$. Wir können allerdings das Cantorsche Diagonalverfahren anwenden, um eine einzige Teilfolge zusammenzubasteln: Betrachte nämlich die Funktionenfolge $(f_{n_k^k})_{k\in\mathbb{N}}$, welche nach Konstruktion die Eigenschaft hat, dass

$$(f_{n_k^k}(x_j))_{k\in\mathbb{N}} \subset \mathbb{R}^m \qquad \text{für alle } j \in \mathbb{N} \text{ konvergiert.} \tag{1.9}$$

Jetzt, wo wir ein Kandidat für eine konvergente Teilfolge haben, müssen wir zeigen, dass $(f_{n_k^k})_{k\in\mathbb{N}}$ tatsächlich gleichmäßig konvergiert, und nicht nur punktweise auf $\mathbb{Q}\cap[a,b]$.

2) Bevor wir beweisen, dass die Folge $(f_{n_k^k})_{k\in\mathbb{N}}$ die gleichmäßige Cauchy-Eigenschaft genügt, lass uns anmerken, dass $[a,b]$ folgendermaßen zerlegt werden kann:

Zu jedem $\delta>0$ gibt es $q_1,\ldots,q_J\in\mathbb{Q}$, so dass $[a,b]$ in $\bigcup_{j=1}^J(q_j-\delta,q_j+\delta)$ enthalten ist.

Betrachte nämlich die ganze Zahlen $a_0 := \lfloor a\rfloor$ und $b_0 := \lfloor b\rfloor+1$ – so dass $[a,b]\subset[a_0,b_0]$ – , eine rationale Zahl $\delta_0<\delta$ und finde ein $\ell\in\mathbb{N}$, so dass $\ell\delta_0 > b_0-a_0$. Dann gilt nämlich

$$\delta > \frac{b_0-a_0}{\ell} \in \mathbb{Q}$$

und

$$[a,b]\subset[a_0,b_0]\subset\bigcup_{j=1}^{\ell}\left(a_0+\frac{j}{\ell}(b_0-a_0)-\delta, a_0+\frac{j}{\ell}(b_0-a_0)+\delta\right).$$

Man kann nun o.B.d.A. die in (1) eingeführte Folge $(x_j)_{j\in\mathbb{N}}$ so durchnummerieren, dass $x_1=q_1$, …, $x_J=q_J$ und somit gerade die Inklusion

$$[a,b]\subset\bigcup_{j=1}^{J}(x_j-\delta,x_j+\delta)$$

gilt.

3) Man will eine (in x gleichmäßige!) Abschätzung

$$\|f_{n_p^p}(x)-f_{n_q^q}(x)\| < 3\varepsilon, \qquad x\in[a,b], \tag{1.10}$$

für p,q groß genug beweisen.

Es sei nun $\varepsilon>0$ beliebig. Wegen der gleichgradigen Stetigkeit der Folge $(f_n)_{n\in\mathbb{N}}$, und somit insbesondere ihrer Teilfolge $(f_{n_k^k})_{k\in\mathbb{N}}$, gibt es ein $\delta>0$ so, dass

$$\|f_{n_k^k}(x)-f_{n_k^k}(y)\| < \varepsilon \tag{1.11}$$

für alle $k\in\mathbb{N}$ und alle $x,y\in[a,b]$ mit $|x-y|<\delta$.

Betrachte jetzt $x\in[a,b]$. Aufgrund der Aussage in (2) gibt es zu dem gerade erwähnten δ ein $j\in\{1,\ldots J\}$, so dass

$$|x-x_j|<\delta. \tag{1.12}$$

Man zerlege also den abzuschätzenden Term als

$$\|f_{n_p^p}(x)-f_{n_q^q}(x)\| \le \|f_{n_p^p}(x)-f_{n_p^p}(x_j)\| + \|f_{n_p^p}(x_j)-f_{n_q^q}(x_j)\| + \|f_{n_q^q}(x_j)-f_{n_q^q}(x)\|$$

Nun gibt es zu ε Dank (1) ein $N\in\mathbb{N}$ so, dass

$$\|f_{n_p^p}(x_j)-f_{n_q^q}(x_j)\| < \varepsilon$$

für alle $p, q > N$ gilt. Auch liefert aufgrund von (1.12) die Abschätzung (1.11)

$$\text{sowohl}\quad \|f_{n_p^p}(x) - f_{n_p^p}(x_j)\| < \varepsilon \qquad \text{als auch} \qquad \|f_{n_q^q}(x_j) - f_{n_q^q}(x)\| < \varepsilon$$

Somit ist die Ungleichung (1.10) vollständig bewiesen. □

1.2. Funktionenreihen

Wie wir bereits wissen, sind Reihen nichts anderes als Spezialfälle von Folgen. Somit kann man auch Funktionenreihen

$$\sum_{n=0}^{\infty} f_n$$

betrachten, was sich in vielen Hinsichten als nützlich erweist.

Betrachtet man eine Funktionenreihe, so kann man verschiedene Konvergenzbegriffe betrachten. Es gibt natürlich die gleichmäßige bzw. die punktweise Konvergenz (im Sinne der Definition 1) für die Folge ihrer Partialsummen, aber auch weitere Begriffe, die wir hier zusammenfassen.

DEFINITION 10. *Es seien X, Y Normierte Räume, A eine Teilmenge von Y und $(f_n)_{n\in\mathbb{N}}$ eine Folge beschränkter Funktionen von A nach X. Dann heißt die Reihe*

$$\sum_{k=0}^{\infty} f_k$$

- punktweise konvergent, *falls die Funktionenfolge $\left(\sum_{k=0}^{n} f_k\right)_{n\in\mathbb{N}}$ punktweise (im Sinne der Definition 1) konvergiert, d.h., falls $\sum_{k=0}^{\infty} f_k(x)$ für alle $x \in A$ in X konvergiert;*
- gleichmäßig konvergent, *falls die Funktionenfolge $\left(\sum_{k=0}^{n} f_k\right)_{n\in\mathbb{N}}$ gleichmäßig (im Sinne der Definition 1) konvergiert, d.h., falls $\left\|\sum_{k=0}^{n} f_k - F\right\|_{\infty}$ gegen 0 konvergiert für n gegen ∞, für ein passendes $F : A \to X$;*
- absolut konvergent, *falls $\sum_{k=0}^{\infty} \|f_k(x)\|_X$ für alle $x \in A$ in $\mathbb{R}$ konvergiert;*
- normkonvergent, *falls $\sum_{k=0}^{\infty} \|f_k\|_{\infty}$ für alle $x \in A$ in $\mathbb{R}$ konvergiert;*

LEMMA 1.9. *Gleichmäßige Konvergenz impliziert punktweise Konvergenz. Normkonvergenz impliziert absolute Konvergenz.*

Ist X ein Banachraum, so folgt aus der absoluten Konvergenz auch die punktweise Konvergenz; und aus der Normkonvergenz auch die gleichmäßige Konvergenz

BEWEIS. Dass die gleichmäßige Konvergenz auch die punktweise Konvergenz impliziert, ist ein allgemeines Phänomen, dass wir bereits für Folgen erwähnt haben.

Es folgt aus

$$\|f(x)\|_X \le \|f\|_{\infty} \qquad \forall x \in X$$

und dem Majorantenkriterium (für reell-wertige Reihen: somit braucht X kein Banachaum zu sein), dass Normkonvergenz auch die absolute Konvergenz impliziert.

Ist X ein Banachraum, so folgt dank dem Majorantenkriterium für Reihen, die mit Folge von Vektoren aus X assoziiert sind, aus der absoluten Konvergenz auch die punktweise Konvergenz.

Betrachtet man schließliche eine Folge beschränkter Funktionen, so kann man $\left(\sum_{k=0}^{n} f_k\right)_{n\in\mathbb{N}}$ als Folge von Vektoren in $B(A; X)$ anschauen. Somit liefert das Majorantenkriterium für Reihen (wieder im Fall, dass X ein Banachraum ist), dass Normkonvergenz auch die gleichmäßige Konvergenz impliziert:

denn die absolute Konvergenz (resp. die Konvergenz) einer Reihe im Banachraum $B(A, X)$ bedeutet gerade die Normkonvergenz (resp. die gleichmäßige Konvergenz) im Sinn der Definition 10. □

Anmerkung 11. Eine spezielle Klasse von Funktionenreihen kennen wir schon: Es sind die Potenzreihen, welche die obere Form haben, wenn alle f_n Monome sind, also

$$f_n : \mathbb{C} \ni z \mapsto a_n z^n \in \mathbb{C},$$

für ein $a_n \in \mathbb{N}$. Man betrachte eine Potenzreihe P und die beiden Aussagen

(i) P hat Konvergenzradius $R > 0$
(ii) P (als Reihe von Funktionen von $B_R(0)$ nach X aufgefasst) konvergiert punktweise.

Dann besagt [**6**, Satz 6.41] gerade, dass (i)⇒(ii).

Übungsaufgabe 12. Formuliere Cauchy-Bedingungen für die Konvergenz von Funktionenreihen.

Das folgende Konvergenzkriterium geht auf Karl Weierstraß zurück.

Satz 1.10. *Es seien X, Y Normierte Räume, A eine Teilmenge von Y und $(f_n)_{n\in\mathbb{N}}$ eine Folge beschränkter Funktionen von A nach X. Gibt es eine Folge $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ so, dass*

- *$\|f_n\|_\infty \leq \alpha_n$ für alle n*[3] *und*
- *die assoziierte Reihe $\sum_{n=0}^\infty \alpha_n$ konvergiert.*

Dann ist die Reihe

$$\sum_{k=0}^{\infty} f_k$$

normkonvergent (und insbesondere absolut und gleichmäßig konvergent).

Beweis. Diese Reihe ist mit der Folge $(f_n)_{n\in\mathbb{N}} \subset B(A; X)$ assoziiert. Dann folgt die absolute Konvergenz in $B(A; X)$ (in der Sprache der Definition 10: die Normkonvergenz) nach dem Majorantenkriterium für Reihen, s. [**6**, Satz 6.16], da $B(A; X)$ nach dem Korollar 1.2 ein Banachraum ist. Folglich liefert die Anmerkung 10 die behaupteten Konvergenzaussagen in absoluten und gleichmäßigen Sinne. □

Im folgenden wenden wir den Satz 1.6 an Potenzreihen an. Dabei wollen wir notationell nicht zwischen der formalen Potenzreihe P (die nicht zu Konvergieren braucht) und die assoziierte Funktion, deren Definitionsmenge sicherlich die offene Kugel mit Radius $R(P)$ enthält – wobei $R(P)$ der Konvergenzradius von P bezeichnet.

Beispiel 13. Es seien $A = Y = X = \mathbb{R}$ und definiere $(f_n)_{n\in\mathbb{N}}$ durch

$$f_n(x) := \frac{x}{x^4 + 3n^4}, \qquad n \in \mathbb{N},\ x \in \mathbb{R}.$$

Für alle $n \in \mathbb{N}$ gilt

$$f_n'(x) = \frac{3(n^4 - x^4)}{(x^4 + 3n^4)^2} \qquad \forall n \in \mathbb{N},\ \forall x \in \mathbb{N},$$

[3] Wir wissen, dass Konvergenz nicht vom Verhalten von endlich vielen Termen einer Folge abhängt: Somit könnten wir diese Bedingung schwächen und lediglich fordern, dass $\|f_n\|_\infty \leq \alpha_n$ für alle n größer als ein passender Index N.

und somit sind $+n$ und $-n$ die einzigen kritischen Punkte von f_n (übungsaufgabe). Man kann leicht überprüfen, dass $+n$ bzw. $-n$ tatsächlich das globale Maximum bzw. das globale Minimum der Funktion darstellen, und somit gilt

$$\|f_n\|_\infty = f_n(n) = -f_n(-n) = \frac{1}{4n^3}.$$

Dann kann man das obige Konvergenzkriterium von Weierstraß anwenden, denn die Reihe

$$\sum_{n=0}^{\infty} \frac{1}{4n^3}$$

konvergiert nach [**6**, Beispiel 6.18]. □

Beispiel 14. Es seien $X = Y = \mathbb{C}$, $A = B_1(0)$. Die Exponentialreihe

$$\sum_{k=0}^{\infty} \frac{z^k}{k!}$$

definiert eine auf A normkonvergente Funktionenreihe – der Grenzwert ist selbstverständlich die Exponentialfunktion. □

Die Aussage im obigen Beispiel kann angesichts des folgenden Resultats deutlich gestärkt werden.

Korollar 1.11. *Es sei*

$$P(z) := \sum_{k=0}^{\infty} a_k z^k$$

eine Potenzreihe mit Konvergenzradius $R(P) > 0$ und es sei $r \in (0, R(P))$. Dann ist die dadurch definierte Funktionenreihe auf $\{z \in \mathbb{C} : |z| \le r\}$ normkonvergent.

Beweis. Nach [**6**, Satz 6.41] ist $P(z)$ für alle $z \in \{w \in \mathbb{C} : |w| < R(P)\}$ absolut konvergent. Also insbesondere (als Funktionenreihe $\sum_{k=0}^{\infty} f_k : \{z \in \mathbb{C} : |z| \le r\} \to \mathbb{C}$ aufgefasst) gilt

$$\sum_{k=0}^{\infty} \|f_k\|_\infty = \sum_{k=0}^{\infty} |a_k| r^k < \infty.$$

Dies vervollständigt den Beweis. □

Beispiel 15. Es seien wieder $X = Y = \mathbb{C}$, $A = B_1(0)$. Dank [**6**, Satz 6.44] hat Exponentialreihe Konvergenzradius ∞, somit ist die damit assoziierte Funktionenreihe laut dem Korollar 1.11 auf $B_R(0)$ normkonvergent, für $R > 0$ beliebig groß. □

Es bietet sich nun an, die im § 1.1 erhaltenen Konvergenzsätze speziell an Funktionenreihen anzuwenden.

Satz 1.12. *Es seien X, Y Banachräume, $A \subset X$, und $(f_n)_{n\in\mathbb{N}}$ einer Folge Funktionen von A nach X, welche an einer Stelle $x_0 \in A$ stetig sind. Konvergiert die mit der Folge $(f_n)_{n\in\mathbb{N}}$ assoziierte Reihe gleichmäßig, etwa gegen F, so ist auch F an der Stelle x_0 stetig.*

Beweis. Es reicht, die Folge $(\phi_n)_{n\in\mathbb{N}}$ der Partialsummen der Reihe

$$F = \sum_{k=0}^{\infty} f_k$$

zu betrachten, und dann den Satz 1.1 anzuwenden. Das ist möglich, da jedes ϕ_n in x_0 stetig ist. □

Satz 1.13. *Es seien $A \subset \mathbb{R}^n$ kompakt und $(f_n)_{n\in\mathbb{N}}$ eine Folge stetiger Funktionen von A nach $[0,\infty)$. Konvergiert die assoziierte Funktionenreihe Punktweise gegen eine stetige Funktion f, so ist die Konvergenz bereits gleichmäßig.*

Beweis. Es reicht wieder, die Folge der Partialsummen zu betrachten. Da die Funktionen positive Werte annehmen, ist die Folge der Partialsummen monoton wachsend. Die Aussage folgt nun als Korollar vom Satz von Dini. □

Satz 1.14. *Es seien $a, b \in \mathbb{R}$, $a < b$, X ein Banachraum und $(f_n)_{n\in\mathbb{N}}$ einer Folge differenzierbarer*[4] *Funktionen von $[a,b]$ nach X. Gibt es ein $x_0 \in [a,b]$ so, dass die mit der Folge $(f_n(x_0))_{n\in\mathbb{N}} \subset \mathbb{R}$ assoziierte Reihe konvergiert, und konvergiert auch die mit der Folge $(f_n')_{n\in\mathbb{N}}$ assoziierte Reihe gleichmäßig, etwa gegen F, so konvergiert die mit der Folge $(f_n)_{n\in\mathbb{N}}$ assoziierte Reihe gleichmäßig, etwa gegen F. Der Grenzwert F ist auf $[a,b]$ differenzierbar und*

$$F'(x) = \sum_{k=1}^{\infty} f_k'(x).$$

Besteht die Folge aus stetig *differenzierbaren Funktionen, so ist auch F* stetig *differenzierbar.*

Beweis. Betrachte die Folge $(\phi_n)_{n\in\mathbb{N}}$ der Partialsummen der Reihe

$$F = \sum_{k=0}^{\infty} f_k$$

und wende den Satz 1.6 an. Das ist möglich, da jedes ϕ_n auf $[a,b]$ differenzierbar ist. □

Eine verlockend anschauliche Möglichkeit, um den obigen Satz auszudrucken, ist, dass

$$\left(\sum_{k=1}^{\infty} f_k(x)\right)' = \sum_{k=1}^{\infty} f_k'(x), \qquad x \in [a,b],$$

gilt. Man kann also die Reihe *summandenweise ableiten.*

Insbesondere kann man den obigen Satz zum Spezialfall von Potenzenreihen

$$P : z \mapsto \sum_{n=0}^{\infty} a_n z^n,$$

anwenden: denn Potenzenreihen sind mit Folgen von Funktionen assoziiert, welche sicherlich unendlich oft stetig differenzierbar sind. Dann kann man hoffentlich summandenweise ableiten und somit die neue, *abgeleitete* Potenzreihe

$$P' : z \mapsto \sum_{n=0}^{\infty} n a_n z^{n-1},$$

betrachten.

Korollar 1.15. *Eine Potenzreihe und ihre abgeleitete Potenzreihe haben übereinstimmende Konvergenzradien.*

[4] Aber nicht notwendigerweise *stetig* differenzierbar!

Beweis. Man beachte, dass der Konvergenzradius der Reihen

$$\sum_{n=0}^{\infty} n a_n z^{n-1} \qquad \text{und} \qquad \sum_{n=0}^{\infty} n a_n z^n = z \sum_{n=0}^{\infty} n a_n z^{n-1}$$

übereinstimmen. Somit ist der Satz bewiesen, wenn wir die Aussage für die ursprügliche Reihe und für diese modifizierte abgeleitete Reihen beweisen können. Wollen wir die Cauchy-Formel für die Bestimmung des Konvergenzradius ([**6**, Satz 6.42]) anwenden, so merken wir, dass

$$\limsup_{n\to\infty} \sqrt[n]{|a_n|} = \lim_{n\to\infty} \sqrt[n]{n} \limsup_{n\to\infty} \sqrt[n]{|a_n|} = \limsup_{n\to\infty} \sqrt[n]{n|a_n|},$$

da $\lim_{n\to\infty} \sqrt[n]{n} = 1$ (übungsaufgabe!). □

Anmerkung 16. Es ist klar, dass genau dann eine Potenzreihe

$$P(z) := \sum_{n=0}^{\infty} a_n z^n,$$

konvergiert, wenn die verschobene Potenzreihe

$$\tilde{P}(z) := \sum_{n=0}^{\infty} a_n (z - z_0)^n,$$

konvergiert. Allgemeiner gilt: Die Potenzreihe um ein $z_0 \neq 0$ konvergiert in $B_{R(P)}(z_0) = \{z \in \mathbb{C} : |z - z_0| < R(P)\}$ und divergiert in $\{z \in \mathbb{C} : |z - z_0| > R(P)\}$, wobei $R(P)$ der Konvergenzradius der ursprünglichen Reihe P ist. Diese einfache Beobachtung wird im folgenden Abschnitt oft gebraucht.

1.3. Approximierbarkeit und Taylorentwicklung

Im ersten Teilabschnitt haben wir uns mit der Frage befasst, im welchen Sinn eine Funktionenfolge konvergieren kann, und welche Eigenschaften die Grenzfunktion von den Folgenglieder erben kann. Hier wollen wir uns mit dem etwas dualen Thema beschäftigen, inwiefern eine gegebene stetige Funktion sich als Grenzwert eine passende Folge (z.B. einer Reihe) darstellen lässt. Die Bedeutung dieser Aufgabe wird gleich klar, wenn man bedenkt, das eine ganz allgemeine Funktion – salopp gesagt – unendlich viel Information enthält, genauer gesagt: die Beschreibung einer Funktion $f : \mathbb{R} \to \mathbb{R}$ benötigt im Allgemeinen überabzählbare bits. Ist jedoch eine Funktion z.B. als Potenzreihe darstellbar, dann zählen für ihre Beschreibung ausschließlich ihre (abzählbar viele) Koeffizienten. Noch besser wäre, wenn wir anstatt der ganzen Potenzreihe nur eine Partialsumme – d.h., ein Polynom – betrachten könnten und dabei gut wüssten, welchen Fehler dabei maximal entsteht.

Satz 1.16. *Es seien $a, b \in \mathbb{R}$, $a < b$, X ein normierter Vektorraum und $f : [a, b] \to X$ eine n-mal stetig differenzierbare Funktion, wobei $n \in \mathbb{N}$.*

(i) Dann gibt es zu jedem $x_0 \in [a, b]$ eine Restfunktion $R_n \in C([a, b]; X)$ *(die i.A. von f und x_0 abhängt), so dass*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + R_n^{f,x_0}(x), \qquad x \in (a, b), \tag{1.13}$$

und welche die Abschätzung

$$\|R_n^{f,x_0}(x)\|_X \leq \frac{1}{(n-1)!} \sup_{t\in(0,1)} \|f^{(n)}(x_0+t(x-x_0)) - f^{(n)}(x_0)\|_X |x-x_0|^n, \qquad x \in (a,b), \tag{1.14}$$

erfüllt.

(ii) Ist insbesondere die Funktion sogar $(n+1)$-Mal differenzierbar mit beschränkter $(n+1)$-ter Ableitung , so gibt es ein $L > 0$, für das die Abschätzung

$$\|R_n^{f,x_0}(x)\|_X \leq \frac{L}{(n-1)!}|x-x_0|^{n+1}, \qquad x \in (a,b), \tag{1.15}$$

gilt.

Anders gesagt lässt sich die Funktion f durch ein Polynom approximieren, das alle seine Nullstellen in einem gewünschten Punkt x_0 hat. Je regulärer die Funktion f und je näher das Argument der Funktion an x_0 ist, desto genauer diese Approximation ist. Dabei folgt aber nicht, etwa aus der Abschätzung

$$\left\| f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k \right\|_X \leq \frac{1}{(n-1)!}|x-x_0|^{n+1}, \qquad x \in (a,b),$$

(vgl. (1.15)), dass T_n^{f,x_0} in einer Umgebung von x_0 gegen f konvergiert, nicht mal punktweise. Vielmehr gilt eine solche punktweise Konvergenz genau dann, wenn R_n^{f,x_0} punktweise gegen 0 konvergiert.

Wir wollen auch ausdrucklich erwähnen, dass die Voraussetzun von (b) insbesondere erfüllt ist, falls f auf $[a,b]$ $(n+1)$-Mal stetig differenzierbar. Es sei also $f \in C^\infty\big([a,b];X\big)$ [5]. Dann kann man für jedes n den Wert $f(x)$ durch das *Taylor-Polynom vom Grad n an der Stelle x_0*

$$T_n^{f,x_0}(x) := \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k, \qquad x \in (a,b),$$

abschätzen: man spricht dabei von *Taylor-Formel*.

Beweis. (i) Es sei

$$R_n^{f,x_0}(x) := f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k, \qquad x \in (a,b),$$

und setze für ein festes $x \in (a,b)$

$$h(t) := f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0+t(x-x_0))}{k!}(1-t)^k(x-x_0)^k - \frac{f^{(n)}(x_0)}{n!}(1-t)^n(x-x_0)^n, \qquad t \in [0,1].$$

Diese Funktion $h : (0,1) \to X$ ist gerade so gewählt, dass

$$h(0) = R_n^{f,x_0}(x) \qquad \text{und} \qquad h(1) = 0$$

(da die Summe an der rechten Seite für $k = 0$ nicht verschwindet). Darüber hinaus ist h differenzierbar, da für das obige (festes!) x

$$h'(t) = \Big(f^{(n)}(x_0) - f^{(n)}(x_0+t(x-x_0))\Big)\frac{(1-t)^{n-1}}{(n-1)!}(x-x_0)^n, \qquad t \in (0,1),$$

[5] D.h. (zur Erinnerung): $f \in C^n\big([a,b];X\big)$ für alle $n \in \mathbb{N}$.

(Dank der Kettenregel wird $g'(t)$ zu einer teleskopischen Summe!). Nach der vektorwertigen Version des Satzes von Lagrange ([**6**, Satz 8.39]) gilt

$$\begin{aligned}\|R_n^{f,x_0}(x)\|_X &= \|h(1)-h(0)\|_X \\ &\leq \sup_{\xi\in(0,1)} \|h'(\xi)\|_X \\ &= \sup_{\xi\in(0,1)} \left\| f^{(n)}(x_0) - f^{(n)}(x_0+\xi(x-x_0))\right\|_X \frac{|x-x_0|^n}{(n-1)!}, \qquad x\in(a,b).\end{aligned}$$

(ii) Es sei nun f zusätzlich $(n+1)$-Mal differenzierbar mit beschränkter $(n+1)$-ter Ableitung. Dann kann mann [**6**, ÜA 8.21] an $f^{(n)}$ anwenden und somit bekommen, dass $f^{(n)}$ Lipschitz-stetig ist, d.h., es gibt $L>0$ mit

$$\|f^{(n)}(x)-f^{(n)}(y)\|_X \leq L|x-y|, \qquad \text{für alle } x,y\in[a,b].$$

Wendet man diese Abschätzung an (1.14) an, so erhält man

$$\|R_n^{f,x_0}(x)\| \leq \frac{1}{(n-1)!}\sup_{t\in(0,1)} \left(Lt|x-x_0|\right)|x-x_0|^n = \frac{L}{(n-1)!}|x-x_0|^{n+1}, \qquad x\in(a,b),$$

wie man zeigen wollte. □

Im Folgenden verwenden wir die folgende Verallgemeinerung der Notation, die in [**6**, Definition 4.65] eingeführt wurde:

Es sei f eine Funktion von (a,b) nach X und $\alpha\in\mathbb{R}$. Dann sagt man, dass f *eine Nullstelle von Ordnung* α *an der Stelle* x_0 hat und schreibt $f=o(|x-x_0|^\alpha)$, wenn

$$\lim_{x\to x_0}\frac{f(x)}{|x-x_0|^\alpha}=0$$

[6]

Anmerkung 17. Für die Anwendungen ist eine Abschätzung der Art von (1.13) wesentlich, denn sie liefert eine *quantitative* Aussage über den Fehler, den wir machen, wenn wir den Wert $f(x)$ mit $T_n^{f,x_0}(x)$ annähern. Man beachte, dass nur für $x\in[x_0-1,x_0+1]$ wird der Term $|x-x_0|^n$ beschränkt, und somit $R_n^{f,x_0}(x)$ klein.

Unter den Voraussetzungen des Satzes 1.16 gilt im Allgemeinen nur

$$f(x)=\sum_{k=0}^n\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k+o(|x-x_0|^n) \text{ für } x\to x_0 \qquad \text{für alle } x\in(a,b).$$

Beispiel 18. Für viele Zwecke lässt sich also eine Funktion durch Polynome niedriger Ordnung approximieren. So beruhen z.B. die meisten Herleitungen in der Physik des Pendels darauf, dass

$$\sin x\simeq x \quad \textit{für } x \textit{ klein}$$

gilt. Diese oft angewandte Approximations entspricht der Tatsache, dass x der erste Term der Taylorreihe von $\sin x$ um 0 ist. Eine genauere Approximation wäre

$$\sin x\simeq x+\frac{x^3}{6} \quad \textit{für } x \textit{ klein.}$$

[6] Wir fügen hinzu: Auch sagt man, dass f *beschränkt von Ordnung* α *an der Stelle* x_0 ist und schreibt $f=O(|x-x_0|^\alpha)$, wenn die Funktion $\frac{f}{|\cdot-x_0|^\alpha}$ in einer Umgebung von x_0 beschränkt ist.

ähnlich approximiert man gelegentlich

$$\cos x \simeq \frac{x^2}{2} \quad \textit{für } x \textit{ klein.}$$

□

Beispiel 19. Wir haben in [**6**] die Logarithmusfunktion einfach als Umkehrfunkion der Exponentialfunktion eingeführt. Will man aber die Logarithmusfunktion auswerten, so bietet sich an, die Taylorformel von $\log x$ zu verwenden – zunächst für $x \in (0, 2)$.

Betrachte dazu die Funktion f, die durch

$$f(x) := \log(1 + x)$$

definiert wird und die nach der Kettenregel auf $(-1, 1)$ uendelich oftstetig differenzierbar ist. Dann gilt

$$f'(x) = \frac{1}{1+x}, \quad f''(x) = -\frac{1}{(1+x)^2}, \quad \text{und induktiv} \quad f^{(n)}(x) = (-1)^{n-1}\frac{(n-1)!}{(1+x)^n} \quad \text{für } n = 1, 2, 3, \ldots.$$

Insbesondere

$$f^{(n)}(0) = (-1)^{n-1}(n-1)!, \quad n = 1, 2, 3, \ldots.$$

Durch Einsetzen in (1.13) für $x_0 = 0$ (das entspricht der Taylorschen Entwicklung von log an der Stelle 1) erhält man für jedes $N \in \mathbb{N}$

$$f(x) = \sum_{n=0}^{N}(-1)^{n-1}(n-1)!\frac{x^n}{n!} + R_N(x), \qquad x \in (-1, 1).$$

d.h.,

$$\log(1 + x) = \sum_{n=0}^{N}(-1)^{n-1}\frac{x^n}{n} + R_N(x), \qquad x \in (-1, 1).$$

Somit ist für alle $N \in \mathbb{N}$

$$\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \ldots + (-1)^{N-1}\frac{x^N}{N} + o(|x|^N), \qquad x \in (0, 2).$$

□

Korollar 1.17. *Es seien $a, b \in \mathbb{R}$, $a < b$, und $f : [a, b] \to \mathbb{R}$ auf (a, b) n-Mal differenzierbar. Gilt für ein $x_0 \in (a, b)$*

$$f'(x_0) = f''(x_0) = \ldots = f^{(n-1)}(x_0) = 0 \quad \textit{und} \quad f^{(n)}(x_0) \neq 0,$$

so gelten die folgenden Aussagen.

(a) Ist n ungerade, so hat f weder ein lokales Minimum noch ein lokales Maximum an der Stelle x_0.
(b) Ist n gerade, so hat f an der Stelle x_0
- *ein lokales Minimum, falls $f^{(n)}(x_0) > 0$, oder*
- *ein lokales Maximum, falls $f^{(n)}(x_0) < 0$.*

Beweis. Nach dem Satz 1.16 gilt

$$f(x) - f(x_0) = \left(\frac{f^{(n)}(x_0)}{n!} + \frac{o(|x - x_0|^n)}{(x - x_0)^n}\right)(x - x_0)^n \quad \text{für } x \to x_0. \tag{1.16}$$

Da nach Definition des Landau-Symbols o

$$\lim_{x\to x_0} \frac{o(|x-x_0|^n)}{(x-x_0)^n} = 0$$

kann man insbesondere zur Konstante

$$\varepsilon := \frac{|f^{(n)}(a)|}{(2n)!}$$

ein $\delta > 0$ so klein finden, dass

$$\frac{|o(|x-x_0|)^n|}{|x-x_0|^n} \leq \varepsilon \qquad \text{für alle } x \in (a,b) \cap (x_0-\delta, x_0+\delta),$$

d.h. falls n ungerade ist

$$\frac{f^{(n)}(x_0)}{n!} + \frac{o(|x-x_0|^n)}{(x-x_0)^n} \geq \varepsilon \qquad \text{für alle } x \in (a,b) \cap (x_0-\delta, x_0+\delta) \tag{1.17}$$

sowie auch, falls n gerade ist und $f^{(n)}(x_0) > 0$, dass

$$\frac{f^{(n)}(x_0)}{n!} + \frac{o(|x-x_0|^n)}{(x-x_0)^n} \geq \varepsilon \qquad \text{für alle } x \in (a,b) \cap (x_0-\delta, x_0+\delta) \tag{1.18}$$

und, falls n gerade ist und $f^{(n)}(x_0) < 0$, dass

$$\frac{f^{(n)}(x_0)}{n!} + \frac{o(|x-x_0|^n)}{(x-x_0)^n} \leq -\varepsilon \qquad \text{für alle } x \in (a,b) \cap (x_0-\delta, x_0+\delta) \tag{1.19}$$

(Beachte, dass das Symbole $o(|x-x_0|^n)$ nur Informationen über eine Größe in Betrag liefert, jedoch *nicht* über ihr Vorzeichen!)

Jetzt kann die Fallunterscheidung anfangen.

- Es sei n ungerade. Es sei zunächst $f^{(n)}(a) > 0$. Dann folgt aus (1.16) und (1.17), dass

  $$f(x) - f(x_0) \geq \varepsilon (x-x_0)^n \qquad \text{für alle } x \in (a,b) \cap (x_0, x_0+\delta),$$

  sowie auch, dass

  $$f(x) - f(x_0) \leq \varepsilon (x-x_0)^n \qquad \text{für alle } x \in (a,b) \cap (x_0-\delta, x_0).$$

  Somit ist x_0 weder eine Maximum- noch eine Minimumstelle. ähnlich beweist man die Aussage, falls stattdessen $f^{(n)}(a) > 0$ gilt.
- Es sei n gerade und es gelte $f^{(n)}(a) > 0$. Dann folgt aus (1.16) und (1.18), dass

  $$f(x) - f(x_0) \geq \varepsilon (x-x_0)^n \qquad \text{für alle } x \in (a,b) \cap (x_0-\delta, x_0+\delta),$$

  und somit ist x_0 eine Minimumstelle.
- Entsprechend folgt aus (1.16) und (1.19), dass

  $$f(x) - f(x_0) \leq -\varepsilon (x-x_0)^n \qquad \text{für alle } x \in (a,b) \cap (x_0-\delta, x_0+\delta),$$

  falls n gerade ist und $f^{(n)}(a) < 0$ gilt, und somit ist x_0 eine Maximumstelle.

Alle Fälle wurden betrachtet und somit ist der Beweis vervollständigt. $\square$

Man kann formal die Potenzreihe betrachten, die durch

$$T^{f,x_0} : x \mapsto \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k,$$

definiert ist: sie heißt die *Taylor-Reihe von f an der Stelle* x_0. Diese Reihe braucht an beliebige x nicht zu konvergieren, vielmehr soll man ihren Konvergenzradius $R(T^{f,x_0})$ bestimmen.

Innerhalb der Kugel mit Zentrum x_0 und Radius $R(T^{f,x_0})$ definiert definitionsgemäß T^{f,x_0} eine Funktion, die im Allgemeinen nicht mit f übereinstimmen muss (obwohl $(T_n^{f,x_0})_{n\in\mathbb{N}}$ tatsächlich f approximiert).

Beispiel 20. Es sei $f : \mathbb{R} \to \mathbb{R}$ die Funktion, die durch

$$f(x) := \begin{cases} e^{-\frac{1}{x^2}}, & \text{falls } x \neq 0, \\ 0, & \text{sonst,} \end{cases}$$

definiert ist.

Wir zeigen durch Induktion über n, dass $f \in C^\infty(\mathbb{R})$ ist, und deren Ableitungen lassen sich durch

$$f^{(n)}(x) := \begin{cases} p_n(x)e^{-\frac{1}{x^2}}, & \text{falls } x \neq 0, \\ 0, & \text{sonst,} \end{cases}$$

darstellen lassen, wobei p_n passende Polynome sind.

Für $n = 0$ ist die Aussage klar: nimm einfach das Polynom vom Grad 0. Es sei nun die Aussage für n richtig, so unterscheidet man die Fälle $x \neq 0$ und $x = 0$. Für $x \neq 0$ gilt nämlich

$$\begin{aligned} f^{(n+1)}(x) &= \frac{d}{dx} f^{(n)}(x) \\ &= \frac{d}{dx}\left(p_n\left(\frac{1}{x}\right)e^{-\frac{1}{x^2}}\right) \\ &= \left(-p_n'\left(\frac{1}{x}\right)\frac{1}{x^2}) + 2p_n\left(\frac{1}{x}\right)\frac{1}{x^3}\right)e^{-\frac{1}{x^2}} \\ &=: p_{n+1}\left(\frac{1}{x}\right)e^{-\frac{1}{x^2}}. \end{aligned}$$

Andererseits gilt auch für $x = 0$

$$\begin{aligned} f^{(n+1)}(0) &= \lim_{x\to 0} \frac{f^{(n)}(x) - f^{(n)}(0)}{x} \\ &= \lim_{x\to 0} \frac{p_n\left(\frac{1}{x}\right)e^{-\frac{1}{x^2}}}{x}. \end{aligned}$$

Nun kann dieser Grenzwert bestimmt mithilfe von [**6**, üA 7.12], welche

$$\lim_{x\to 0} \frac{p_n\left(\frac{1}{x}\right)e^{-\frac{1}{x^2}}}{x} = \lim_{R\to\pm\infty} Rp_n(R)e^{-R^2} = 0.$$

liefert. Somit ist bewiesen, dass $f^{(n)}(0) = 0$ für alle $n \in \mathbb{N}$, also nimmt die Taylor-Reihe von f an der Stelle 0 Wert

$$T^{f,0}(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k \equiv 0 \qquad \text{für alle } x \in \mathbb{R},$$

obwohl $f \not\equiv 0$. □

Beispiel 21. Betrachte eine Kanonenkugel, die mit fester Anfangsgeschwindigkeit v und Abschusswinkel α geschossen wird. Berücksichtigt man den Luftwiderstand nicht, so liefern die klassischen Gesetze der Newtonschen Mechanik unmittelbar die Flugbahn der Kugel in Abhängigkeit von der Zeit ab dem Schuss. Dann liegt die Flugbahn in einer Ebene, die wir in kartesischen Koordinaten mit x bzw. y (horizontaler bzw. vertikaler Abstand vom Schusspunkt) beschreiben. Dann sind die dementsprechenden Gleichungen

$$\begin{aligned} x(t) &= vt\cos(\alpha) \\ y(t) &= -\frac{gt^2}{2} + vt\sin(\alpha) \end{aligned}$$

solang $y(t) \geq 0$, wobei g die Fallbeschleunigung bezeichnet. Diese Gleichungen gelten auf jedem Planet, sobald man die spezifische Fallbeschleunigung g kennt ($g = 9,81 \, \frac{m}{s^2}$ auf der Erde).

Hat übrigens die Funktion y ein Maximum an einem Zeitpunkt t_0? Wir wissen schon, dass eine notwendige Bedingung hierfür ist, dass $y'(t_0) = 0$, also dass

$$gt_0 = v\sin(\alpha) \qquad \text{und somit} \qquad t_0 = \frac{v\sin(\alpha)}{g}.$$

Es ist $y''(t_0) = -g < 0$, also ist t_0 eine Maximumstelle: dann erreicht *für festen Anfangsgeschwindigkeit und Abschusswinkel* die Kugel ihr Maximum

$$y(t_0) = \frac{v^2\sin(\alpha)^2}{2g}.$$

□

Satz 1.18. *Ist P eine Potenzreihe*

$$P : x \mapsto \sum_{k=0}^{\infty} a_k (x - x_0)^k \tag{1.20}$$

mit Konvergenzradius $R(P) > 0$, so ist $P \in C^\infty\big((-R(P), R(P)); X\big)$ und P stimmt mit seiner eigener Taylor-Reihe an jeder Stelle $x \in (-R(P), R(P))$ überein [7]*. Innsbesondere stimmt das k-te Koeffizient a_k von P mit*

$$\frac{P^{(k)}(0)}{k!}$$

überein.

[7] D.h., $(T_n^{f,x_0})_{n\in\mathbb{N}}$ konvergiert auf $(-R(P), R(P))$ punktweise gegen die mit P assoziierte Funktion f.

Beweis. Es sei $k \in \mathbb{N}$. Man kann rekursiv aus dem Satz 1.15 folgern, dass $P \in C^k\big((x_0 - R(P), x_0 + R(P)\big)$ und zusätzlich

$$\begin{aligned} P^{(k)}(x) &= \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)a_n(x-x_0)^{n-k} \\ &= k!\, a_k + \sum_{h=1}^{\infty}(h-k)(h-k-1)\cdots 2a_{h-k}(x-x_0)^h, \end{aligned}$$

und insbesondere (durch Einsetzen von $x = x_0$)

$$P^{(k)}(0) = k!\, a_k.$$

Somit

$$\frac{P^{(k)}(0)}{k!} = a_k,$$

und deshalb stimmt P notwendigerweise mit seiner Potenzreihe.

Dies zeigt die Aussage. □

Wir haben im Beispiel 20 gesehen, dass eine Funktion nicht notwendigerweise mit ihrer Taylor-Reihe übereinstimt – selbst wenn sie C^∞ ist. In der Tat kann man eine Klasse von Funktionen einführen, die noch regulärer ist und diese Lücke schließt.

Definition 22. *Es seien $a, b \in \mathbb{R}$, $a < b$, X ein normierter Vektorraum und $f \in C^\infty\big((a,b); X\big)$. Dann heißt f* **reell-analytisch**, *falls es für jedes $x_0 \in (a,b)$ ein R_0 gibt mit $B_{R_0}(x) \subset (a,b)$ und eine Potenzreihe*

$$P : x \mapsto \sum_{k=0}^{\infty} a_k(x-x_0),$$

so dass

$$f(x) = P(x) \qquad \text{für alle } x \in B_{R_0}(x).$$

Korollar 1.19. *Es seien $a, b \in \mathbb{R}$, $a < b$, X ein normierter Vektorraum. Dann ist $f : (a,b) \to X$ genau dann reell-analytisch, wenn $f \in C^\infty\big((a,b); X\big)$ und zu jedem $x_0 \in (a,b)$ gibt es ein $\rho > 0$, so dass*

- *$(x_0 - \rho, x_0 + \rho) \subset (a,b)$,*
- *die Taylor-Reihe an der Stelle x_0 konvergiert punktweise und*
- *$f(x) = T^{f,x_0}(x)$ für alle $x \in (x_0 - \rho, x_0 + \rho)$.*

Beweis. Die Aussage folgt direkt vom Satz 1.18. □

Anmerkung 23. Unter den Voraussetungen des Satzes 1.16, ist f auf (a,b) $(n+1)$-mal differenzierbar und ist $X = \mathbb{R}$, so gibt es zu jedem $x \in (a,b)$ ein $\xi \in (a,b)$, so dass die Restfunktion R_n in (1.13) die Form

$$R_n^{f,x_0}(x) = \frac{f^{(n+1)}(\xi)}{(n-1)!}(x-x_0)^{n+1}, \qquad x \in (a,b). \tag{1.21}$$

nimmt: Denn man findet

$$\begin{aligned} R_n^{f,x_0}(x) &= h(1) - h(0) \\ &\overset{!}{=} h'(\zeta) \\ &= \Big(f^{(n)}(x_0) - f^{(n)}(x_0 + \xi(x - x_0))\Big) \frac{(x - x_0)^n}{(n-1)!}, \qquad x \in (a, b), \\ &\overset{!}{=} f^{(n+1)}(\eta) \frac{(x - x_0)^{n+1}}{(n-1)!}, \end{aligned}$$

wobei ζ, η zwei passende Elemente von (a, b) sind (beide mit ! gekennzeichneten Ausdrucke folgen aus dem Satz von Lagrange).

Wir werden aber gleich sehen, dass die Abschätzung (1.21) sich ein wenig verbessern lässt, wenn man den Satz von Cauchy (und nicht den Satz von Lagrange!) im Beweis verwendet – der leider im vektorwertigen Fall nicht zur Verfügung steht. Diese Verbesserung ist aber für den Beweis des Satzes 1.21 wesentlich.

Satz 1.20. *Es seien $a, b \in \mathbb{R}$, $a < b$, $n \in \mathbb{N}$ und $f : [a, b] \to \mathbb{R}$ eine n-mal stetig differenzierbare Funktion, welche zudem auf (a, b) $(n + 1)$-mal differenzierbar ist. Es sei $x_0 \in (a, b)$. Dann gelten die folgenden Aussagen.*

(a) Zu jedem $x \in (a, b) \setminus \{x_0\}$ gibt es ein $\xi \in (a, b)$ (genauer: $\xi \in (\min\{x, x_0\}, \max\{x, x_0\})$), so dass die Restfunktion R_n in (1.13) die Form

$$R_n^{f,x_0}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}, \qquad x \in (a, b).$$

nimmt.

(b) Ist zudem die $(n + 1)$-te Ableitung beschränkt, so gilt die Abschätzung

$$|f(x) - T_n^{f,x_0}(x)| \le \sup_{\xi \in (a,b)} |f^{(n+1)}(\xi)| \frac{|x - x_0|^{n+1}}{(n+1)!}, \qquad x \in (a, b).$$

Beachte den Term $\frac{1}{(n+1)!}$, im Gegensatz zu $\frac{1}{(n-1)!}$ in (1.21).

Beweis. (a) Betrachte für festes x die Funktionen g, h, die durch

$$g(t) := \sum_{k=0}^{n} f^{(k+1)}(t) \frac{(x - t)^k}{k!}$$

bzw.

$$h(t) := -(x - t)^{n+1}$$

definiert sind. Sie sind so gewählt worden, dass

$$g(x_0) = f(x), \quad g(x) = T_n^{f,x_0}(x), \qquad \text{und somit} \qquad g(x) - g(x_0) = R_n^{f,x_0}(x), \tag{1.22}$$

sowie

$$h(x) - h(x_0) = -(x - x_0)^{n+1} \tag{1.23}$$

Dann sind g, h auf $[\min\{x, x_0\}, \max\{x, x_0\}]$ stetig und auf $(\min\{x, x_0\}, \max\{x, x_0\})$ differenzierbar mit

$$g'(t) = f^{(n+1)}(t) \frac{(x - t)^n}{n!}$$

(Dank der Kettenregel wird $g'(t)$ zu einer teleskopischen Summe!) bzw.

$$h'(t) := -(n+1)(x-t)^n.$$

Nach dem Satz von Cauchy gibt es somit $\xi \in (\min\{x, x_0\}, \max\{x, x_0\})$ mit

$$g(x) - g(x_0) = \frac{g'(\xi)}{h'(\xi)}(h(x) - h(x_0)).$$

definiert werden. Durch Einsetzen von (1.22) folgt die Aussage sofort.

(b) Die Abschätzung ist eine unmittelbare Folgerung des obigen Resultats und der Definition von R_n^{f,x_0}. □

Übungsaufgabe 24. Es seien $a, b \in \mathbb{R}$, $a < b$, und $f : [a, b] \to \mathbb{R}$ eine n-mal stetig differenzierbare Funktion, die zusätzlich auf (a, b) $n+1$-mal differenzierbar ist. Es sei $x_0 \in [a, b]$. Zeige: Es gibt für alle $p > 0$ und alle $x \in [a, x_0) \cup (x_0, b]$ ein $\xi \in (\min\{x_0, x\}, \max\{x_0, x\})$, so dass die Restfunktion R_n aus dem Satz 1.16 durch

$$R_n^{f,x_0}(x) = \frac{f^{(n+1)}(\xi)}{pn!}\left(\frac{x-\xi}{x-x_0}\right)^{n-p+1}(x-x_0)^{n+1}.$$

gegeben ist. *(Hinweis: Betrachte Funktionen g, h, welche durch*

$$g(s) := \sum_{k=0}^{n} \frac{f^{(k)}(s)}{k!}(x-s)^k, \qquad h(s) := (x-s)^p, \qquad s \in [a, b],$$

gegeben sind und wende daran den Satz von Cauchy an).

Beispiel 25. Selbst für Funktionen, die als Potenzreihen eingeführt wurden, liefert die Taylorsche Theorie interessante Informationen. Betrachte etwa die Cosinusfunktion $\cos : \mathbb{R} \to \mathbb{R}$,

$$\cos x := \sum_{k=0}^{\infty}(-1)^k \frac{x^{2k}}{(2k)!}, \qquad x \in \mathbb{R}.$$

Möchte man wissen, welchen Fehler man macht, wenn man die Potenzreihe etwa bei $(N+1)$-ten Term abschneidet, d.h. wenn man die Approximation

$$\cos x \simeq \sum_{k=0}^{2N+2}(-1)^k \frac{x^{2k}}{(2k)!}$$

betrachet, so erhält man Dank dem Satz 1.20.(b) für $x_0 = 0$ und angesichts von

$$|\cos^{2k}(x)| = |\cos x|, \qquad k \in \mathbb{N},\ x \in \mathbb{R},$$

dass

$$|f(x) - T_{2N+2}^{f,0}(x)| \leq \sup_{\xi \in (-\pi,\pi)} |\cos \xi| \frac{|x|^{2N+2}}{(2N+2)!} \leq \frac{\pi^{2N+2}}{(2N+2)!}, \qquad x \in [-\pi, \pi].$$

Schon für $N = 10$ erhält man damit einen Fehler von nur ca. 10^{-12}. □

Der Satz 1.20 sagt somit, dass jede Potenzreihe mit strikt positivem Konvergenzradius eine reell analytische Funktion definiert. Ein weiteres nützliches Kriterium für die reelle Analytizität ergibts sich im Folgenden.

Satz 1.21. *Es seien $a, b \in \mathbb{R}$, $a < b$ und $f : (a,b) \to \mathbb{R}$ eine unendlich oft differenzierbare Funktion. Gibt es für alle $n \in \mathbb{N}$ Zahlen $M, L \geq 0$ so, dass*

$$\|f^{(n)}\|_\infty \leq ML^n,$$

so ist die Taylor-Reihe $T^{f,\cdot}$ (als Reihe von Funktionen von (a,b) nach $\mathbb{R}$ aufgefasst) gegen f normkonvergent.

Beweis. Wie wir schon aus dem Satz 1.20 wissen, gibt es zu jedem $n \in \mathbb{N}$ ein $R_n \in C((a,b);\mathbb{R})$ so, dass

$$\left| f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k \right| = R_n^{f,x_0}(x), \qquad x \in (a,b), \tag{1.24}$$

und welche für ein $\xi \in (\min\{x_0, x_1\}, \max\{x_0, x_1\})$ die Abschätzung

$$|R_n^{f,x_0}(x)| \leq \frac{1}{(n+1)!} f^{(n+1)}(\xi)|x-x_0|^n, \qquad x \in (a,b),$$

erfüllt. Die Aussage ist dann zur Konvergenz von R_n^{f,x_0} gegen 0 äquivalent. Dank der Voraussetzung gilt somit

$$|R_n^{f,x_0}(x)| \leq \frac{ML^{n+1}}{(n+1)!}|x-x_0|^{n+1} \leq M\frac{(L|b-a|)^{n+1}}{(n+1)!}, \qquad x \in (a,b),$$

und der rechtere Term definiert eine Funktionenfolge. Dank dem Majorantenkriterium ist der linkere Term konvergent: Dies zeigen wir dadurch, dass wir die deutlich stärkere Aussage beweisen, die Reihe

$$\sum_{n=0}^{\infty} \frac{(L|b-a|)^{n+1}}{(n+1)!}$$

sei konvergent – das kann z.B. mittels des Quotientenkriteriums überprüft werden. □

Übungsaufgabe 26. Formuliere ein ähnliches Kriterien für die absolute, und somit punktweise Konvergenz der Taylor-Reihe.

Übungsaufgabe 27. Betrachte die Taylor-Entwicklung im Beispiel 19 und zeige, dass für

$$\lim_{n\to\infty} T_n^{f,1}(x) = \log(1+x) \qquad \text{für alle } |x| < 1.$$

KAPITEL 2

Integralrechnung

DEFINITION 28. *Es seien $a, b \in \mathbb{R}$ mit $a < b$. Eine endliche Punktenmenge $\zeta := \{x_0, \ldots, x_n\}$ heißt* Zerlegung *von (a, b), falls $a = x_0$, $x_n = b$ und $x_k < x_{k+1}$ für alle $k = 0, \ldots, n + 1$.*

ANMERKUNG 29. Die Punktenmenge $\{a, b\}$ ist selbstverständlich eine (triviale) Zerlegung des Intervalls (a, b). Es seien ζ, ζ' zwei Zerlegungen eines Intervalls. Dann ist auch $\zeta \cup \zeta'$ eine Zerlegung desselben Intervalls. Allgemeiner gilt: Durch Zufügen von endlich vielen Punkten bei einer gegebenen Zerlegung ζ bekommt man eine neue Zerlegung, welche eine *Verfeinerung* von ζ heißt.

DEFINITION 30. *Es seien a, b reelle Zahlen, mit $a < b$ und X ein normierter Raum. Eine Abbildung $\phi : (a, b) \to X$ heißt* Treppenfunktion*, falls es eine Zerlegung $\zeta := \{x_0, \ldots, x_n\}$ ihres Definitionsbereichs (a, b) gibt, so dass die Einschränkungen von ϕ auf (x_k, x_{k+1}) für alle $k = 0, \ldots, n - 1$ konstant sind, etwa*

$$\phi_{|(x_k, x_{k+1})} \equiv: c_k \in X.$$

Die Menge der Treppenfunktionen von (a, b) nach X bezeichnet man mit $\mathcal{T}(a, b; X)$.

Die Summe

$$\int_a^b \phi(x)\, dx := \sum_{k=0}^{n-1} c_k (x_{k+1} - x_k)$$

nennen wir (bestimmtes) Integral von ϕ zwischen a und b.

Mann kann offensichtlich (a, b) in unendlich vielen Weisen derart zerlegen, dass eine gegebene Treppenfunktion f auf jeden Teilintervall konstant ist. Doch ist die obige Definition berechtigt, da folgendes gilt.

LEMMA 2.1. *Es seienb a, b reelle Zahlen, mit $a < b$ und X ein normierter Raum. Es sei $\phi : (a, b) \to X$ eine Treppenfunktion. Das Integral $\int_a^b f(x)\, dx$ hängt nicht von der gewählten Zerlegung ab.*

BEWEIS. Es seien ζ, ζ' zwei Zerlegungen. Es reicht aus, die Aussage im Fall, dass $\zeta \subset \zeta'$ zu zeigen; denn dann würde es im allgemeinen Fall einfach die neue Zerlegung $\zeta'' := \zeta \cup \zeta'$ zu betrachten: denn nach dem ersten fall würde dann beide Intergrale bezüglich ζ bzw. bezüglich ζ' mit dem Intergral bezüglich ζ'' übereinstimmen. In der Tat reicht es sogar den Fall zu betrachten, dass $\zeta' \setminus \zeta$ aus einem einzigen Punkt besteht, da der allgemeine Fall wohl induktiv behandelt werden kann.

Es sei also $\zeta := \{x_0, \ldots, x_n\}$ und $\zeta' := \{x_0, \ldots, x_m, y, x_{m+1}, x_n\}$ mit $x_0 < x_1 \ldots x_m < y < x_{m+1} < \ldots x_n$, so dass

$$\phi_{|(x_k, x_{k+1})} \equiv: c_k \in X, \qquad k = 0, 1, \ldots, n,$$

sowie auch

$$\phi_{|(x_m, y)} \equiv: d_+ \in X, \qquad \phi_{|(y, x_{m+1})} \equiv: d_- \in X.$$

Da aber die Einschränkung von $\phi_{|(x_k,x_{k+1})}$ auf (x_m, y) mit $\phi_{|(x_m,y)}$, bzw. die Einschränkung von $\phi_{|(x_k,x_{k+1})}$ auf (y, x_{m+1}) mit $\phi_{|(y.x_{m+1})}$ übereinstimmen, muss notwendigerweise

$$d_- = c_m = d_+$$

gelten. Die Aussage folgt dann unmittelbar. $\square$

Anmerkung 31. Man sieht, dass die Definitionen von Zerlegungen und Treppenfunktionen sich auch im Fall nichtoffener (beschränker) Intervalle problemlos formulieren lässt. Allerdings zeigt die Definition vom bestimmten Integral einer Funktion $\phi : (a, b) \to X$ unmittelbar, dass die Integrale zwischen a und b von ϕ und von ihren Einschränkungen auf $[a, b)$, $(a, b]$, oder (a, b) übereinstimmen. Somit werden wir im folgenden immer wieder der Einfachheit halber nur den Fall von auf offenen Intervallen definierten Funktionen explizit betrachten.

Jetzt wollen wir einige Rechenregel für die Integrale von Treppenfunktionen beweisen.

Satz 2.2. *Es seien $f, g : (a, b) \to X$ Treppenfunktionen und $\alpha, \beta \in \mathbb{K}$. Dann gelten die folgenden Rechenregel.*

(1) Für alle $c \in (a, b)$ sind die Einschränkungen $f_{|(a,c)}, f_{|(c,b)}$ Treppenfunktionen und

$$\int_a^b f(x)\, dx = \int_a^c f(x)\, dx + \int_c^b f(x)\, dx.$$

(2) $\alpha f + \beta g$ ist eine Treppenfunktion und

$$\int_a^b (\alpha f + \beta g)(x)\, dx = \alpha \int_a^b f(x)\, dx + \beta \int_a^b g(x)\, dx.$$

(3) $\|f\|_X$ ist eine (reellwertige) Treppenfunktion und

$$\left\| \int_a^b f(x)\, dx \right\| \leq \int_\alpha^\beta \|f(x)\|_X\, dx \leq \|f\|_\infty |b - a|.$$

(4) Ist $X = \mathbb{R}$ und gilt $f \leq g$, so ist auch

$$\int_a^b f(x)\, dx \leq \int_a^b g(x)\, dx.$$

Zur Erinnerung: $f \leq g$ bedeutet, dass

$$f(x) \leq g(x) \qquad \text{für alle } x \in (a, b).$$

Beweis. Da die Integrale zwischen a und b definitionsgemäß nichts anderes als endliche Summen sind, gelten alle Aussagen selbstverständlich: Z.B. beruht die erste Gleichung darauf, dass die Vereinigung zweier Zerlegungen ζ von (a, c) und ζ' von (b, c) auch eine Zerlegung $\zeta \cup \zeta'$ von (a, b) liefert, so dass die Aussage eine Folgerung der Assoziativität der Addition ist. Die zweite Aussage folgt aus der Distributivität der Multiplikation. Die dritte Aussage folgt aus der Dreiecksungleichung für $| \cdot |$. Schließlich gilt die letzte Aussage, da $\mathbb{R}$ ein geordneter Körper ist, und somit aus $m, n, p, q \in \mathbb{R}$ mit $m \leq n$ und $p \leq q$ folgt $m + p \leq n + p \leq n + q$. $\square$

Beispiel 32. Es sei $c \in \mathbb{R}$ und betrachte die Funktion $f : [a, b] \to \mathbb{R}$ mit konstantem Wert c. Dann gilt

$$\int_a^b f(x)\, dx = c(b - a).$$

Denn von der gewählte Zerlegung von $[a,b]$ unabhängig sind offensichtlich auch die approximierenden Funktionen ϕ_u und ϕ_o konstant mit Wert c, und somit gilt

$$\begin{aligned}\inf_{\zeta_n\in\mathcal{P}}\sum_{k=0}^{n-1}\phi_o(x)(x_{k+1}-x_k) &= c\inf_{\zeta_n\in\mathcal{P}}\sum_{k=0}^{n-1}(x_{k+1}-x_k)=c(b-a)\\ &= c\sup_{\zeta_n\in\mathcal{P}}\sum_{k=0}^{n-1}(x_{k+1}-x_k)=\sup_{\zeta_n\in\mathcal{P}}\sum_{k=0}^{n-1}\phi_u(x)(x_{k+1}-x_k).\end{aligned}$$

□

Wir wollen nun die obige Konstruktion auf allgemeinere Funktionen verallgemeinern. Dazu benötigen wir erst eine technische Definition.

Definition 33. *Es seien X ein normierter Raum, $a,b\in\mathbb{R}$ mit $a<b$ und $x_0\in[a,b]$. Es seien $f:[a,b]\to X$ und $y\in X$. Man sagt, y* ist linksseitiger *(bzw.* rechtsseitiger*)* Grenzwert von f für x gegen x_0 *falls y der Grenzwert von $f_{|\{x\in[a,b]:x<x_0\}}$ (bzw. von $f_{|\{x\in[a,b]:x>x_0\}}$) für x gegen x_0 ist, d.h.,*

$$\forall\varepsilon>0\ \exists\delta>0 \text{ so dass } |y-f(x)|<\varepsilon \text{ für alle } x\in[a,x_0) \text{ mit } |x-x_0|<\delta$$

bzw.

$$\forall\varepsilon>0\ \exists\delta>0 \text{ so dass } |y-f(x)|<\varepsilon \text{ für alle } x\in(x_0,b] \text{ mit } |x-x_0|<\delta.$$

Man schreibt dann

$$\lim_{x\to x_0-}f(x)=y\quad(\text{bzw. }\lim_{x\to x_0+}f(x)=y).$$

Besitzt eine Funktion an einer Stelle x_0 sowohl linksseitigen als auch rechtsseitigen Grenzwert, so sagt man, dass sie beide einsseitigen Grenzwerte *besitzt.*

Beispiel 34. Die Gaußsche Klammerfunktion ist unstetig, hat aber an jeder Stelle beide einsseitigen Grenzwerte: für $x\notin\mathbb{Z}$ stimmen beide einsseitigen Grenzwerte mit $\lfloor x\rfloor$ überein, während für $x\in\mathbb{Z}$ ist der linksseitiger Grenzwert $\lfloor x-1\rfloor$ und der rechtsseitiger Grenzwert $\lfloor x\rfloor$.

Dagegen hat die Funktion $\mathbb{R}\ni x\mapsto\frac{1}{x}\in\mathbb{R}$ an der Stelle $x_0=0$ keinen der beiden einsseitigen Grenzwerte, da $\lim_{x\to0\pm}\frac{1}{x}\notin\mathbb{R}$. □

Anmerkung 35. Auch in diesem Fall ist es klar, dass man die obige Definition sich auch an Funktionen erweitern kann, die auf einem nichtabgeschlossenen Intervall definiert sind. Allerdings bevorzugen wir auf technischen Gründen, Regelfunktionen immer als auf abgeschlossenen Intervallen definiert zu betrachten.

Eine wichtige Beispielklasse liefern die monotonen Funktionen.

Satz 2.3. *Es sei $A\subset\mathbb{R}$ ein Intervall und $f:A\to\mathbb{R}$ eine monotone Funktion. Dann besitzt sie an jeder Stelle beide einsseitigen Grenzwerte.*

Beweis. Wir betrachten o.B.d.A. den Fall, dass f monoton wachsend ist. Es sei dazu $x_0\in A$. Ist $x_0=\inf A$, dann ist $inf f(A)=\lim_{x\to x_0+}f(x)$. Ist jedoch $x_0>\inf A$, so wollen wir zeigen, dass

$$\lim_{x\to x_0-}f(x)=s:=\sup\{f(x)\in\mathbb{R}:x\in A\cap(-\infty,x_0)\}.$$

Es sei $\varepsilon > 0$ und $\xi \in A \cap (-\infty, x_0)$ mit $s - \varepsilon < f(\xi)$. Somit ist $s - \varepsilon < f(x) \leq s$ für alle $x \in (\xi, x_0)$, für $\varepsilon > 0$ beliebig klein. Dies zeigt, dass $\lim_{x \to x_0 -} f(x) = s$. Ähnlich zeigt man, dass

$$\lim_{x \to x_0 +} f(x) = i := \inf\{f(x) \in \mathbb{R} : x \in A \cap (x_0, +\infty)\}.$$

Damit ist der Beweis vervollständigt. □

Definition 36. *Es seienb a, b reelle Zahlen, mit $a < b$ und X ein normierter Raum. Eine Abbildung $\phi : (a, b) \to X$ heißt* Regelfunktion *oder* sprungstetig, *falls sie an jeder Stelle $x_0 \in (a, b)$ beide einseitigen Grenzwerte besitzt, und zusätzlich den linken einseitigen Grenzwert an der Stelle $x_0 = b$ und den rechten einseitigen Grenzwert an der Stelle $x_0 = a$. Die Menge der Regelfunktionen von (a, b) nach X bezeichnet man mit $\mathcal{R}([a, b]; X)$.*

Beispiel 37. Nach Definition ist jede stetige Funktion eine Regelfunktion. Nach dem Lemma 2.3 sind monotone Funktione auch Regelfunktionen. Die Funktion $x \mapsto \frac{1}{x}$ ist eine Regelfunktion, wenn sie auf einem Intervall eingeschränkt wird, welches die 0 nicht enthält, wohl aber keine Regelfunktion, wenn die Endpunkte des Intervalls jeweils strikt positiv und strikt negativ sind. Jede Einschränkung der Dirichlet-Funktion auf einem Intervall, dass mehr als einen Punkt enthält, ist keine Regelfunktion. □

Anmerkung 38. Die Gaußsche Klammerfunktion, deren Sprungstetigkeit wir direkt bewiesen haben, ist natürlich eine monotone Funktion. Ihr Betrag allerdings nicht, so dass man den obigen Satz nicht direkt anwenden dürfte, Eine kurze überlegung zeigt allerdings, dass die Srpungstetigkeit eine rein *lokale* Eigenschaft ist – also eine Eigenschaft, die an jeder Stelle nur durch beliebig kleine Umgebungen derselben Stelle beinträchtigt werden kann. Um ein Beispiel eine *nicht*sprungstetigen Funktion zu finden, müssen wir also an Funktionen denken, die nicht mal lokal monoton sind. Mindestens eine solche Funktion kennen wir schon: die Dirichlet-Funktion aus [**6**, Bsp. 7.58] – welche tatsächlich keine Regelfunktion ist.

Anmerkung 39. Eine Funktion $f : [a, b] \to \mathbb{R}$ heißt *stückweise stetig*, falls eine Zerlegung $\zeta_n := \{x_0, \ldots, x_n\}$ von $[a, b]$ existiert so, dass jede Einschränkung $f_{|(x_k, x_{k+1})}$ stetig ist. Selbstverständlich ist jede stückweise stetige Funktion eine Regelfunktion.

Übungsaufgabe 40. Es seienb a, b reelle Zahlen, mit $a < b$ und X ein normierter Raum. Zeige: Genau dann ist $f : [a, b] \to X$ eine Regelfunktion, wenn alle folgende Bedingungen erfüllt sind.

- Es gibt $x \in X$ so, dass für *jede* gegen a konvergente Folge $(x_n)_{n \in \mathbb{N}} \subset (a, b]$ auch $\lim_{n \to \infty} f(x_n) = x$ gilt.
- Für *jedes* $x_0 \in (a, b)$ gibt es $x \in X$ so, dass für *jede* gegen x_0 konvergente Folge $(x_n)_{n \in \mathbb{N}} \subset [a, x_0)$ auch $\lim_{n \to \infty} f(x_n) = x$ gilt.
- Für *jedes* $x_0 \in (a, b)$ gibt es $y \in X$ so, dass für *jede* gegen x_0 konvergente Folge $(y_n)_{n \in \mathbb{N}} \subset (x_0, b]$ auch $\lim_{n \to \infty} f(y_n) = y$ gilt.
- Es gibt $y \in X$ so, dass für *jede* gegen b konvergente Folge $(y_n)_{n \in \mathbb{N}} \subset [a, b)$ auch $\lim_{n \to \infty} f(y_n) = y$ gilt.

Theorem 2.4. *Es seienb a, b reelle Zahlen, mit $a < b$ und X ein Banachraum. Dann ist $f : [a, b] \to X$ genau dann eine Regelfunktion, wenn sie gleichmäßiger Grenzwert einer Folge $(f_n)_{n \in \mathbb{N}}$ von Treppenfunktionen ist.*

Beweis. Es seien $f : [a, b] \to X$ eine Regelfunktion und $n \in \mathbb{N} \setminus \{0\}$. Dann gibt e zu jedem $x \in [a, b]$ zwei Punkte $\alpha(x), \beta(x) \in \mathbb{R}$ mit $\alpha(x) < x < \beta(x)$ und so, dass mindestens eine der Aussagen

- $\|f(s) - f(t)\|_X \leq \frac{1}{n}$ für alle $s, t \in (\alpha(x), x) \cap [a, b]$,
- $\|f(s) - f(t)\|_X \leq \frac{1}{n}$ für alle $s, t \in (x, \beta(x)) \cap [a, b]$,

gilt. Somit ist $[a, b] \subset \bigcup_{x \in [a,b]} (\alpha(x), \beta(x))$. Aus der Kompaktheit von $[a, b]$ [1] folgt, dass es eine endliche Teilmenge (o.B.d.A. sogar eine Zerlegung) ζ von $[a, b]$ gibt, so dass

$$[a, b] \subset \bigcup_{x \in \zeta} (\alpha(x), \beta(x)).$$

Wir können nun ζ so verfeinern, dass wir eine neue Zerlegung $\tilde{\zeta} := (\xi_0, \ldots, \xi_m)$ finden, mit

- $\|f(s) - f(t)\|_X \leq \frac{1}{n}$ für alle $s, t \in (\xi_{j-i}, \xi_j)$, $j = 1, \ldots, m$:

das wird z.B. dadurch erreicht, dass wir ζ ergänzen durch alle (endlich viele) Punkte der Form $\alpha(z), \beta(z)$, wobei $z \in \zeta$ (so lang $\alpha(z), \beta(z) \in [a, b]$): dann gehören je zwei Punkte $s, t \in (\xi_{j-i}, \xi_j)$ *mindestens* einem gemeinsamen Intervall der Form $(\alpha(z), z)$ oder $(z, \beta(z))$ an, und somit gilt die behauptete Abschätzung. Betrachte nun eine Funktionenfolge $(f_n)_{n \in \mathbb{N}}$, die durch

$$f_n(x) : \begin{cases} f(x), & \text{falls } x \in \{\xi_0, \ldots, \xi_m\}, \\ f(\frac{(\xi_{j-i} + \xi_j}{2}), & \text{falls } x \in (\xi_{j-i}, \xi_j),\ j = 1, \ldots, m. \end{cases}$$

definiert wird. Dann ist jedes $f_n : [a, b] \to \mathbb{X}$ nach Konstruktion eine Treppenfunktion und es gilt auch

$$\|f(x) - f_n(x)\|_X < \frac{1}{n}, \qquad x \in [a, b].$$

Somit haben wir bewiesen, dass $(f_n)_{n \in \mathbb{N}}$ gleichmäßig gegen f konvergiert.

Es sei umgekehrt eine Funktionenfolge $(f_n)_{n \in \mathbb{N}}$ gegeben, die gleichmäßig gegen f konvergiert. Wir wollen zeigen, dass f eine Regelfunktion ist, indem wir das Kriterium in der Übungsaufgabe 40 anwenden.

Als Vorbereitung zeigen wir erst, dass f lokal „fast konstant" sein muss. Es seien dazu $\varepsilon > 0$ und $x \in (a, b)$ beliebig. Dann gilt

$$\|f - f_n\|_\infty < \varepsilon \qquad \text{für alle } n \geq N, \tag{2.1}$$

für ein N groß genug. Da f_N eine Treppenfunktion ist, kann man zu jedem $x_0 \in (a, b)$ ein $a' \in [a, x_0)$ finden, so dass $f_{N|(a', x_0)}$ konstant ist. Somit findet man, dass

$$\|f(s) - f(t)\|_X \leq \|f(s) - f_N(s)\|_X + \|f_N(s) - f_N(t)\|_X + \|f_N(t) - f(t)\| < 2\varepsilon \qquad \text{für alle } s, t \in (a', x_0), \tag{2.2}$$

wobei der erste und dritte Term in der rechten Seite mittels (2.1) abgeschätzt werden können, und der mittlere Term auf (a', x_0) identisch verschwindet.

Betrache nun eine beliebige Folge $(s_k)_{k \in \mathbb{N}} \subset [a, x_0)$, die gegen x_0 konvergiert. Für ein beliebiges $\varepsilon > 0$ kann man ein $M \in \mathbb{N}$ finden so, dass

$$s_k \in (a', x_0) \qquad \text{für alle } k \geq M.$$

Angesichts von (2.2) folgt somit, dass

$$\|f(s_j) - f(s_k)\|_X < 2\varepsilon \qquad \text{für alle } j, k \geq M, \tag{2.3}$$

d.h., $(f(s_k))_{k \in \mathbb{N}} \subset X$ ist eine Cauchy-Folge. Da X ein Banachraum ist, ist $(f(s_k))_{k \in \mathbb{N}}$ bereits konvergent, etwa gegen $x \in X$. Wir müssen nun zeigen, dass $(f(t_k))_{k \in \mathbb{N}}$ auch für alle weitere gegen x_0 konvergente Folgen $(t_k)_{k \in \mathbb{N}} \subset [a, x_0)$ gegen $x \in X$ konvergiert. Es sei dazu $\varepsilon > 0$. Das obige Argument zeigt nämlich

[1] Anschaulich ist diese Aussage wahrscheinlich klar, formal wird sie erst im Satz 6.4 bewiesen.

die Konvergenz von $(f(t_k))_{k\in\mathbb{N}}$, etwa gegen ein $y \in X$. Da aber es ein $\tilde{M} \in \mathbb{N}$ gibt so, dass $t_k \in (a', x_0)$ für alle $k \geq \tilde{M}$, Wieder aus (2.1) folgt somit

$$\|f(s_j) - f(t_k)\|_X < 2\varepsilon \qquad \text{für alle } j, k \geq M,$$

und im Grenzwert für $j, k \to \infty$

$$\|x - y\|_X < 2\varepsilon.$$

Daher gilt $x = y$. Dies zeigt, dass f einen linksseitigen Grenzwert an der Stelle x_0 hat. Genauso zeigt man die restlichen drei Bedingungen in der Übungsaufgabe 40. Dies vervollständigt den Beweis. □

Anmerkung 41. Die Abänderung einer Regelfunktion an endlich vielen Stellen ändert ihre Sprungsstetigkeit nicht. An jeder Stelle stimmen sogar die einsseitigen Grenzwerte der alten und der neuen Funktion überein.

Insbesondere können wir die Konvergenz einer Folge von Treppenfunktionen gegen eine Regelfunktion f betrachten, selbst wenn die ersteren nur im offenen Intervall (a, b) definiert sind und die letzteren im abgeschlossenen Intervall $[a, b]$, da man bei Bedarf auch die Treppenfunktionen am Intervallrand wie $f(a)$ bzw. $f(b)$ definieren kann, ohne dass ihre Integrale sich ändern.

Lemma 2.5. *Es seien a, b reelle Zahlen, mit $a < b$ und X ein normierter Raum. Dann gelten die folgenden Aussagen.*

(1) $\mathcal{R}([a, b]; X)$ ist ein Vektorraum, ja ein normierter Unterraum des Raumes $\mathcal{B}([a, b]; X)$ der beschränkten Funktionen von $[a, b]$ nach X.
(2) $\mathcal{T}(a, b; X)$ ist ein Unterraum von $\mathcal{R}([a, b]; X)$.
(3) Ist X ein Banachraum, so ist $\mathcal{R}([a, b]; X)$ selbser ein Banachraum.

Beweis. 1) Für die einsseitigen Grenzwerte gelten die üblichen Rechenregeln (Übungsaufgabe!), und somit haben Summen zweier Regelfunktionen und Produkte eines Skalars und einer Funktion an jeder Stelle einsseitige Grenzwerte. Lass uns jetzt annehmen, es gibt eine unbeschränlte Funktion $\phi \in \mathcal{R}([a, b]; X)$. Somit gibt es für jedes $n \in \mathbb{N}$ ein x_n so, dass $\|f(x_n)\|_X \geq n$. Da $[a, b]$ kompakt ist, ist $(x_n)_{n\in\mathbb{N}}$ beschränkt und hat somit – nach dem Satz von Bolzano–Weierstraß – eine (etwa gegen $x_0 \in [a, b]$) konvergente Teilfolge $(x_{n_k})_{k\in\mathbb{N}}$. Ist $x_0 = b$, so kann man $(x_{n_k})_{k\in\mathbb{N}}$ o.B.d.A. als monoton wachsend voraussetzen (denn i.A. hat jede reellwertige Folge eine monotone Teilfolge – ÜA). Ist jedoch $x_0 \in [a, b)$, so können wir ähnlich o.B.d.A. als monoton fallend voraussetzen. Weil ϕ eine Regelfunktion ist, gibt es dann in beiden Fällen den einseitigen (linksseitigen im ersteren, rechtsseitig im zweiteren Fall) Grenzwert

$$\lim_{k\to\infty} \phi(x_{n_k}) =: y \in X,$$

und, wegen der Stetigkeit der Norm, auch

$$\lim_{k\to\infty} \|\phi(x_{n_k})\|_X = \|y\|.$$

Somit ist die Folge $(\|\phi(x_{n_k})\|_X)_{n\in\mathbb{N}}$ einerseits konvergent, andererseits unbeschränkt – ein Widerspruch.

2) Die zweite Aussage gilt offensichtlich, so bald wir jede Treppenfunktion $\phi : (a, b) \to X$ mit einer Funktion von $\tilde{\phi} : [a, b] \to X$ identifizieren, indem wir sie willkürlich auf $[a, b]$ fortsetzen – etwa durch $\phi(a) = \phi(b) := 0$, vgl. die Anmerkung 41.

3) WIr wissen schon aus , dass $\mathcal{R}([a, b]; X)$ ein Unterraum von $\mathcal{B}([a, b]; X)$ ist und auch, dass $\mathcal{T}(a, b; X)$ ein Unterraum von $\mathcal{R}([a, b]; X)$ ist. Aufgrund vom Satz 2.4 stimmt der Abschluss von $\mathcal{T}(a, b; X)$ (bzgl. der Norm von $\mathcal{B}([a, b]; X)$) mit $\mathcal{R}([a, b]; X)$ überein. Somit ist $\mathcal{R}([a, b]; X)$ abgeschlossener Unterraum eines Banachraums, und somit selber ein Banachraum. □

Übungsaufgabe 42. Zeige: Jede reellwertige Folge hat eine monotone Teilfolge.

Theorem 2.6. *Es seienb a, b reelle Zahlen, mit $a < b$ und X ein Banachraum. Es sei $f : [a, b] \to X$ eine Regelfunktion, und betrachte eine Folge $(f_n)_{n\in\mathbb{N}} \subset \mathcal{T}(a, b; X)$, welche gleichmäßig gegen f konvergiert. Dann existiert der Grenzwert*

$$\lim_{n\to\infty} \int_a^b f_n(x)\, dx \in X$$

und er hängt nicht von der Wahl der approximierenden Folge $(f_n)_{n\in\mathbb{N}}$.

Beweis. Lass uns erst die Vektoren

$$I_n := \int_a^b f_n(x)\, dx$$

betrachten, die als Integrale von Treppenfunktionen sicherlich wohl definiert sind. Darüber hinaus bildet $(I_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge. Um dies zu sehen, beachte, dass für alle $m, n \in \mathbb{N}$ $f_n - f_m$ eine Treppenfunktion ist (dank dem Satz 2.2.(2)) und somit I_n überhaupt wohl definiert ist und (dank dem Satz 2.2.(3)) die Abschätzung

$$\|I_n - I_m\|_X \le (b-a)\|f_n - f_m\|_\infty \le (b-a)(\|f_n - f\|_\infty + \|f - f_m\|_\infty)\,, \qquad n, m \in \mathbb{N},$$

gilt. Weil nach Voraussetzung $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$ folgt auch, dass $\|I_n - I_m\|_X$ beliebig klein wird, wenn m, n groß genug werden. Somit konvergiert die Folge $(I_n)_{n\in\mathbb{N}}$, da X ein Banachraum ist.

Wir zeigen jetzt, dass der Grenzwert wohl nicht von der Folge $(f_n)_{n\in\mathbb{N}}$ abhängt. Betrachte nämlich eine neue Folge $(g_n)_{n\in\mathbb{N}} \subset \mathcal{T}(a, b; X)$ mit $\lim_{n\to\infty} \|g_n - f\|_\infty = 0$ und definiere die Folge $(h_n)_{n\in\mathbb{N}}$ durch

$$h_{2n} := f_n \quad \text{und} \quad h_{2n+1} := g_n, \qquad n \in \mathbb{N}.$$

Dann ist auch $\lim_{n\to\infty} \|h_n - f\|_\infty = 0$ (Übungsaufgabe!)und somit muss die Folge $\left(\int_a^b h_n(x)\, dx\right)_{n\in\mathbb{N}}$ konvergieren. Insbesondere haben ihre beide Teilfolgen

$$\left(\int_a^b f_n(x)\, dx\right)_{n\in\mathbb{N}} \quad \text{bzw.} \quad \left(\int_a^b g_n(x)\, dx\right)_{n\in\mathbb{N}}$$

den selben Grenzwert. □

Definition 43. *Es seienb a, b reelle Zahlen, mit $a < b$ und X ein Banachraum. Es sei $\phi : [a, b] \to X$ eine Regelfunktion. Der eindeutige Grenzwert, der im Satz 2.6 eingeführt wird, heißt* (Riemannsches) Integral von ϕ zwischen a und b.

Somit haben wir insbesondere für jede stückweise stetige Funktion und jede stückweise monotone Funktion ein Integral definiert. Weiter gelten die folgenden Rechenregeln, welche genau den Regeln für Treppenfunktionen entsprechen.

Satz 2.7. *Es seienb a, b reelle Zahlen, mit $a < b$ und X ein Banachraum. Es seien $f, g : [a, b] \to X$ Regelfunktionen und $\alpha, \beta \in \mathbb{K}$. Dann gelten die folgenden Rechenregel.*

(1) Für alle $c \in (a, b)$ sind die Einschränkungen $f_{|(a,c)}, f_{|(c,b)}$ Regelfunktionen und

$$\int_a^b f(x)\, dx = \int_a^c f(x)\, dx + \int_c^b f(x)\, dx.$$

(2) $\alpha f + \beta g$ *ist eine Regelfunktion und*

$$\int_a^b (\alpha f + \beta g)(x)\, dx = \alpha \int_a^b f(x)\, dx + \beta \int_a^b g(x)\, dx.$$

(3) $\|f\|_X$ *ist eine (reellwertige) Regelfunktion und*

$$\left\| \int_a^b f(x)\, dx \right\| \leq \int_\alpha^\beta \|f(x)\|_X\, dx \leq \|f\|_\infty |b-a|.$$

(4) Ist $X = \mathbb{R}$ *und gilt* $f \leq g$*, so ist auch*

$$\int_a^b f(x)\, dx \leq \int_a^b g(x)\, dx.$$

Beweis. (1) Betrachte zwei Folgen $(g_n)_{n\in\mathbb{N}}, (h_n)_{n\in\mathbb{N}}$ von Treppenfunktionen, welche gegen $f_{|(a,b)}$ bzw. $f_{|(b,c)}$. Dann definiert

$$f_n(x) := \begin{cases} g_n(x), & \text{falls } x \in (a,c), \\ f(c), & \text{falls } x = c, \\ h_n(x), & \text{falls } x \in (c,b), \end{cases}$$

eine neue Folge von Treppenfunktionen von (a,b) nach X, welche gleichmäßig gegen f konvergiert. Somit folgt die Aussage aus dem Satz 2.2.(1), denn für die approximierende Folge von Treppenfunktionen $(f_n)_{n\in\mathbb{N}}$

$$\begin{aligned}
\int_a^b f(x)dx &= \lim_{n\to\infty} \int_a^b f_n(x)dx \\
&= \lim_{n\to\infty} \left(\int_a^c f_n(x)dx + \int_c^b f_n(x)\, dx \right) \\
&= \lim_{n\to\infty} \left(\int_a^c g_n(x)dx + \int_c^b h_n(x)\, dx \right) \\
&= \int_a^c f_{|(a,c)}(x)dx + \int_c^b f_{|(c,b)}(x)\, dx \\
&= \int_a^c f(x)dx + \int_c^b f(x)\, dx
\end{aligned}$$

(2) Es seien $(f_n)_{n\in\mathbb{N}}$ und $(g_n)_{n\in\mathbb{N}}$ Folgen von Treppenfunktionen, welche gegen f bzw. g konvergieren. Dann konvergiert auch $(\alpha f_n + \beta g_n)_{n\in\mathbb{N}}$ gegen $\alpha f + \beta g$. Angesichts vom Theorem 2.6 kann die Aussage mittels $(\alpha f_n + \beta g_n)_{n\in\mathbb{N}}$ untersucht werden. Tatsächlich gilt aufgrund vom Satz 2.2.(2)

$$\begin{aligned}
\int_a^b (\alpha f + \beta g)(x)dx &= \lim_{n\to\infty} \int_a^b (\alpha f_n + \beta g_n)(x)dx \\
&= \alpha \lim_{n\to\infty} \int_a^b f_n(x)dx + \beta \lim_{n\to\infty} \int_a^b g_n(x)dx \\
&= \alpha \int_a^b f(x)dx + \beta \int_a^b g(x)dx.
\end{aligned}$$

(3) Überlassen wir als Übungsaufgabe.

(4) Man kann o.B.d.A. den Fall von $f \equiv 0$ betrachten, da sonst kann man $\phi := f - g \geq 0$ betrachten, welches aufgrund von (1) auch eine Regelfunktion ist. Der erste Teil im Beweis vom Satz 2.4 zeigt, dass zu jeder Regelfunktion $f \geq 0$ eine approximierende Folge von Treppenfunktionen gefunden werden kann, die zudem alle positiv-wertig sind. Somit ist laut dem Satz 2.2.(4) auch ihr integral positiv, und auch der Grenzwert ihrer Integrale – d.h., das Integral von f. $\square$

Korollar 2.8. *Es seienb a, b reelle Zahlen, mit $a < b$ und X ein Banachraum. Dann ist die Abbildung*

$$\mathcal{I} := \mathcal{R}([a,b]; X) \ni f \mapsto \int_a^b f(x)dx \in X$$

stetig.

Beweis. Wir werden sogar zeigen, dass die obige Abbildung Lipschitzstetig ist – Stetigkeit wird dann aus [**6**, ÜA 7.90] folgen. Es seien nämlich $f, g \in \mathcal{R}([a,b]; X)$. Dann gilt aufgrund von Satz 2.7.(2)–(4)

$$\|\mathcal{I}(f) - \mathcal{I}(g)\|_\infty = \| \int_a^b (f-g)(x)dx\|_\infty \leq \|f-g\|_\infty (b-a).$$

Dies zeigt die Behauptung. $\square$

Übungsaufgabe 44. Beweise den Satz 2.7.(3)

Anmerkung 45. Beachte: In unserem Zugsang gibt es keine *nicht*integrierbare Funktionen, sondern nur Funktionen, die nicht die Voraussetzungen des Satzes 2.6 erfüllen und somit für die kein Integral im Rahmen der Definition 43 definiert ist. In diesem Sinne unterscheidet sich dieser Zugang von jenem, der auf Verwendung der sogenannten Untersummen und Obersummen beruht, vgl. z.B. [**3**] – und deshalb für die Integration Vektorwertiger Funktionen nicht geeignet ist. Den Zugang, den wir folgen, stimmt im Wesentlichen mit dem in [**5**] überein. Riemannsche Integrale werden in [**2**] in einer leicht unterschiedlichen Weise definiert – dort wird u.A. auch der Begriff der *Integrierbarkeit* in [**2**, Bemerk. VI.3.5] definiert.

Tatsächlich gibt es die Möglichkeit, eine weitere Fortsetzung des Integralbegriffs einzuführen. Henri Léon Lebesgue hat 1902 eine Integrierbarkeitsbedingung definiert, die strikt schwächer als Sprungstetigkeit ist – und somit eine (strikt) allgemeinere Integrationstheorie einleitet. Nach seiner Theorie ist z.B. die Dirichletfunktion tatsächlich (Lebesgue-) integrierbar und ihr Integral beträgt 0. Die Lebesguesche Theorie ist jedoch technisch deutlich komplizierter und wird deshalb erst in einem späteren Semester vorgeführt.

Anmerkung 46. Man verwendet auch zwei zusätzliche Notationen, die eigentlich nicht aus der Definition 43 oder aus dem Satz 2.7 folgen. Man setzt nämlich

$$\int_a^a f(x)dx := 0$$

und auch

$$\int_b^a f(x)dx := -\int_a^b f(x)dx,$$

falls $a < b$ und $f : [a,b] \to \mathbb{R}$ eine Regelfunktion ist.

Wir fassen im Folgenden die ersten wesentlichen Eigenschaften des Riemannschen Integrals zusammen. Der Folgende ist als *Mittelwertsatz der Integralrechnung* bekannt.

Satz 2.9. *Es seienb a, b reelle Zahlen, mit $a < b$ und $f : [a, b] \to \mathbb{R}$ eine Regelfunktion mit $f \geq 0$. Es sei $g : [a, b] \to \mathbb{R}$ eine zusätzliche stetige Funktion. Dann gibt es $\xi \in [a, b]$ so, dass*

$$\int_a^b f(x)g(x)dx = g(\xi) \int_a^b f(x)dx.$$

Beachte insbesondere den Spezialfall für $f \equiv 1$:

$$\int_a^b g(x)dx = g(\xi)(b-a)$$

für ein $\xi \in [a, b]$.

Beweis. Da g stetig auf einem kompakten Intervall ist, hat sie ein Maximum M und ein Minimum m. Somit gilt für alle x aufgrund der Eigenschaften des geordneten Körpers $\mathbb{R}$, dass

$$f(x)m \leq f(x)g(x) \leq f(x)M \qquad \text{für alle } x \in [a, b].$$

Es folgt aus dem Satz 2.7.(1)–(4), dass

$$m \int_a^b f(x)dx \leq \int_a^b f(x)g(x)dx \leq M \int_a^b f(x)dx.$$

Somit gibt es

$$\mu \in [m, M] \tag{2.4}$$

mit

$$\mu \int_a^b f(x)dx = \int_a^b f(x)g(x)dx.$$

Wiederum folgt aus (2.4) und dem Zwischenwertsatz, an die stetige Funktion g angewendet, dass es $\xi \in [a, b]$ gibt mit $g(\xi) = \mu$. Dies vervollständigt den Beweis. □

Satz 2.10. *Es seienb a, b reelle Zahlen, mit $a < b$ und $f : [a, b] \to \mathbb{R}$ eine Regelfunktion mit $f \geq 0$. Gilt*

$$\int_a^b f(x)dx = 0,$$

so gilt $f(x_0) = 0$ an jeder Stelle x_0, wo f stetig ist.

Man kann in der Tat zeigen, dass eine Regelfunktion höchstens abzählbar viele Unstetigkeitsstellen haben darf.

Beweis. Es sei x_0 eine Stelle, an der f stetig ist. Wäre $f(x_0) \neq 0$, o.B.d.A. $f(x_0) > 0$, so kann man aufgrund der Stetigkeit von f eine Umgebung $(x_0 - \varepsilon, x_0 + \varepsilon) \subset (a, b)$ von x_0 finden, so dass

$$f(x) > \frac{1}{2} f(x_0) \qquad \text{für alle } x \in (x_0 - \varepsilon, x_0 + \varepsilon).$$

Betrachte nun die Treppenfunktion $\phi : [a, b] \to \mathbb{R}$, welche durch

$$\phi(x) := \begin{cases} \frac{1}{2} f(x_0), & x \in (x_0 - \varepsilon, x_0 + \varepsilon), \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt $f \geq \phi$ und angesichts von Satz 2.7.(4) $\int_a^b f(x)dx \geq \int_a^b \phi(x)dx$. Somit

$$0 < \varepsilon f(x_0) = \int_a^b \phi(x)dx \leq \int_a^b f(x)dx,$$

ein Widerspruch. □

Wir wollen jetzt die Resultate des ersten Kapitels über die Stetigkeit und Differenzierbarkeit von konvergenten Funktionenfolgen- und Reihen vervollständigen.

SATZ 2.11. *Es seienb a, b reelle Zahlen mit $a < b$, X ein Banachraum und $(f_n)_{n\in\mathbb{N}}$ eine Folge von Regelfunktionen von $[a, b]$ nach X. Dann gelten die folgenden Aussagen.*

(1) Ist $(f_n)_{n\in\mathbb{N}}$ gleichmäßig konvergent, etwa gegen f, so ist f eine Regelfunktion und

$$\int_a^b f(x)dx = \lim_{n\to\infty} \int_a^b f_n(x)dx.$$

(2) Ist $\sum_{n=0}^\infty f_n$ gleichmäßig konvergent, etwa gegen F, so ist F eine Regelfunktion und

$$\int_a^b F(x)dx = \sum_{n=0}^\infty \int_a^b f_n(x)dx.$$

BEWEIS. Wir wissen, dass Regelfunktionen bzgl. der $\|\cdot\|_\infty$-Norm einen Banachraum bilden. Insbesondere bleibt der gleichmäßiger Grenzwert einer Folge in $\mathcal{R}([a, b]; X)$ selber in $\mathcal{R}([a, b]; X)$. Da aber die Abbildung $\int_a^b : \mathcal{R}([a, b]; X) \to X$ nach dem Korollar 2.8 stetig ist, kommutiert sie mit dem Grenzwert – somit folgen beide Aussagen. □

ÜBUNGSAUFGABE 47. Zeige durch ein Gegenbeispiel, dass die Aussagen im Satz 2.11 i.A. nicht gelten, wenn die Konvergenz nur punktweise gilt.

ANMERKUNG 48. Beachte, dass die Benennung des Argumentes der Funktion keinerlei Rolle spielt: ist also $f : [a, b] \to X$ eine Regelfunktion, so bezeichnen etwa die Symbole

$$\int_a^b f(x)\,dx, \qquad \int_a^b f(y)dy, \qquad \int_a^b f(s)ds$$

den gleichen Vektor.

ANMERKUNG 49. Beachte dass das bestimmte Integral einer Funktion kann wohl eine negative Zahl sein. Interpretieren wir das Integral einer Funktion wie der Flächeninhalt zwischen dem Graphen von f und der x-Achse, so gilt die Konvention, dass Fläche *unterhalb* der x-Achse einen negativen Beitrag leistet.

ANMERKUNG 50. Es sei $f : [a, b] \to \mathbb{R}$. Ein sehr populärer Zugang zur Definition von Integral läuft über die sogenannten Darbouxschen Integrale: Definiere für jede Zerlegung $\zeta_n := \{x_0, \ldots, x_n\}$ von $[a, b]$ zwei approximierende Funktionen durch

$$\phi_u(x) := \inf_{x\in[x_k,x_{k+1})} f(x) \qquad \text{und} \qquad \phi_o(x) := \sup_{x\in[x_k,x_{k+1})} f(x) \qquad \text{falls } x \in [x_k, x_k + 1).$$

Sind nun ϕ_u, ϕ_o Funktionen von $[a,b] \to \mathbb{R}$ (d.h., $\phi_u(x), \phi_o(x)$ sind reelle Zahlen für alle $x \in [a,b]$) und stimmen die sogenannten *unteres* bzw. *oberes Darbouxschen Integrale* überein, d.h., gilt

$$\inf_{\zeta_n \in \mathcal{P}} \sum_{k=0}^{n-1} \phi_o(x)(x_{k+1} - x_k) = \sup_{\zeta_n \in \mathcal{P}} \sum_{k=0}^{n-1} \phi_u(x)(x_{k+1} - x_k),$$

so heißt f *Riemann-integrierbar* oder einfach *integrierbar*. Der gemeinsame Wert wird mit dann $\int_a^b f(x)dx$ bezeichnet und heißt *bestimmtes Integral* von f zwischen a und b.

Dieser Zugang ist allerdings nur mittelbar auf $\mathbb{C}^n$-wertige Funktionen anwendbar und läßt sich nicht auf allgemeinen vektorwertige Funktionen erweitern.

Definition 51. *Es seien $f, F : [a,b] \to X$ Funktionen. Dann heißt F* Stammfunktion *von f, falls es für alle $x \in [a,b]$ gilt:*

- *F ist an der Stelle x differenzierbar und*
- *$F'(x) = f(x)$.*

Satz 2.12. *Es seien $a, b \in \mathbb{R}$, X ein normierter Raum, $f, F : [a,b] \to \mathbb{R}$, und es sei F eine Stammfunktion von f. Dann gilt: Genau dann ist eine weitere Funktion $G : [a,b] \to X$ eine Stammfunktion von f, wenn $F - G$ eine Konstante ist.*

Beweis. Es sei G eine Stammfunktion von f. Dann gilt $(F-G)'(x) = 0$ für alle $x \in [a,b]$. Nun wende [**6**, Kor. 8.40] an. Umgekehrt: da die Ableitung einer konstanten Funktion $x \mapsto c$ identisch 0 ist, sieht man sofort, dass $(F+c)' = F' = f$. □

Anmerkung 52. Manche Autoren betrachten lieber eine allgemeinere Definiton, welche fordert, eine Stammfunktion F von f soll nur *fast überall* – also an allen Stelle, außer höchstens abzählbar vielen – sein, und dort eine Ableitung F' haben, welche mit f übereinstimmt. Dieser Begriff erlaubt tatsächlich, einige Fälle zu betrachten, die man sonst mit dem Hauptsatz nicht bewältigen können. Doch ist diese Theorie mit erheblichen technischen Komplikationen verbunden. Außerdem sollte man bedenken, dass die Bestimmung einer Stammfunktion grundsätzlich zur Anwendung des Hauptsatzes – genauer gesagt: des Korollars 2.14. Liegt aber eine nichtstetige Regelfunktion f vor, welche nur endlich viele Unstetigkeitsstellen $x_0, \ldots, x_N$ hat mit $a < x_0 < x_1 < \ldots < x_N < b$, so kann man angesichts vom Satz 2.7.(1) ihr bestimmtes Integral als

$$\int_a^b f(x)dx = \int_a^{x_0} f(x)dx + \sum_{k=0}^{N-1} \int_{x_k}^{x_{k+1}} f(x)dx + \int_{x_N}^b f(x)dx$$

darstellen. Da aber auf $(a, x_0), (x_0, x_1), \ldots, (x_{N-1}, x_N), (x_N, b)$ die Funktion f wohl stetig ist, hat sie dort eine Stammfunktion. Den gesamten Wert des bestimmten Integrals erhält man also durch wiederholte Anwendung des Korollars 2.14. Dies zeigt, dass eine Erweiterung des Hauptsatzes auf allgemeine Regelfunktionen nur für den Fall einer Funktion mit unendlich vielen Unstetigkeitsstellen dringend notwendig ist. Dazu verweisen wir auf [**5**, Kap. 11] – die Resultate dort hängen nicht wesentlich davon ab, dass die betrachteten Funktionen reellwertig sind.

Beispiel 53. Angesischts der bekannten Differenziationsregeln gelten folgende Aussagen:

- $\frac{1}{a}\exp_a$ ist eine Stammfunktion von $\exp_a : \mathbb{R} \to \mathbb{R}$;
- $x \mapsto \frac{x^{k+1}}{k+1}$ ist eine Stammfunktion von $x \mapsto x^k : \mathbb{R} \to \mathbb{R}$ für alle $k \in \mathbb{Z} \setminus \{-1\}$;
- sin ist eine Stammfunktion von $\cos : \mathbb{R} \to \mathbb{R}$;

- $-\cos$ ist eine Stammfunktion von $\sin : \mathbb{R} \to \mathbb{R}$;
- $\log$ ist eine Stammfunktion von $x \mapsto \frac{1}{x} : (0, \infty) \to \mathbb{R}$;
- $\tan$ ist eine Stammfunktion von $x \mapsto \frac{1}{\cos(x)^2} : \mathbb{R} \to \mathbb{R}$.

□

Beispiel 54. Es seien $a_0, a_1, \ldots, a_n \in \mathbb{R}$. Dann ist

$$x \mapsto a_0 x + a_1 \frac{x^2}{2} + \ldots + a_n \frac{x^{n+1}}{n+1}$$

eine Stammfunktion von

$$x \mapsto a_0 + a_1 x + \ldots + a_n x^n.$$

□

Der Folgende ist als *Hauptsatz der Differential- und Integralrechnung* bekannt.

Satz 2.13. *Es seien $a, b \in \mathbb{R}$, X ein normierter Raum. Es sei $f : [a, b] \to X$ eine Regelfunktion. Dann ist die Funktion*

$$F : [a, b] \ni x \mapsto \int_a^x f(s) ds \tag{2.5}$$

stetig und an jeder Stetigkeitsstelle von f differenzierbar.

Insbesondere ist F eine Stammfunktion von f im Sinne der Definition 51, falls f stetig ist.

Beweis. (1) Es reicht zu zeigen, dass F Lipschitzstetig ist. Dazu verwenden wir den Satz 2.7.(1)–(3) und erhalten für alle $x, y \in [a, b]$

$$\|F(x) - F(y)\|_X = \left\| \int_a^x f(s) ds - \int_a^y f(s) ds \right\|_X = \left\| \int_y^x f(s) ds \right\|_X \leq |x - y| \|f\|_\infty.$$

(bedenke dabei die Konvention, die in der Anmerkung 46 eingeführt wurde.

Es sei nun $\varepsilon > 0$ und wähle $\delta > 0$ so, dass $\|f(x) - f(x_0+)\|_X < \varepsilon$ für alle $x \in [a, b] \cap (x_0, x_0 + \delta)$ – vgl. den Anfang vom Beweis vom Satz 2.4. Dann gilt

$$\left\| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0+) \right\|_X = \left\| \frac{1}{x - x_0} \int_{x_0}^x (f(s) - f(x_0+)) ds \right\|_X < \frac{\varepsilon |x - x_0|}{|x - x_0|} = \varepsilon,$$

wieder wegen des Satzes 2.7.(3). Dabei verwenden wir die Schreibweise $f(x_0+) := \lim_{x \to x_0+} f(x)$. Ähnlich zeigt man die Existenz des anderen einseitigen Grenzwerts $f(x_0-) := \lim_{x \to x_0-} f(x)$. Dies vervollständigt den Beweis, da in den Stetigkeitsstellen x_0 von f gilt $f(x_0+) = f(x_0-)$. □

Die folgende konkrete Regel zur Berechnung eines Integrals folgt unmittelbar.

Korollar 2.14. *Es seien $a, b \in \mathbb{R}$, X ein normierter Raum. Es sei $f : [a, b] \to X$ eine stetige Funktion und Φ eine Stammfunktion von f. Dann gilt*

$$\int_a^b f(x)\, dx = \Phi(b) - \Phi(a).$$

Man bezeichnet oft

$$\Big[\Phi(x)\Big]_{x=a}^{x=b} := \Big[\Phi(x)\Big]_a^b := \Phi(b) - \Phi(a).$$

Beweis. i) Betrachte erst die Stammfunktion F, die in (2.5) definiert ist. Für sie gilt die Aussage offensichtlich, denn laut der Anmerkung 46 ist $\int_a^a f(x)\,dx = 0$.

ii) Es sei nun G eine weitere Stammfunktion, so existiert laut Satz 2.12 eine Konstante c so, dass $G(x) = F(x) + c$ für alle $x \in [a,b]$. Somit ist dank i)

$$G(b) - G(a) = F(b) - c - F(a) + c = F(b) - F(a) = \int_a^b f(x)\,dx.$$

Dies vervollständigt den Beweis. $\square$

Man kann den Satz so umformulieren: Ist $g : [a,b] \to \mathbb{R}$ eine stetige, auf (a,b) differenzierbare Funktion mit stetiger Ableitung, so gilt für alle $c,d \in (a,b)$ mit $c < d$, dass

$$\int_c^d g'(x)\,dx = g(d) - g(c).$$

Übungsaufgabe 55. Es seien $a,b \in \mathbb{R}$, $a < b$. Betrachte eine Folge $(f_n)_{n\in\mathbb{N}} \subset C^1([a,b];X)$. Es sei $(f_n')_{n\in\mathbb{N}}$ gleichmäßig konvergent und es gebe $x_0 \in [a,b]$, für das die Folge $(f_n(x_0))_{n\in\mathbb{N}}$ konvergiert. Verwende den Hauptsatz, um einen einfachen Beweis des Satzes 1.6 zu liefern.

Definition 56. *Die Menge der Stammfunktionen einer gegebener, integrierbarer Funktion f bezeichnet man mit*

$$\int f(x)\,dx \qquad \textit{(oder } \int f(y)dy,\ \int f(s)ds\ \int f(t)dt\ldots\textit{)}$$

und nennt man unbestimmtes Integral von f.

Anmerkung 57. Es seien $f,g : [a,b] \to \mathbb{R}$ integrierbar. Gilt $f(x) \neq g(x)$ in höchstens abzählbar vielen Punkten, so haben f,g das selbe unbestimmte Integral. Zum Beweis s. [**5**, S. 201]. Insbesondere folgt, dass das bestimmte Integral zwischen a und b einer Funktion f bereits eindeutig bestimmt ist, falls f auf (a,b) stetig und beschränkt ist.

Für eine gegebene integrierbare Funktion f stimmen laut Satz 2.14 alle Elemente der Menge $\int f(x)\,dx$ bis auf eine Konstante überein; man schreibt also z.B. salopp

$$\int \exp(x)\,dx = \exp + c$$

statt

$$\int \exp(x)\,dx = \{f : \mathbb{R} \to \mathbb{R} : f(x) = \exp(x) + c \text{ für alle } x \in \mathbb{R} \text{ und ein } c \in \mathbb{R}\}.$$

Satz 2.15. *Es seien $a,b \in \mathbb{R}$. Es seien $f,g : [a,b] \to \mathbb{R}$ zwei stetig differenzierbaren Funktionen. Dann gilt*

$$\int_a^b f'(x)g(x)\,dx = \Big[(fg)(x)\Big]_a^b - \int_a^b f(x)g'(x)\,dx.$$

Dabei spricht man von *partieller Integration.*

Beweis. Die Rechenregel zeigen, dass die Funktion $f \cdot g$ auf (a,b) differenzierbar und es gilt $(fg)' = f'g + fg'$, d.h.

$$f'(x)g(x) = (fg)'(x) - fg'(x) \tag{2.6}$$

für alle $x \in (a,b)$. Weil f', g' und $(fg)'$ nach Voraussetzung stetig sind kann man (2.6) zwischen a und b integrieren und

$$\int_a^b f'(x)g(x)\,dx = \int_a^b (fg)'(x)\,dx - \int_a^b f(x)g'(x)\,dx.$$

Die Funktion fg ist aber eine Stammfunktion von $(fg)'$ und nach dem Hauptsatz der Differential- und Integralrechnung gilt

$$\int_a^b (fg)'(x)\,dx = \Big[(fg)(x)\Big]_a^b.$$

Somit folgt die Aussage. □

Beispiel 58. Um das Integral

$$\int_0^{\frac{\pi}{2}} x\cos(x)\,dx$$

zu berechnen, interpretiere die Funktion unter dem Integralzeichen wie das Produkt zweier Funktionen f, g, wobei f die Sinus- und g die identische Funktion ist. Dann folgt aus Satz 2.15, dass

$$\int_0^{\frac{\pi}{2}} x\cos(x)\,dx = \Big[(fg)(x)\Big]_0^{\frac{\pi}{2}} - \int_0^{\frac{\pi}{2}} f(x)g'(x)\,dx = \frac{\pi}{2}\sin\frac{\pi}{2} - \int_0^{\frac{\pi}{2}} \sin(x)\,dx = \frac{\pi}{2} + [\cos]_0^{\frac{\pi}{2}} = \frac{\pi}{2} - 1.$$

□

Beispiel 59. Um die Logarithmusfunktion zu integrieren, stelle log als $f \cdot g'$ dar, wobei $f : x \mapsto \log(x)$ und $g : x \mapsto x$, sodass $g'(x) = 1$ für alle x sind. Somit gilt für alle a, b mit $0 < a < b$

$$\begin{aligned}
\int_a^b \log(x)\,dx &= \int_a^b f(x)g'(x)\,dx \\
&= \Big[(fg)(x)\Big]_a^b - \int_a^b f'(x)g(x)\,dx \\
&= b\log(b) - a\log(a) - \int_a^b \frac{x}{x}\,dx \\
&= b\log(b) - a\log(a) - (b-a) \\
&= b(\log(b) - 1) - a(\log(a) - 1).
\end{aligned}$$

Also ist $x \mapsto x(\log(x) - x)$ eine Stammfunktion der Logarithmusfunktion. □

Anmerkung 60. Es sei $f : [a,b] \to \mathbb{R}$ eine stetige, auf (a,b) differenzierbare Funktion, sodass f' auf (a,b) stetig ist. Es folt aus Satz 2.15, dass

$$\int_a^b f'(x)f(x)\,dx = \frac{(f(b))^2 - (f(a))^2}{2}.$$

Die obige Regel der partiellen Integration ist eine Folgerung der Produktregel der Differenzialrechnung. Genauso folgt sie sogenannte *Substitutionsregel* aus der Kettenregel.

Satz 2.16. *Es seien $a, b, c, d \in \mathbb{R}$ mit $a < b$ und $c < d$. Es seien $f : (a,b) \to \mathbb{R}$ und $g : (c,d) \to \mathbb{K}$ stetig differenzierbare Funktionen mit $f(x) \in (c,d)$ für alle $x \in (a,b)$, sodass die Verkettung $j \circ f$:*

$(a,b) \to \mathbb{K}$ wohldefiniert und stetig differenzierbar ist. Dann gilt: Ist J eine Stammfunktion von j, so ist $J \circ f$ eine Stammfunktion von $(j \circ f) \cdot f'$, und es gilt

$$\int_a^b j(f(x))f'(x)\, dx = \int_{f(a)}^{f(b)} j(u)du = \Big[J(u)\Big]_{u=f(a)}^{u=f(b)}.$$

Als (formale) Eselsbrücke bietet sich an zu beachten, dass für $u := f(x)$ dann

$$\frac{du}{dx} = f'(x)$$

gilt, also

$$du = f'(x)dx.$$

Beweis. Substituiere x durch u, d.h., setze $u := f(x)$. Laut den Rechenregeln für die Ableitungen und dem Hauptsatz der Differential- und Integralrechnung ist die Funktion $x \mapsto J \circ f(x) = J(f(x)) = J(u)$ auf (a,b) differenzierbar und es gilt $(J \circ f)' = (J' \circ f)f' = (j \circ f)f'$. Durch Integration zwischen a und b folgt

$$\int_a^b j(f(x))f'(x)\, dx = \int_a^b (J \circ f)'(x)\, dx = \Big[J(f(x))\Big]_a^b = \Big[J(u)\Big]_{u=f(a)}^{u=f(b)} = \int_{f(a)}^{f(b)} J'(u)du = \int_{f(a)}^{f(b)} j(u)du$$

Dies liefert die Aussage. □

Als direkte Folgerung der Substitutionsregel ergibt sich der folgende Fall.

Korollar 2.17. *Es sei $f : [a,b] \to \mathbb{R}$ eine stetige, auf (a,b) differenzierbare Funktion mit $f(x) > 0$ für alle $x \in [a,b]$. Dann ist $\log \circ f$ eine Stammfunktion von $\frac{f'}{f}$.*

Beweis. Wende die Substitutionsregel mit $j : y \mapsto \frac{1}{y}$ an. □

Übungsaufgabe 61. Verallgemeinere die Regel der partiellen Integration und die Substitutionsregel zum Fall zweier stetig differenzierbare Funktionen $f : [a,b] \to \mathbb{K}$ und $g : [c,d] \to X$, wobei X ein Banachraum ist.

Beispiel 62. Um das bestimmte Integral

$$\int_0^{\frac{\pi}{2}} x \sin(x^2)\, dx$$

zu berechnen, setze $f : x \mapsto x^2$ und $j := \sin$. Dann kann man die Substitutionsregel unmittelbar anwenden, also ist $(-\cos) \circ f$ eine Stammfunktion von $((-\cos) \circ f) \circ f'$. Somit ist

$$\int_0^{\frac{\pi}{2}} x \sin(x^2)\, dx = \frac{1}{2}\Big[-\cos(x^2)\Big]_0^{\frac{\pi}{2}} = \frac{1}{2}\left(1 - \cos\left(\frac{\pi^2}{4}\right)\right).$$

□

Beispiel 63. Man findet für jede Regelfunktion $j : (a,b) \to \mathbb{R}$ mittels der Substitution $f(x) := x+c$, dass

$$\int_a^b j(t+c)dt = \int_{a+c}^{b+c} j(u)du,$$

und ähnlich mittels $f(x) := cx$, dass

$$\int_a^b j(ct)dt = \frac{1}{c}\int_{ac}^{bc} j(u)du,$$

sowie schließlich mittels $f(x) := x^2$, dass

$$\int_a^b tj(ct)dt = \frac{1}{2}\int_{a^2}^{b^2} j(u)du,$$

$\square$

Beispiel 64. Es seien $a, b \in \mathbb{R}$, $a \neq 0$. Um

$$\int_0^1 \cos(ax+b)dx$$

zu berechnen, führe die Substitution $u := ax + b$ ein. Daher kann man die Substitutionsregel mit

$$j := \cos, \qquad J := \sin, \qquad f(x) := ax+b, \qquad f'(x) = a,$$

anwenden. Man findet somit

$$\int_0^1 \cos(ax+b)dx = \frac{1}{a}\int_0^1 \cos(ax+b)adx = \frac{1}{a}\int_0^1 j(f(x))f'(x)dx = \frac{1}{a}\Big[J(u)\Big]_{u=a\cdot 0+b}^{u=a\cdot 1+b} = \frac{\sin(a+b)-\sin(b)}{a}.$$

$\square$

Interessanter ist der Fall, dass die zu integrierende Funktion (noch) nicht in Form einer Verkettung zweier Funktionen auftritt. Manchmal ist es jedoch günstig, eine solche Darstellung als Verkettung *künstlich* einzuführen.

Beispiel 65. Wir wollen die Theorie der Integration dazu verwenden, um zu überprüfen, dass der Flächeninhalt des Einheitskreises π ist. Das ist natürlich zur Aussage äquivalent, dass der Flächeninhalt der obere hälfte des Einheitskreises $\frac{\pi}{2}$ ist – anders gesagt: Der Flächeninhalt unterhalb dem Graphen der Funktion $f : x \mapsto 1 - x^2$ und oberhalb der x-Achse ist $\frac{\pi}{2}$ (hier benutzen wir, dass der Einheitskreis die Menge aller Punkten $(x, y) \in \mathbb{R}^2$ ist, die $x^2 + y^2 = 1$ erfüllen). Berechne also

$$\frac{1}{2}\int_{-1}^1 \sqrt{1-x^2}\, dx.$$

Führe die Substitution $x = \sin(y)$ durch, sodass $y = \arcsin(x)$ und $\sqrt{1-x^2} = \cos(y)$. Dann ist

$$\begin{aligned}
\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{1-\sin(y)^2}\cos(y)dy &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{\cos(y)^2}\cos(y)dy \\
&= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(y)^2 dy \\
&= \frac{1}{2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (1+\cos(2y))dy \\
&= \frac{1}{2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dy + \frac{1}{2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(2y)dy \\
&= \frac{\pi}{2} + \frac{1}{4}\Big[\sin(2y)\Big]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \\
&= \frac{\pi}{2}.
\end{aligned}$$

□

Übungsaufgabe 66. Ähnlich wollen wir den Flächeninhalt innerhabl der Hyperbel mit Koordinaten

$$x^2 - y^2 = 1.$$

Wie vorhin berechne aus Symmetriegründen nur das Integral

$$2\int_1^x \sqrt{y^2-1}dy.$$

Beispiel 67. Zur Erinnerung: die *Tangens-Funktion* ist definiert durch

$$\tan := \frac{\sin}{\cos}.$$

Da sie (direckte Berechnung!) zwischen $-\frac{\pi}{2}$ und $\frac{\pi}{2}$ streng monoton wachsend ist, ist sie dort invertierbar – ihre Inverse bezeichnet man arctan. Man zeigt leicht, dass

$$\arctan' x = \frac{1}{1+x^2}, \qquad x \in \mathbb{R}.$$

Um eine Stammfunktion von arctan zu finden, d.h., um

$$\int \arctan x dx$$

zu bestimmen, verwendet man die partielle Integration, so dass

$$\begin{aligned}
\int \arctan x dx &= x\text{arctanx} - \int x\arctan' x dx \\
&= x\text{arctanx} - \int \frac{x}{1+x^2}dx.
\end{aligned}$$

Mithilfe vom Korollar 2.17 sieht man nun, dass

$$\int \arctan x dx = x\text{arctanx} - \frac{1}{2}\log(1+x^2).$$

□

Übungsaufgabe 68. Bestimme
(1) $\int \exp(ax)\, dx$;
(2) $\int a^x\, dx$, wobei $a > 0$;
(3) $\int \cos(ax)\, dx$.

Übungsaufgabe 69. Zeige, dass
(1) $\int \frac{dx}{\sqrt{1+x^2}} = \log(x + \sqrt{1+x^2})$;
(2) $\int \sinh(x) = \cosh(x)$;
Bestimme jeweils ein Intervall, auf dem die Formel gültig ist.

Übungsaufgabe 70. Bestimme
(1) $\int_a^b \frac{\log(x)}{x}\, dx$;
(2) $\int_a^b x^\alpha \log(x)\, dx$ für $\alpha \in \mathbb{R}$, $\alpha \neq -1$, wobei $0 < a < b$.

Beispiel 71. Betrachte die Funktion $\tan : (\frac{\pi}{6}, \frac{\pi}{3}) \ni x \mapsto \frac{\sin(x)}{\cos(x)} \in \mathbb{R}$. Dann ist die Funktion überall strikt positiv. Um eine Stammfunktion zu bestimmen merke, dass $\cos' = -\sin$. Setze $f := \cos$. Somit ist laut Satz 2.17 $-\log \circ \cos$ eine Stammfunktion der tan, also ein Element von

$$\int \tan(x)\, dx = -\int \frac{f'(x)}{f(x)}\, dx.$$

□

Die partielle Integration und die Substitutionsregel haben wir nur für skalarwertige Funktionen definiert. Doch ist es manchmal anhand des folgenden Resultats möglich, sie auch in allgemeineren Fälle zu verwenden.

Satz 2.18. *Es seien $a, b \in \mathbb{R}$, $a < b$, und $f : [a, b] \to \mathbb{K}^d$, $d \in \mathbb{N}$. Genau dann ist f eine Regelfunktion, wenn alle ihre Komponenten $f^j : [a, b] \to \mathbb{K}$, $j = 1, \ldots, d$, das sind. In diesem Fall gilt*

$$\int_a^b f(x)dx = \left(\int_a^b f^1(x)dx, \ldots, \int_a^b f^d(x)dx\right).$$

Beweis. Die Funktion f ist genau dann eine Regelfunktion, wenn sie gleichmäßiger Grenzwert einer Folge von Treppenfunktionen $(\phi_n)_{n\in\mathbb{N}}$ ist. Da aber eine vektorwertige Funktion offensichtlich genau dann eine Treppenfunktion, wenn alle ihre Komponenten das sind, sieht man gleich, dass genau dann $f = \lim_{n\to\infty} \phi_n$, wenn $f^j = \lim_{n\to\infty} \phi_n^j$ für alle $j = 1, \ldots, d$ (alle Grenzwerte in gleichmäßigen Sinne). Daher folgt die Aussage, da die behauptete Identität für Treppenfunktionen sicherlich gilt. □

Beispiel 72. Wir wollen das Integral der stetigen Funktion $f : [0, 2\pi] \ni x \mapsto \exp(x + ix) \in \mathbb{C}$ bestimmen. Aufgrund der Eulerschen Identität folgt, dass

$$\int_0^{2\pi} f(x)\, dx = \int_0^{2\pi} \exp(x)\cos(x)\, dx + i\int_0^{2\pi} \exp(x)\sin(x)\, dx$$

existiert. Durch mehrfache partielle Integration erhält man

$$\begin{aligned}
\int_0^{2\pi} \exp(x)\cos(x)\ dx &= \Big[\exp(x)\cos(x)\Big]_0^{2\pi} + \int_0^{2\pi} \exp(x)\sin(x)\ dx \\
&= \Big[\exp(x)\cos(x)\Big]_0^{2\pi} + \Big[\exp(x)\sin(x)\Big]_0^{2\pi} - \int_0^{2\pi} \exp(x)\cos(x)\ dx \\
&= \exp(2\pi)\cos(2\pi) - \exp(0)\cos(0) + \exp(2\pi)\sin(2\pi) - \exp(0)\sin(0) \\
&\qquad - \int_0^{2\pi} \exp(x)\cos(x)\ dx
\end{aligned}$$

und somit

$$\int_0^{2\pi} \exp(x)\cos(x)\ dx = \frac{\exp(2\pi)-1}{2}.$$

ähnlich gilt

$$\begin{aligned}
\int_0^{2\pi} \exp(x)\sin(x)\ dx &= \Big[\exp(x)\sin(x)\Big]_0^{2\pi} - \int_0^{2\pi} \exp(x)\cos(x)\ dx \\
&= \Big[\exp(x)\sin(x)\Big]_0^{2\pi} - \Big[\exp(x)\cos(x)\Big]_0^{2\pi} - \int_0^{2\pi} \exp(x)\sin(x)\ dx
\end{aligned}$$

also

$$\int_0^{2\pi} \exp(x)\sin(x)\ dx = \frac{-\exp(2\pi)+1}{2}.$$

Insgesamt erhält man

$$\int_0^{2\pi} f(x)\ dx = \frac{\exp(2\pi)-1}{2} + i\frac{1-\exp(2\pi)}{2}.$$

□

Da man $z = x + iy \in \mathbb{C}$ mit $(x, y) \in \mathbb{R}^2$ identifizieren kann, gilt das folgende Resultat unmittelbar.

KOROLLAR 2.19. *Es seien $a, b \in \mathbb{R}$ mit $a < b$. Genau dann ist ein $f : [a, b] \to \mathbb{C}$ eine Regelfunktion, wenn sowohl ihr Real- als auch ihr Imaginärteil (d.h., $\mathrm{Re}\, f : [a, b] \to \mathbb{R}$ und $\mathrm{Im}\, f : [a, b] \to \mathbb{R}$) das sind. In diesem Fall gilt*

$$\int_a^b f(x)dx = \int_a^b \mathrm{Re} f(x)dx + i\int_a^b \mathrm{Im} f(x)dx.$$

BEISPIEL 73. Berechne

$$\int_1^2 \frac{1}{4x+1}\ dx.$$

Definiere

$$u \equiv f(x) := 4x + 1, \qquad j(y) := \frac{1}{y}$$

sodass die Integrationsextrema $u = 5$ bzw. $u = 9$ (falls $x = 1$ bzw. $x = 2$) werden. Weil $f'(x) = 4$ für alle $x \in \mathbb{R}$ kann man das Integral auch als

$$\int_5^9 \frac{1}{u} \cdot \frac{1}{4} du$$

darstellen. Für dieses Integral gilt offensichtlich

$$\int_5^9 \frac{1}{4u} du = \frac{1}{4} \Big[\log(u) \Big]_5^9 = \frac{\log 9 - \log 5}{4}.$$

□

Was, wenn die zu integrierende Funktion zwar rational wie im vorigen Beispiel ist, aber einer komplizieretern Gestalt? Um diesen Abschnitt zu beenden erwähnen wir den folgenden wichtigen Satz, der aber eher eine algebraische Aussage enthält und dessen Beweis deshalb unseren Rahmen sprengen würde (s. z.B. [**2**, Sätze I.8.15 und IV.5.8]).

SATZ 2.20. *Es seien P, Q polynome vom Grad m bzw. n. Dann gelten folgende Aussagen.*

- *Es gibt eindeutig bestimmte Polynome S, T, für die*
$$\frac{P}{Q} = S + \frac{T}{Q}$$
gilt.
- *Es sei nun $m < n$ und es sei 1 der höchste Koeffizient von Q. Dann gibt es eindeutig bestimmten $a_{jk} \in \mathbb{C}$ so, dass*
$$\frac{P(x)}{Q(x)} = \sum_{j=1}^{N} \sum_{h=1}^{n_j} \frac{a_{jh}}{(x - z_j)^k},$$
wobei N die Anzahl der verschiedenen Nullstellen $z_1, \dots, z_N$ von Q bezeichnet und $n_1, \dots, n_N$ deren Vielfachkeiten, so dass insbesondere $n_1 + \ldots + n_N = n$. Insbesondere gilt, falls alle Nullstellen nicht *im Integrationsintervall (a, b) liegen, dass*
$$\int_a^b \frac{P(x)}{Q(x)} dx = \sum_{j=1}^{N} \sum_{h=1}^{n_j} a_{jh} \int_a^b \frac{1}{(x - z_j)^k} dx.$$
- *Sind alle Nullstellen von Q einfach, so kann man die obige Darstellung vereinfachen. Es gilt nämlich*
$$\frac{P(x)}{Q(x)} = \sum_{j=1}^{N} \frac{P(z_j}{Q'(z_j)} \frac{1}{(x - z_j)},$$
und insbesondere
$$\int_a^b \frac{P(x)}{Q(x)} dx = \sum_{j=1}^{N} \frac{P(z_j}{Q'(z_j)} \log |x - z_j|,$$
falls alle Nullstellen nicht *im Integrationsintervall (a, b) liegen.*

BEISPIEL 74. Betrachte das Polynom

$$Q(x) := x^2 + 2\alpha x + \beta.$$

Gilt $\Delta := \alpha^2 - \beta > 0$, so hat Q zwei verschiedenen reelle Nullstellen $z_1 := -a + \sqrt{\Delta}$ und $z_2 := -a - \sqrt{\Delta}$. Eine Zerlegung der rationalen Funktion $1/Q$ wie im obigen Satz ist möglich mit Wahl von $a_{11} := \frac{1}{2}\sqrt{\Delta}$ und $a_{21} := -\frac{1}{2}\sqrt{\Delta}$, d.h.,

$$\frac{1}{Q(x)} = \frac{a_{11}}{x - z_1} + \frac{a_{21}}{x - z_2}.$$

Liegen z_1, z_2 *nicht* im Integrationsintervall (a,b), so findet man

$$\int_a^b \frac{1}{Q(x)}dx = \int_a^b \left(\frac{a_{11}}{x-z_1} + \frac{a_{21}}{x-z_2}\right) dx = \Big[a_{11}\log|x-z_1| + a_{21}\log|x-z_2|\Big]_{x=a}^{x=b} = \frac{1}{2\sqrt{\Delta}}\left[\log\frac{|x-z_1|}{|x-z_2|}\right]_{x=a}^{x=b}.$$

□

2.1. Uneigentliche Integrale

Schließlich möchten wir den Integrationsbegriff auch für möglicherweise unbeschränkte Funktionen auf möglichereweise unbeschränkten Integrationsintervallen definieren.

Definition 75. *Es seien $a, b \in \mathbb{R} \cup \{\pm\infty\}$, $a < b$, und $J \subset \mathbb{R}$ ein Intervall mit $a = \inf J$ und $b = \sup J$. Es seien X ein Banachraum und $f : J \to X$ eine Funktion*

- *deren Einschränkung auf jedes kompakte Teilintervall $[\alpha, \beta] \subset J$ eine Regelfunktion ist,*
- *für die ein $c \in (a,b)$ existiert, so dass beide Grenzwerte*

$$\lim_{\zeta \to a+} \int_\zeta^c f(x)dx \qquad \textit{und} \qquad \lim_{\eta \to b-} \int_c^\eta f(x)dx$$

existieren (bzgl. der Norm von X!).

So heißt f auf J uneigentlich integrierbar *und*

$$\int_J f(x)dx := \int_a^c f(x)\, dx + \int_c^b f(x)\, dx$$

das uneigentliche Integral *von f auf J, wobei*

$$\int_a^c f(x)\, dx := \lim_{\zeta \to a+} \int_\zeta^c f(x)dx \quad \textit{sowie} \quad \int_c^b f(x)\, dx := \lim_{\eta \to b-} \int_c^\eta f(x)dx.$$

Da man auch $a, b \in \mathbb{R}$ wählen darf, ist diese eine strikte Verallgemeinerung des Begriffes einer Regelfunktion. Beachte, dass mit f erfüllt auch ihre Normfunktion $\|f\|_X : J \to \mathbb{R}$ die erste Bedingung in der Definition.

Dass die obige Definition, und insbesondere die nicht von c abhängige Notation des uneigentlichen Integrals konsistent ist, zeigt das folgende Resultat.

Lemma 2.21. *Es seien $J \subset \mathbb{R}$ ein Intervall mit $a := \inf J$ und $b := \sup J$ und X ein Banachraum.*
(1) Ist f eine uneigentlich integrierbare Funktion, so existieren gleich für alle *$c \in (a,b)$ die Grenzwerte*

$$\lim_{\zeta \to a+} \int_\zeta^c f(x)dx \qquad \textit{und} \qquad \lim_{\eta \to b-} \int_c^\eta f(x)dx.$$

(2) Darüber hinaus sind je zwei Paare solcher Grenzwerte konsistent, denn für je zwei $c, c' \in (a,b)$ gilt

$$\lim_{\zeta \to a+} \int_\zeta^c f(x)dx + \lim_{\eta \to b-} \int_c^\eta f(x)dx = \lim_{\zeta \to a+} \int_\zeta^{c'} f(x)dx + \lim_{\eta \to b-} \int_{c'}^\eta f(x)dx.$$

Beweis. Führe für $c \in (a,b)$ die Notationen

$$I_{ac} := \lim_{\zeta \to a+} \int_\zeta^c f(x)dx, \qquad I_{cb} := \lim_{\eta \to b-} \int_c^\eta f(x)dx$$

ein – nach Voraussetzung gibt es mindestens ein $c \in (a,b)$, für das sowohl I_{ac} als auch I_{cb} existieren.

(1) Es folgt dann aus dem Satz 2.7.(1), dass für alle weitere $c', \zeta \in (a,b)$ auch

$$\int_{\zeta}^{c'} f(x)dx = \int_{\zeta}^{c} f(x)dx + \int_{c}^{c'} f(x)dx$$

gilt. Da $\int_c^{c'} f(x)dx$ ein wohl definierter Vektor in X ist, existiert mit I_{ac} jedes weitere uneigentliche Integral $I_{ac'}$ – dann gilt offensichtlich auch die Identität

$$I_{ac'} = I_{ac} + \int_{c}^{c'} f(x)dx.$$

Ähnlich existiert mit I_{cb} jedes weitere uneigentliche Integral $I_{c'b}$ und die Identität

$$I_{c'b} = I_{cb} + \int_{c}^{c'} f(x)dx.$$

gilt. Dies zeigt die erste Aussage.

(2) Es gilt nun

$$\begin{aligned} \lim_{\zeta\to a+}\int_{\zeta}^{c} f(x)dx + \lim_{\eta\to b-}\int_{c}^{\eta} f(x)dx &= I_{ac} + I_{cb} \\ &= I_{ac'} + \int_{c'}^{c} f(x)dx + \int_{c}^{c'} f(x)dx + I_{c'b} \\ &= \lim_{\zeta\to a+}\int_{\zeta}^{c'} f(x)dx + \lim_{\eta\to b-}\int_{c'}^{\eta} f(x)dx. \end{aligned}$$

Dies vervollständigt den Beweis. □

Übungsaufgabe 76. Inbesondere: Es seien $a, b \in \mathbb{R}$, $a < b$, X ein Banachraum, und $f : [a,b] \to X$ eine Regelfunktion. Dann stimmt das uneigentliche Integral von f mit dem Riemannschen Integral von f überein. Liefere die Details.

Beispiel 77. Es sei $c > 0$. Dann ist die stetige Funktion $x \mapsto \exp(-cx)$ auf $J := [0,\infty)$ uneigentlich integrierbar. Denn eine Stammfunktion von $x \mapsto \exp(-cx)$ ist $x \mapsto \frac{\exp(-cx)}{c}$. Also gilt für $s > 0$

$$\int_0^s - \exp(-cx)\, dx = \left[\frac{\exp(-cx)}{c}\right]_0^s = \frac{1}{c}(1-\exp(-cs)).$$

Somit ist

$$\int_0^\infty - \exp(-cx)\, dx = \lim_{s\to\infty}\frac{1}{c}(1-\exp(-cs)) = \frac{1}{c}.$$

□

Beispiel 78. Es sei $a > 0$ und $s \in \mathbb{R}$. Dann ist

$$(a,\infty) \ni x \mapsto \frac{1}{x^s} \in \mathbb{R}$$

uneigentlich integrierbar, da diese Funktion stetig ist. Ist $s > 1$, so gilt für jedes $\eta > a$

$$\int_a^\eta \frac{1}{x^s}dx = \frac{1}{1-s}\left[x^{1-s}\right]_{x=a}^{x=\eta} = \frac{1}{1-s}\left(\frac{1}{\eta^{s-1}} - \frac{1}{a^{s-1}}\right),$$

und daher

$$\int_a^\infty \frac{1}{x^s} dx = \lim_{\eta \to +\infty} \frac{1}{1-s} \left(\frac{1}{\eta^{s-1}} - \frac{1}{a^{s-1}} \right) = -\frac{1}{1-s} \frac{1}{a^{s-1}}.$$

Ist jedoch $s = 1$, so gilt

$$\int_a^\eta \frac{1}{x} dx = \Big[\log x \Big]_{x=a}^{x=\eta} = (\log \eta - \log a) ,$$

und daher

$$\int_a^\infty \frac{1}{x} dx = \lim_{\eta \to +\infty} (\log \eta - \log a) = +\infty.$$

$\square$

BEISPIEL 79. Die stetige Funktion $x \mapsto x^{-\frac{1}{2}}$ ist auf dem Intervall $[0,1]$ uneigentlich integrierbar. Dies kann man sehen, indem man die Substitution $f(x) := \frac{1}{x}$ durchführt und dann das Beispiel 78 anwendet (Übungsaufgabe!). Oder auch direkt, wie folgt. Da die Funktion stetig ist, soll man nur beachten, dass für jedes $\zeta \in (0,1)$

$$\int_\zeta^1 x^{-\frac{1}{2}} dx = \Big[2x^{\frac{1}{2}} \Big]_{x=\zeta}^{x=1} = 2 - 2\sqrt{\zeta},$$

und daher

$$\int_0^1 x^{-\frac{1}{2}} dx = \lim_{\zeta \to 0+} 2 - 2\sqrt{\zeta} = 2.$$

$\square$

BEISPIEL 80. Es gilt

$$\begin{aligned}
\int_{-1}^1 \frac{1}{\sqrt{1-x^2}} dx &= \lim_{\zeta \to -1+} \int_\zeta^0 \frac{1}{\sqrt{1-x^2}} dx + \lim_{\eta \to 1-} \int_0^\eta \frac{1}{\sqrt{1-x^2}} dx \\
&= -\lim_{\zeta \to -1+} \arcsin\zeta + \lim_{\eta \to 1-} \arcsin\eta \\
&= -\left(-\frac{\pi}{2}\right) + \frac{\pi}{2}.
\end{aligned}$$

$\square$

KAPITEL 3

Einige Anwendungen der Integralrechnung

3.1. Konvergenz von Reihen

SATZ 3.1. *Es sei* $f : [1, +\infty) \to [0, \infty)$ *eine monoton fallende Funktion. Genau dann konvergiert das uneigentliche Integral*

$$\int_1^\infty f(x)dx,$$

wenn die Reihe

$$\sum_{n=1}^\infty f(n),$$

es tut.

BEWEIS. Wir führen zwei Treppenfunktionen ω (wie *oberhalb*) und υ (wie *unterhalb*) von $(1, \infty)$ nach $\mathbb{R}$ ein durch

$$\omega(x) := f(n) \qquad \text{falls } x \in [n, n+1),$$

bzw.

$$\upsilon(x) := f(n+1) \qquad \text{falls } x \in [n, n+1).$$

Aufgrund der Monotonie von f gilt für alle x

$$f(\lfloor x \rfloor) \geq f(x) \geq f(\lfloor x \rfloor + 1),$$

d.h.,

$$\omega(x) \geq f(x) \geq \upsilon(x) \qquad \text{für alle } x \in [1, \infty).$$

Somit gilt nach Definition von υ und ω

$$\sum_{n=2}^N f(n) = \int_1^N \upsilon(x)dx \leq \int_1^N f(x)dx \leq \int_1^N \omega(x)dx = \sum_{n=1}^{N-1} f(n) \qquad \text{für alle } N \in \mathbb{N}.$$

Nun sind wir in der Lage, die Aussage zu beweisen. Denn ist das uneigentliche Integral konvergent, so ist die Folge $(\sum_{n=2}^N f(n))_{N\in\mathbb{N}}$ beschränkt (eben durch $\int_1^\infty f(x)dx$) und monoton wachsend (da f nur positive Werte annimmt). Somit ist die Folge der Partialsummen konvergent, d.h., $\sum_{n=1}^\infty f(n)$ konvergiert.

Es sei umgekehrt die Reihe konvergent, und deshalb insbesondere auch ω uneigentlich integrierbar. So ist auch die Familie $\left(\int_1^y f(x)dx\right)_{y>1}$ monoton wachsend und durch $\int_1^\infty \omega(x)dx$ beschränkt, und daher ihre Konvergenz – d.h., die uneigentliche Integrierbarkeit von f. $\square$

3.2. Taylor-Entwicklung einer Stammfunktionenreihe

LEMMA 3.2. *Es seien $a, b \in \mathbb{R}$, $a < b$, und $x_0 \in (a,b)$. Es seien $f, g : [a,b] \to X$ differenzierbare Funktionen, X ein Banachraum. Gilt $f'(x) = g'(x)$ für alle $x \in [a,b]$ und $f(x_0) = g(x_0)$, so gilt*

$$f(x) = g(x) \qquad \text{für alle } x \in [a,b].$$

BEWEIS. Es folgt aus dem Hauptsatz, dass

$$f(x) = f(x_0) + \int_{x_0}^{x} f'(s)ds = g(x_0) + \int_{x_0}^{x} g'(s)ds = g(x).$$

□

SATZ 3.3. *Es seien $a, b \in \mathbb{R}$, $a < b$, und $f : [a,b] \to \mathbb{R}$ eine unendlich oft stetig differenzierbare Funktion und $x_0 \in (a,b)$. Konvergiert die von der Taylorreihe T^{f,x_0} abgeleitete Potenzreihe*

$$T^{f,x_0} = \sum_{k=1}^{\infty} \frac{f^{(k)}(x_0)}{(k-1)!}(x-x_0)^{k-1}, \qquad x \in (a,b),$$

punktweise gegen f', so ist $T^{f,\cdot}$ (als Reihe von Funktionen von (a,b) nach $\mathbb{R}$ aufgefasst) gegen f normkonvergent.

BEWEIS. Es seien $c, d \in \mathbb{R}$ mit $x_0 \in [c,d] \subset (a,b)$. Dann ist nach dem Korollar 1.11 die von T^f abgeleitet Reihe auf $[c,d]$ normkonvergent. Nun folgt aus dem Satz 1.6, dass die Taylorreihe $T^{f,\cdot}$ auf $[c,d]$ konvergiert (da sie bereits in x_0 konvergiert), und ihr Grenzwert ergibt eine Funktion $g : (a,b) \to \mathbb{R}$, welche

$$g(x_0) = f(x_0) \qquad \text{sowie} g'(x) = f'(x) \quad \text{für alle } x \in (a,b),$$

erfüllt. Somit folgt aus dem Lemma 3.2, dass $f \equiv g$, was die Aussage vollständig beweist. □

ÜBUNGSAUFGABE 81. Zeige, dass für die Funktion $f := \arctan$ die Taylorreihe $T^{f,\cdot}$ auf $(-1,1)$ durch

$$x \mapsto x - \frac{x^3}{3} + \frac{x^5}{5} + \dots$$

gegeben wird, und dass sie gegen f normkonvergent ist.

ÜBUNGSAUFGABE 82. Zeige, dass für die Funktion $f := \log(1+x)$ die Taylorreihe $T^{f,\cdot}$ auf $[-1,1]$ durch

$$x \mapsto x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

gegeben wird, und dass sie gegen f normkonvergent ist.

3.3. Der Satz von Weierstraß

Eine interessante Anwendung der Integralrechnung besteht darin, eine explizite Konstruktion einer Polynomenfolge durchzuführen, welche eine gegebene aber beliebige stetige Funktion approximieren. Zum Beweis von diesem sogenannten *Satz von Stone–Weierstraß*, der tatsächlich in dieser einfacher Version 1885 von Karl Weierstraß bewiesen wurde, werden wir das folgende Lemma benötigen.

Lemma 3.4. *Es sei $f : \mathbb{R} \to \mathbb{R}$ eine stetige Funktion und Q ein Polynom. Dann ist auch die Funktion P, welche durch*

$$P(x) = \int_0^1 f(u)Q(u-x)du, \qquad x \in [0,1],$$

definiert ist, ein Polynom.

Beweis. Wir führen erst die folgende Notation an: Ist $f : [0,1] \to \mathbb{R}$ eine stetige Funktion, so dass f eine Stammfunktion

$$F(u) = \int_0^u f(s)ds, \qquad u \in [0,1],$$

hat, so kann man auch aufgrund der Stetigkeit von F eine Stammfunktion von F betrachten, also eine *zweite Stammfunktion von f*, welche wir mit $F^{(2)}$ bezeichnen werden, also

$$F^{(2)}(u) = \int_0^u \int_0^{u_1} f(s)ds\, du_1, \qquad u \in [0,1],$$

und so weit rekursiv die $(n+1)$-te Stammfunktion

$$F^{(n+1)}(u) = \underbrace{\int_0^u \int_0^{u_1} \dots \int_0^{u_n}}_{n+1 \text{ Integrale}} f(s)ds\, du_1\, \dots du_n, \qquad u \in [0,1],$$

wobei $F^{(1)} := F$. Beachte, dass $F^{(n)}(0) = 0$ für alle $n \in \mathbb{N}$. Die Funktion P können wir auch als

$$P(x) = \int_0^1 f(u)Q(u-x)du = \int_0^1 \frac{d}{du}F(u)Q(x-u)du, \qquad x \in [0,1],$$

darstellen. Partielle Integration liefert nun für alle $x \in [0,1]$

$$\begin{aligned} P(x) &= \int_0^1 \frac{d}{du}F(u)Q(x-u)du \\ &= \Big[F(u)Q(x-u)\Big]_{u=0}^{u=1} - \int_0^1 F(u)Q'(x-u)du \\ &= F(1)Q(x-1) - \int_0^1 F(u)Q'(x-u)du. \end{aligned}$$

Weiter sieht man, dass

$$\begin{aligned} P(x) &= F(1)Q(x-1) - \int_0^1 \frac{d}{du}F^{(2)}(u)Q'(x-u)du \\ &= F(1)Q(x-1) - F^{(2)}(1)Q'(x-1) + \int_0^1 F^{(3)}(u)Q''(x-u)du, \end{aligned}$$

und rekursiv

$$\begin{aligned} P(x) &= \sum_{k=0}^{n}(-1)^{k+1}F^{(k+1)}(1)Q^{(k)}(x-1)+(-1)^n\int_0^1 F^{(n+2)}(u)Q^{(n+1)}(x-u)du \\ &= \sum_{k=0}^{n}(-1)^{k+1}F^{(k+1)}(1)Q^{(k)}(x-1) \\ &= \sum_{k=0}^{n}a_kF^{(k+1)}(1)Q^{(k)}(x-1), \end{aligned}$$

wobei n das Grad vom Polynom Q ist, da die $(n+1)$-te Ableitung eines Polynoms n-ten Grades identisch verschwindet. Allgemeiner ist $Q^{(k)}(\cdot-1)$ – wie jede k-te Ableitung eines Polynoms n-ten Grades – selber ein Polynom $(n-k)$-ten Grades, was endlich die Aussage liefert, da $a_0,\dots,a_n\in\mathbb{R}$. □

THEOREM 3.5. *Es seien $a,b\in\mathbb{R}$ mit $a<b$ und $f:[a,b]\to\mathbb{R}$ stetig. Dann gibt es eine Folge $(P_n)_{n\in\mathbb{N}}$ von Polynomen, welche gleichmäßig gegen f konvergiert.*

BEWEIS. Man kann o.B.d.A. annehmen, dass $a=0$ und $b=1$ – sonst führe die Substitution

$$x\mapsto\frac{x-a}{b-a}$$

durch. Man wird den Satz nur für den Fall $f(0)=f(1)=0$ beweisen müssen, denn sonst kann man für eine allgemeine Funktion f die Abbildung

$$g:x\mapsto P(x)-P(0)-x(P(1)-P(0))$$

betrachten, welches wohl $g(0)=g(1)=0$ erfüllt: Ist nun g gleichmäßiger Grenzwert einer Polynomenfolge $(P_n)_{n\in\mathbb{N}}$, so ist

$$f=(f-g)+\lim_{n\to\infty}P_n=\lim_{n\to\infty}Q_n$$

mit

$$Q_n:=(f-g)+P_n,\qquad n\in\mathbb{N},$$

ebenfalls Grenzwert einer Polynomenfolge.

Mit diesen Voraussetzungen können wir jetzt die Funktion f auf $\mathbb{R}$ durch 0 außerhalb $[0,1]$ fortsetzen: die neue, erweiterte Funktion ist ebenfalls stetig, und wir werden sie mit einem leichten Notationsmissbrauch auch mit f bezeichnen.

Betrachte nun $(Q_n)_{n\in\mathbb{N}}$ definiert durch

$$Q_k(x):=c_k(1-x^2)^k,\qquad k\in\mathbb{N},\ x\in\mathbb{R},$$

wobei $c_k\in\mathbb{R}$ ist derart gewählt, dass

$$\int_{-1}^{1}Q_k(x)dx\equiv c_k\int_{-1}^{1}(1-x^2)^kdx\stackrel{!}{=}1, \tag{3.1}$$

d.h.,

$$c_k:=\frac{1}{\int_{-1}^{1}(1-x^2)^kdx},\qquad k\in\mathbb{N}. \tag{3.2}$$

Nun folgt aus der Bernoullischen Ungleichung (vgl. [**6**, Übungsaufgabe 1.61]), dass

$$\int_{-1}^{1}(1-x^2)^k dx = 2\int_0^1 (1-x^2)^k dx \geq 2\int_0^{\frac{1}{\sqrt{k}}}(1-x^2)^k dx \geq 2\int_0^{\frac{1}{\sqrt{k}}}(1-kx^2)dx = \frac{4}{3\sqrt{k}} > \frac{1}{\sqrt{k}}, \qquad k \in \mathbb{N}.$$

Es folgt aus dieser Ungleichung sowie aus (3.2), dass

$$c_k < \sqrt{k}, \qquad k \in \mathbb{N}.$$

Somit folgt für alle $\delta \in (0,1)$, dass

$$0 \leq Q_k(x) = c_k(1-x^2)^k \leq \sqrt{k}(1-\delta^2)^k \qquad \text{für alle } k \in \mathbb{N} \text{ und } |x| \in [\delta, 1]. \tag{3.3}$$

Daher konvergiert $(Q_k)_{k\in\mathbb{N}}$ gleichmäßig gegen 0 im Kreisring $\{x \in \mathbb{R} : \delta \leq |x| \leq 1\}$. Definiere nun eine neue Funktionenfolge $(P_k)_{k\in\mathbb{N}}$ durch

$$P_k(x) := \int_{-1}^{1} f(x+t)Q_k(t)dt, \qquad k \in \mathbb{N}, x \in [0,1]. \tag{3.4}$$

Für festes $x \in [0,1]$ findet man durch die Substitution $u := x + t$ die äquivalente Darstellung

$$P_k(x) = \int_{x-1}^{x+1} f(u)Q_k(u-x)du, \qquad k \in \mathbb{N}, x \in [0,1].$$

und da der Integrationsbereich $[x-1, x+1]$ das Intervall $[0,1]$ enthält, außer dem die Funktion f identisch verschwindet, kann man diesen Ausdruck weiter zu

$$P_k(x) = \int_0^1 f(u)Q_k(u-x)du, \qquad k \in \mathbb{N}, x \in [0,1].$$

vereinfachen. Das Lemma 3.4 liefert nun, dass $(P_k)_{k\in\mathbb{N}}$ ebenfalls eine Polynomenfolge ist.

Nun folgt von (3.2), dass

$$\int_{-1}^{1} f(x)Q_k(t)dt = f(x)\int_1^1 Q_k(t)dt = f(x), \qquad k \in \mathbb{N},\ x \in [0,1].$$

Beachte, dass aus (3.3) sowie (3.4) gilt

$$P_k(x) - f(x) = \int_{-1}^{1} \left(f(x+t) - f(x)\right) Q_k(t)dt, \qquad k \in \mathbb{N}, x \in [0,1]. \tag{3.5}$$

Nun können wir endlich die Aussage beweisen. Es sei dazu $\varepsilon > 0$ und bestimme $\delta > 0$ so, dass

$$|f(x) - f(y)| < \varepsilon \qquad \text{für alle } x, y \text{ so, dass } |x-y| < \delta \tag{3.6}$$

(dabei haben wir den Satz verwendet, dass jede stetige Funktion auf einem kompakten Intervall auch gleichmäßig stetig ist – vgl. Satz 6.7). Da f stetig auf einem kompakten Intervall ist, ist weiterhin f

beschränkt, d.h., $\|f\|_\infty < \infty$. Es gilt nun für alle $x \in [0,1]$

$$
\begin{aligned}
|P_k(x) - f(x)| &\overset{(3.5)}{\leq} \int_{-1}^{1} |\,(f(x+t) - f(x))\,Q_k(t)|dt \\
&\overset{(3.3)}{=} \int_{-1}^{1} |f(x+t) - f(x)|\,Q_k(t)dt \\
&= \int_{-1}^{-\delta} |f(x+t) - f(x)|\,Q_k(t)dt + \int_{-\delta}^{\delta} |f(x+t) - f(x)|\,Q_k(t)dt + \int_{\delta}^{1} |f(x+t) - f(x)|\,Q_k(t)dt \\
&\leq 2\|f\|_\infty \int_{-1}^{-\delta} Q_k(t)dt + \int_{-\delta}^{\delta} |f(x+t) - f(x)|\,Q_k(t)dt + 2\|f\|_\infty \int_{\delta}^{1} Q_k(t)dt \\
&\overset{(3.6)}{\leq} 2\|f\|_\infty \int_{-1}^{-\delta} Q_k(t)dt + \frac{\varepsilon}{2}\int_{-\delta}^{\delta} Q_k(t)dt + 2\|f\|_\infty \int_{\delta}^{1} Q_k(t)dt \\
&\overset{(3.3)}{\leq} 2\|f\|_\infty \sqrt{k}(1-\delta^2)^k \int_{-1}^{-\delta} dt + \frac{\varepsilon}{2}\int_{-\delta}^{\delta} Q_k(t)dt + 2\|f\|_\infty \sqrt{k}(1-\delta^2)^k \int_{\delta}^{1} dt \\
&\overset{(3.1)}{\leq} 4\|f\|_\infty \sqrt{k}(1-\delta^2)^k(1-\delta) + \frac{\varepsilon}{2}.
\end{aligned}
$$

Diese Abschätzung gilt gleichmäßig für $x \in [0,1]$. Nun konvergiert für $k \to \infty$ der erste Summand gegen 0, was wiederum zeigt, dass $\|f - P_k\|_\infty$ für $k \to \infty$ beliebig klein wird. Somit ist die Aussage bewiesen. □

Übungsaufgabe 83. Kann der obige Satz zum Fall einer vektorwertigen stetigen Funktion verallgemeinert werden? Warum?

3.4. Existenz von transzendenten Zahlen (optional)

In diesem Abschnitt werden wir zeigen, dass es überhaupt transzendente Zahlen gibt – dies hatten wir in [**6**, Anmerkung 5.34] lediglich behauptet, aber nicht bewiesen.

Dazu benötigen wir zuerst das Folgende.

Lemma 3.6. *Es seien $n \in \mathbb{N}$, $(a_1, \ldots, a_n) \in \mathbb{R}^n$, und P ein Polynom vom Grad n, etwa*

$$P : \mathbb{R} \ni z \mapsto \sum_{j=0}^{n} a_j z^j \in \mathbb{R}.$$

Betrachte das assoziierte Betragspolynom

$$\tilde{P} :\mapsto \mathbb{R} \ni z \mapsto \sum_{j=0}^{n} |a_j| z^j \in \mathbb{R}$$

und die Funktion

$$\Phi_P : [0,\infty) \ni t \mapsto \int_0^t e^{t-s} P(s)ds \in \mathbb{R}. \tag{3.7}$$

Dann gilt

$$\Phi_P(t) = e^t \sum_{j=0}^{n} P^{(j)}(0) - \sum_{j=0}^{n} P^{(j)}(t), \qquad t \in [0,\infty), \tag{3.8}$$

und zusätzlich

$$|\Phi_P(t)| \leq |t|e^{|t|}\tilde{P}(|s|), \qquad t \in [0,\infty).$$

Beachte, dass die Funktion Φ_P auf $[0,\infty)$ wohl definiert ist, d.h., das entsprechende Integral konvergiert (Übungsaufgabe!).

Beweis. Wir integrieren Φ_P partiell und erhalten

$$\Phi_P(t) = e^t \sum_{j=0}^{\infty} P^{(j)}(0) - \sum_{j=0}^{\infty} P^{(j)}(t) = e^t \sum_{j=0}^{n} P^{(j)}(0) - \sum_{j=0}^{n} P^{(j)}(t), \qquad t \in [0,\infty),$$

wobei n das grad von P ist. So sieht man, dass

$$|\Phi_P(t)| \leq \left| \int_0^t e^{t-s} P(s) ds \right| \leq |t| \max_{s\in[0,t]} |e^{t-s}| \max_{s\in[0,t]} |P(s)| \leq |t|e^{|t|}\tilde{P}(|t|).$$

□

Theorem 3.7. *Die Zahl e ist transzendent.*

Beweis. Wir nehmen an, e sei nicht transzendent, d.h., e sei die Nullstelle eines Polynoms mit rationalen koeffizienten – etwa von Q, wobei

$$Q(x) := \sum_{k=0}^{n} a_k x^k$$

für Koeffizienten $a_0, \ldots, a_n$. Man kann o.B.d.A. annehmen, dass $a_0 \neq 0$, da sonst bekäme man

$$0 = Q(e) = \sum_{k=1}^{n} a_k e^k = e \sum_{k=0}^{n-1} a_{k+1} e^k,$$

d.h., e wäre auch Nullstelle eines Polynoms niedrigeren Grades.

Wähle eine Primzahl p, welche strikt größer als $\max\{k, |a_0|\}$ ist, und betrachte eine neue Funktion f, die durch

$$P(x) := x^{p-1}(x-1)^p(x-2)^p \cdots (x-k)^p$$

definiert ist. Da f ein Polynom (vom Grad $p(k+1)-1$) ist, kann man die zugehörige Funktion Φ_P wie in (3.7) betrachten. Beachte nun, dass

$$\sum_{k=0}^{n} a_k \left(e^k \sum_{j=0}^{n} P^{(j)}(0) \right) = \left(\sum_{k=0}^{n} a_k e^k \right) \left(\sum_{j=0}^{n} P^{(j)}(0) \right) = Q(e) \left(\sum_{j=0}^{n} P^{(j)}(0) \right) = 0.$$

Somit gilt angesichts von (3.8)

$$\sum_{k=0}^{n} a_k \Phi_P(k) = -\sum_{k=0}^{n} \sum_{j=0}^{n} f^{(j)}(t).$$

□

KAPITEL 4

Differenzialrechnung von Funktionen mehrerer Variablen

Betrachte Funktionen $f : A \to X$, wobei $A \subset Y$ eine offene Menge und X, Y reelle Banachräume sind. Wir kennen schon die Differenzierbarkeit von f definieren, wenn $A \subset \mathbb{R}$. Es stellt sich heraus, dass der allgemeine Fall deutlich komplizierter ist.

Beispiel 84. Man sieht an Beispiel 21, dass die Wahl der Abhängigkeiten einer Funktion am Modell liegt. Wären z.B. die Anfangsgeschwindigkeit, der Abschusswinkel und ja sogar die Fallbeschleinigung nicht fest, so könnte man sich fragen, wie jetzt die Flugbahn von diesen mehreren Variablen abhängen würde. Die Antwort wird von einer Funktion

$$f : (0,\infty) \times \left(0, \frac{\pi}{2}\right) \times (0,\infty) \times (0,\infty) \ni (v,\alpha,g,t) \mapsto \begin{pmatrix} x(v,\alpha,g,t) \\ y(v,\alpha,g,t) \end{pmatrix}$$

gegeben, wobei die Funktionen x, y vierer Variablen durch

$$x : (v,\alpha,g,t) \mapsto vt\cos(\alpha) \quad \text{und} \quad y : (v,\alpha,g,t) \mapsto -\frac{gt^2}{2} + vt\sin(\alpha)$$

definiert sind.

Diese Funktionen x und y (und somit auch f) sind stetig, wie man schon anhand der Kenntnisse der Analysis 1 hätte beweisen können. Die Flugbahn einer Kugel ist aber eine sehr glatte Kurve in der Luft: ist unsere anschauliche Vorstellung der Differenzierbarkeit einer Funktion korrekt, so würde man erwarten, f ist – in einem noch zu präzisierenden Sinne – differenzierbar. $\square$

4.1. Beschränkte lineare Operatoren

Definition 85. *Eine Abbildung $T : Y \to X$ heißt* linear, *falls*

$$T(\alpha v + \beta w) = \alpha T(v) + \beta T(w) \qquad \text{für alle } v, w \in Y \text{ und alle } \alpha, \beta \in \mathbb{R}.$$

Man schreibt dann Tv statt $T(v)$ und spricht oft von einem linearen Operator. *Ein linearer Operator heißt* beschränkt, *und man schreibt*

$$T \in \mathcal{L}(Y, X),$$

falls

$$\|T\|_{\mathcal{L}(Y,X)} := \sup_{\|v\|_Y \le 1} \|Tv\|_X < \infty. \tag{4.1}$$

Ein Isomorphismus *ist ein invertierbarer linearer Operator. Gibt es zwischen zwei normierten Räumen ein Isomorphismus, so heißen die beiden Räume* Isomorph.

Insbesondere gilt

$$T0 = 0 \qquad \text{für jeden linearen Operator } T, \tag{4.2}$$

wobei das Argument von T das neutrale Element von Y und der Wert von T an dieser Stelle das neutrale Element von X ist.

Anmerkung 86. Beachte, dass ein linearer Operator genau dann beschränkt ist, wenn es ein $M > 0$ gibt, so dass

$$\|Tv\|_X \leq M\|f\|_Y \qquad \text{für alle } f \in Y.$$

Das Infimum aller solcher M ist genau $\|T\|$. Die Linearität von T liefert übrigens

$$\|T\|_{\mathcal{L}(Y,X)} = \sup_{\|v\|_Y=1} \|Tv\|_X. \tag{4.3}$$

Umgekehrt ist T genau dann unbeschränkt, wenn es für alle $n \in \mathbb{N}$ ein $f_n \in Y$ mit $\|f_n\|_Y = 1$ gibt, so dass

$$\|Tv_n\|_X \geq n.$$

Beispiel 87. Man hat in der Vorlesung zur linearen Algebra gelernt, dass jede lineare Abbildung T von $\mathbb{R}^n$ nach $\mathbb{R}^m$ sich als Matrixmultiplikation mit einer eindeutig bestimmten Matrix A_T darstellen läßt, welche durch

$$a_{ij} := \sum_{k=1}^{m} (Ae_i)_k \delta_j^k,$$

definiert wird, und die *Darstellungsmatrix von A* heißt. (Dabei bezeichnet δ_j^k das sog. *Kronecker-Delta*, das durch

$$\delta_j^k := \begin{cases} 1 & \text{falls } k = j, \\ 0 & \text{falls } k \neq j, \end{cases}$$

definiert wird). Wiederum definiert jede Matrix A einen linearen Operator T_A vermöge

$$T_A(x) := A_T \cdot x.$$

Somit werden oft den Operator T und die assoziierte Matrix A_T (bzw. die Matrix A und den assoziierten Operator T_A) identifiziert. □

Übungsaufgabe 88. Zeige: Für alle normierte Räume Y, X ist $\mathcal{L}(Y, X)$ ein Vektorraum und die Abbildung

$$\mathcal{L}(Y, X) \ni T \mapsto \|T\|_{\mathcal{L}(Y,X)} \in \mathbb{R}_+$$

definiert eine Norm – die sog. *Operatornorm*.

Anmerkung 89. Man kann ja sogar zeigen, dass $\mathcal{L}(Y, X)$ vollständig ist (und somit ein Banachraum), falls X das ist, s. [**2**, Thm. VII.1.1].

Anmerkung 90. Es sei $T \in \mathcal{L}(Y, X)$. Die Anmerkung 86 zeigt, dass $\|Tv\|_X \leq \|T\|_{\mathcal{L}(Y,X)}\|v\|_Y$ für alle $v \in Y$. Es sei Z ein weiterer normierte Vektorraum und $S \in \mathcal{L}(Z, Y)$. Man bezeichnet herkömmlich $TS := T \circ S$. Dann findet man

$$\|TSv\|_X \leq \|T\|_{\mathcal{L}(Y,X)}\|Sv\|_Y \leq \|T\|_{\mathcal{L}(Y,X)}\|S\|_{\mathcal{L}(Z,Y)}\|v\|_Z,$$

d.h., $TS \in \mathcal{L}(Z, X)$ und

$$\|TS\|_{\mathcal{L}(Z,X)} \leq \|T\|_{\mathcal{L}(X,Z)}\|S\|_{\mathcal{L}(Y,X)}.$$

Man sagt, dass die Operatornorm **submultiplikativ** ist. Insbesondere ist $\mathcal{L}(X) := \mathcal{L}(X, X)$ eine *Algebra* bzgl. der Verkettung, vgl. die Vorlesung zur linearen Algebra.

ÜBUNGSAUFGABE 91. Es sei $T : Y \to X$ ein linearer Operator. Zeige, dass die folgenden Aussagen äquivalent sind.

(i) $\sup_{v\neq 0} \frac{\|Tv\|_Y}{\|v\|_X} < \infty$.
(ii) T ist beschränkt,
(iii) T is stetig,
(iv) T ist Lipschitz-stetig,

(Hinweis zum Beweis von (a) $\Leftrightarrow$ (b): Zeige, dass $\|T\| = \sup_{v\neq 0} \frac{\|Tv\|_Y}{\|v\|_X}$*).*

ANMERKUNG 92. Genau dann zwei lineare Operatoren A, B von Y nach X übereinstimmen, wenn sie auf der Einheitsphäre von Y das tun, d.h.,

$$Av = Bv \qquad \text{für alle } v \in Y \text{ mit } \|v\|_Y = 1. \tag{4.4}$$

Denn $A0 = B0$ wegen der Linearität, und für alle andere $w \in Y$ gibt es $\alpha \in \mathbb{R}$, so dass $\|\alpha w\|_Y = 1$ und somit

$$Aw = A\left(\frac{1}{\alpha}\alpha w\right) = \frac{1}{\alpha}A\,(\alpha w) \overset{(4.4)}{=} \frac{1}{\alpha}B\,(\alpha w) = B\left(\frac{1}{\alpha}\alpha w\right) = Bw.$$

Somit ist bewiesen, dass $Aw = Bw$ für alle $w \in Y$.

4.2. Fréchet-Differenzierbarkeit

DEFINITION 93. *Es sei $x_0 \in A$. Eine Abbildung $f : A \to X$ heißt* an der Stelle x_0 Fréchet-differenzierbar*, oder einfach* differenzierbar*, falls es ein $A_{x_0} \in \mathcal{L}(Y, X)$ existiert, für den*

$$\lim_{x\to x_0} \frac{f(x) - f(x_0) - A_{x_0}(x - x_0)}{\|x - x_0\|_X} = 0 \qquad \textit{(bzgl. der Norm von } X\textit{!).} \tag{4.5}$$

Der lineare Operator A_{x_0}, der i.A. tatsächlich von x_0 abhängt, heißt dann Fréchet-Ableitung *– oder manchmal* Differenzial – von f an der Stelle x_0 *und wird herkömmlich mit*

$$Df(x_0), \qquad df(x_0), \qquad \textit{oder} \qquad f'(x_0). \tag{4.6}$$

bezeichnet.[1]

Ist f an jeder Stelle Fréchet-differenzierbar, so sagt man, dass f Fréchet-differenzierbar ist. Die Abbildung

$$Df : A \ni x_0 \mapsto A_{x_0} \equiv Df(x_0) \in \mathcal{L}(Y, X)$$

heißt dann Differenzial *oder* Fréchet-Ableitung von f*. Den Vektorraum der Fréchet-differenzierbarer Funktionen von A nach X, die zudem eine stetige Fréchet-Ableitung haben, d.h., $Df \in C(A; \mathcal{L}(Y, X))$, bezeichnet man $C^1(A; X)$.*

Einige grundlegenden Eigenschaften von Fréchet-differenzierbaren Funktionen sind im Folgenden zusammengefasst.

LEMMA 4.1. *Es seien f eine Funktion von A nach X und $x_0 \in A$.*

(1) Die folgenden Aussagen sind äquivalent.

(i) Die Funktion f ist an der Stelle x_0 Fréchet-differenzierbar.

[1] Die ersten zwei Notationen sind in der Literatur üblicher, wir werden aber meistens die dritte verwenden.

(ii) Es gibt $A_{x_0} \in \mathcal{L}(Y, X)$ und eine Abbildung $R_{x_0} : A \to X$, welche an der Stelle x_0 stetig ist und dort auch verschwindet, so dass

$$f(x) = f(x_0) + A_{x_0}(x - x_0) + R_{x_0}(x) \qquad \textit{für alle } x \in A.$$

Anders ausgedruckt: Es gibt $A_{x_0} \in \mathcal{L}(Y, X)$ mit

$$f(x) = f(x_0) + A_{x_0}(x - x_0) + o(\|x - x_0\|_Y) \qquad (x \to x_0), \qquad \textit{für alle } x \in A. \tag{4.7}$$

(2) Ist f an der Stelle x_0 Fréchet-differenzierbar, so ist die Ableitung $f'(x_0)$ eindeutig bestimmt.
(3) Ist f an der Stelle x_0 Fréchet-differenzierbar, so ist f dort auch stetig.

Insbesondere rechtfertigt die Aussage von Lemma 4.1.(2), dass die Definition 93 überhaupt sinnvoll ist.

Beweis. (1) Es reicht, eine passende Funktion R_{x_0} zu finden, da die Implikation $(ii) \Rightarrow (i)$ klar ist. Dazu betrachte R_{x_0}, welche durch

$$R_{x_0}(x) := \begin{cases} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{\|x - x_0\|_Y}, & \text{falls } x \neq x_0, \\ 0, & \text{falls } x = x_0. \end{cases}$$

definiert ist. Diese Funktion ist offensichtlich an der Stelle x_0 stetig, da aufgrund der Stetigkeit und der Linearität desOperators $f'(x_0)$

$$\lim_{x \to x_0} f'(x_0)(x - x_0) = f'(x_0)\left(\lim_{x \to x_0}(x - x_0)\right) = f'(x_0)(0) = 0.$$

(2) Wäre B_{x_0} ein weiterer beschränkter linerarer Operator von Y nach X, für den

$$f(x) = f(x_0) + B(x - x_0) + o(\|x - x_0\|_Y) \qquad (x \to x_0).$$

Da auch (4.7) gilt, kann man die beiden Identitäten vergleichen und folgern, dass

$$\lim_{x \to x_0}(A_{x_0} - B_{x_0})\left(\frac{x - x_0}{\|x - x_0\|_Y}\right) = 0,$$

oder äquivalent dazu, dass

$$\lim_{n \to \infty}(A_{x_0} - B_{x_0})\left(\frac{x_n - x_0}{\|x_n - x_0\|_Y}\right) = 0 \qquad \text{für alle } (x_n)_{n \in \mathbb{N}} \subset Y \text{ mit } \lim_{n \to \infty} x_n = x_0. \tag{4.8}$$

Beachte, dass $\frac{x - x_0}{\|x - x_0\|_Y}$ ein Element der Einheitssphäre von Y ist.

Wir wollen nun zeigen, dass $A_{x_0} w = Bw$ für *alle* Elemente der Einheitssphäre von Y. Es sei also $y \in Y$ mit $\|y\|_Y = 1$. Um y als Vektor der Form $\frac{x - x_0}{\|x - x_0\|_Y}$ darzustellen, betrachte

$$x_n := x_0 + \frac{y}{n}, \qquad n \in \mathbb{N}.$$

Nun gilt

$$\frac{x_n - x_0}{\|x_n - x_0\|_Y} = \frac{1}{\|x_n - x_0\|_Y}(x_n - x_0) = \frac{n}{\|y\|_Y}\frac{y}{n} = y \qquad \text{für alle } n \in \mathbb{N},$$

da $\|y\|_Y = 1$. Da $\lim_{n \to 0} x_n = x_0$ und $\frac{x_n - x_0}{\|x_n - x_0\|_Y} = y$ gilt schließlich

$$(A_{x_0} - B_{x_0})y = \lim_{n \to \infty}(A_{x_0} - B_{x_0})\left(\frac{x_n - x_0}{\|x_n - x_0\|_Y})\right) = 0.$$

Die Aussage folgt nun Dank der Anmerkung 92, dass $A_{x_0} = B_{x_0}$.

(3) Die Aussage folgt aus (1), indem man den Grenzwert für $x \to x_0$ der Identität in (4.7) bildet: Denn es gilt für alle $x \in A$

$$\begin{aligned}
\lim_{x\to x_0} f(x) &= \lim_{x\to x_0} \big(f(x_0) + f'(x_0)(x - x_0) + o(\|x - x_0\|_Y)\big) \\
&= f(x_0) + f'(x_0)(\lim_{x\to x_0}(x - x_0)) + \lim_{x\to x_0} o(\|x - x_0\|_Y) \\
&= f(x_0) + f'(x_0)(0) \\
&= f(x_0),
\end{aligned}$$

wobei die erste Identität aus der Folgenstetigkeit von $f'(x_0)$ gilt (Übungsaufgabe 91 und [**6**, Satz 7.57]) folgt, die zweite aus der Definition des Landau-Symbols o, und die letzte Identität aufgrund von (4.2) gilt. □

Die Fréchet Ableitung jeder linearen Abbildung f ist wieder linear und stimmt mit f überein. Allgemeiner gilt Folgendes.

Beispiel 94. Es seien $T \in \mathcal{L}(Y, X)$ und $\tau \in X$ und betrachte die Funktion

$$f : Y \ni x \mapsto Tx + \tau \in X.$$

Dann ist f Fréchet-differenzierbar und $f'(x_0) = T$ für alle $x_0 \in Y$: Da die Bedingung (4.5) für $A_{x_0} \equiv f'(x_0)$ offensichtlich erfüllt ist (es gilt ja wegen der Linearität $Ax - Ax_0 = A(x - x_0)$), folgt die Aussage direkt aus Lemma 4.1.(1). □

Die ersten Rechenregeln folgen unmittelbar aus dem Lemma 4.1.

Satz 4.2. *Es seien f, g zwei Funktionen von A nach X und $x_0 \in A$, die an der Stelle x_0 Fréchet-differenzierbar sind. Dann gelten die folgenden Rechenregel.*

(1) Die Funktion $(f + g)$ ist in x_0 differenzierbar und $(f + g)'(x_0) = f'(x_0) + g'(x_0)$.
(2) Ist $\alpha \in \mathbb{R}$, so ist αf in x_0 differenzierbar und $(\alpha f)'(x_0) = \alpha f'(x_0)$.
(3) Es seien Z ein normierter Raum, $B \subset Z$ offen, $h : B \to Y$ eine Funktion. Ist $y_0 \in B$ eine Stelle, an der h differenzierbar ist und wo $h(y_0) = x_0$ gilt, so ist $f \circ h$ an der Stelle y_0 Fréchet-differenzierbar und es gilt

$$(f \circ h)'(y_0) = f'(h(y_0))h'(y_0)$$

(der Term an der rechten Seite ist die Verkettung zweier Operatoren).

Beweis. (1) Man sieht, dass $f'(x_0) + g'(x_0)$ als Summe zweier beschränkter linearer Operatoren selber ein beschränkter linearer Operator ist, welche auch offensichtlich die Bedingung

$$(f + g)(x) = (f + g)(x_0) + (f'(x_0) + g'(x_0))(x - x_0) + o(\|x - x_0\|_Y) \qquad (x \to x_0), \qquad \text{für alle } x \in A,$$

erfüllt. Somit ist dieser der gesuchte beschränkte lineare Operator, welcher die Fréchet-Ableitung definiert.

(2) Ähnlich sieht man, dass $\alpha f'(x_0)$ der gesuchte beschränkte lineare Operator ist.

(3) Es seien $R_{x_0} : A \to X$ und $S_{y_0} : B \to Y$ so, dass

$$\begin{aligned}
f(x) &= f(x_0) + f'(x_0)(x - x_0) + R_{x_0}(x)\|x - x_0\|_Y \qquad (x \to x_0), \qquad \text{für alle } x \in A, \\
h(y) &= h(y_0) + h'(y_0)(y - y_0) + S_{y_0}(y)\|y - y_0\|_Z \qquad (x \to x_0), \qquad \text{für alle } y \in B.
\end{aligned}$$

Wertet man an der Stelle $x = h(y)$ aus, so findet man erst Dank dem Ausdruck für f und dann Dank dem Ausdruck für h,

$$\begin{aligned} f(h(y)) &= f(h(y_0)) + f'(h(y_0))(x - x_0) + R_{x_0}(h(y))\|x - x_0\|_Y \\ &= f(h(y_0)) + f'(h(y_0))\Big(h'(y_0)(y - y_0) + S_{y_0}(y)\|y - y_0\|_Z\Big) \\ &\qquad + R_{x_0}(h(y))\left\|h'(y_0)(y - y_0) + S_{y_0}(y)\|y - y_0\|_Z\right\|_Y \\ &= f(h(y_0)) + f'(h(y_0))h'(y_0)(y - y_0) + f'(h(y_0))S_{y_0}(y)\|y - y_0\|_Z \\ &\qquad + R_{x_0}(h(y))\left\|h'(y_0)\frac{(y - y_0)}{\|y - y_0\|_Z} + S_{y_0}(y)\right\|_Y \|y - y_0\|_Z \\ &= f(h(y_0)) + f'(h(y_0))h'(y_0)(y - y_0) + T_{y_0}(y)\|y - y_0\|_Z, \end{aligned}$$

wobei $T_{y_0} : B \to X$ durch

$$T_{y_0}(y) := \begin{cases} f'(h(y_0))S_{y_0}(y) + R_{x_0}(h(y))\left\|h'(y_0)\frac{y-y_0}{\|y-y_0\|_Z} + S_{y_0}(y)\right\|_Y, & \text{falls } y \neq y_0, \\ 0, & \text{falls } y = y_0. \end{cases}$$

definiert ist, so dass T_{y_0} an der Stelle y_0 offensichtlich stetitg ist. Dabei haben wir insbesondere $h(y_0) = x_0$ und somit

$$x - x_0 = h(y) - h(y_0) = h'(y_0)(y - y_0) + S_{y_0}(y)\|y - y_0\|_Z.$$

verwendet. Dies vervollständigt den Beweis. □

4.3. Richtungsdifferenzierbarkeit

Aus einer Funktion $f : A \to X$, deren Argumente Vektoren eines höher- oder unendlichdimensionalen normierten Raums sind, kann man leicht eine Funktion einer reellen Variable basteln, indem man einen Vektor $v \in Y$ und einen Punkt $x_0 \in A$ auswählt und dann die Funktion

$$f_{x_0,v} : t \mapsto f(x_0 + tv) \tag{4.9}$$

betrachtet. (Diese Funktion ist sicherlich zumindest auf einem offenen Intervall $(-\varepsilon, \varepsilon) \subset \mathbb{R}$ definiert, da A offen ist und somit eine ε-Umgebung von x_0 enthält). An f_v können wir dann die übliche Theorie der Differenzialrechnung, die wir im [**6**, Kap. 8] eingeführt haben, anwenden. D.h., man kann für alle $v \neq \{0\}$ die Funktion f_v auf Differenzierbarkeit prüfen.

Definition 95. *Es seien f eine Funktion von A nach X, $v \in Y$ und $x_0 \in A$. Ist die Funktion $f_{x_0,v}$, welche in (4.9) eingeführt wurde, an der Stelle 0 differenzierbar, so heißt*

$$f'_{x_0,v}(0) = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t} \in X$$

ihre Richtungsableitung an der Stelle x_0 in Richtung v, *welche man*

$$D_v f(x_0), \qquad d_v f(x_0) \qquad \text{oder} \qquad \frac{\partial f}{\partial v}(x_0)$$

bezeichnet, und f heißt an der Stelle x_0 und in Richtung v partiell differenzierbar. *Existiert die Richtungsableitung von f an der Stelle x_0 in jeder Richtung v, so heißt f* an der Stelle x_0 richtungsdifferenzierbar.

Satz 4.3. *Es seien f eine Funktion von A und $x_0 \in A$. Ist f an der Stelle x_0 Fréchet-differenzierbar, so ist sie auch dort richtungsdifferenzierbar und es gilt*

$$\frac{\partial f}{\partial v}(x_0) = f'(x_0)v \qquad \text{für alle } v \in Y.$$

Beweis. Aufgrund vom [**6**, Satz 8.6] ist die Richtungsdifferenzierbarkeit von f in Richtung v an der Stelle x_0 dazu äquivalent, dass

$$f_{x_0,v}(t) = f_{x_0,v}(0) + \ell_{x_0} t + o(t) \qquad \text{für alle } t \text{ klein genug.}$$

Nun: Ist f an der Stelle x_0 Fréchet-differenzierbar, so gilt Dank dem Satz 4.1.(1) und der linearität von $f'(x_0)$.

$$f(x_0 + tv) = f(x_0) + f'(x_0)(tv) + o(\|tv\|_Y) \qquad (\|tv\|_Y \to 0),$$

und somit

$$f(x_0 + tv) = f(x_0) + tf'(x_0)v + o(|t|) \qquad (t \to 0),$$

was die gewünschte Identität ist. □

Beispiel 96. Anhand vom Satz 4.3 können wir auch zeigen, dass Richtungsdifferenzierbarkeit wohl eine strikt schwächere Eigenschaft als Fréchet-Differenzierbarkeit darstellt. Wir zeigen nämlich, dass es eine an einer Stelle richtungsdifferenzierbare Funktion f gibt, welche dort nicht Fréchet-differenzierbar ist. Betrachte nämlich $f : \mathbb{R}^2 \to \mathbb{R}$, die durch

$$f(x,y) := \begin{cases} \frac{x^2 y}{x^2+y^2} & \text{falls } (x,y) \neq (0,0), \\ 0 & \text{sonst,} \end{cases}$$

definiert ist. Dann gilt

$$f\big(t(x,y)\big) = tf(x,y) \qquad \text{für alle } (x,y) \in \mathbb{R}^2 \text{ und alle } t \in \mathbb{R}, \tag{4.10}$$

(direkte Rechnung!) und daher

$$\frac{\partial f}{\partial v}(0) = \lim_{t\to 0} \frac{f\big((0,0) + tv\big) - f\big((0,0)\big)}{t} = \lim_{t\to 0} \frac{f(tv)}{t} \overset{(4.10)}{\equiv} f(v).$$

Insbesondere ist f an der Stelle 0 richtunsdifferenzierbar.

Dennoch ist f nicht Fréchet-differenzierbar. Denn wenn sie es wäre, so würde ihre Fréchet-Ableitung angan der Stelle 0 z 4.3 an der Stelle 0

$$f'(0)(v) = \frac{\partial f}{\partial v}(0) \equiv f(v) \qquad \text{für alle } v \in \mathbb{R}^2$$

erfüllen. D.h., die Abbildungen an der Stelle 0 , welche beide $\mathbb{R}^2$ nach $\mathbb{R}$ abbilden, würden überall gleiche Werte annehmen – d.h., sie wären dann die selbe Funktion. Da aber $f'(0)$ definitionsgemäß linear sein muss, f aber nicht linear ist, haben wir damit einen Widerspruch gefunden. □

4.4. Der Spezialfall von $Y = \mathbb{R}^n$

Im Folgendem bezeichnen wir für jedes $i = 1, \ldots, n$ mit e_i den Vektor $(0, \ldots, 1, \ldots, 0) \in \mathbb{R}^n$, wobei die 1 an der i-te Stelle sich befindet.

Übungsaufgabe 97. Zeige: Je zwei Normen $\|\cdot\|, |||\cdot|||$ auf $\mathbb{R}^n$ sind *äquivalent*, d.h., es gibt $\mu, M > 0$ mit

$$\mu\|x\| \leq |||x||| \leq M\|x\| \qquad \text{für alle } x \in \mathbb{R}^n.$$

Folgere, dass je zwei Normen auf einem endlichdimensionalen Raum äquivalent sind. *(Hinweis: Man betrachte eine Basis $\{f_1, \ldots, f_n\}$ eines n-dimensionalen Raums, eine Basis $\{e_1, \ldots, e_n\}$ von $\mathbb{R}^n$ und die Abbildung T, welche durch*

$$Tx := \sum_{k=1}^n x_k e_k \qquad \textit{für } x = \sum_{k=1}^n x_k f_k$$

definiert ist, und zeige, dass T ein Isomorphismus ist).

Lemma 4.4. *Jede lineare Abbildung zwischen endlichdimensionalen Räumen ist stetig.*

Tatsächlich könnte man dieses Lemma verallgemeinern zum Fall der Operatoren, welche lediglich auf einem endlichdimensionalen Raum *definiert* sind, vgl. [**2**, Thm. VII.1.6].

Beweis. Angesichts der Aussage in der Übungsaufgabe 97 reicht es o.B.d.A. den Fall einer Abbildung von $\mathbb{R}^n$ nach $\mathbb{R}^m$ für je zwei $n, m \in \mathbb{N}$ zu betrachten, denn jeder andere endlichdimensionale Raum ist ja isomorph zu einem Euklidischen Raum $\mathbb{R}^d$ für ein $d \in \mathbb{N}$ und die Verkettung eines beschränkten linearen Operators mit einem Isomorphismus ist wieder linear und beschränkt.

Insbesondere dürfen wir die Aussage bzgl. der ∞-Norm beweisen, d.h. der Norm, die für alle d durch

$$\|\cdot\|_\infty : \mathbb{R}^d \ni x \mapsto \max_{1 \leq j \leq d} |x_j| \in [0, \infty],$$

definiert ist. Man betrachte also $x \in \mathbb{R}^n$ und eine lineare Abbildung $T : \mathbb{R}^n \to \mathbb{R}^m$, die laut Beispiel 87 in Form einer Multiplikation mit einer Matrix A dargestellt werden kann, also $T \equiv T_A$. Dann gilt

$$Tx = \begin{pmatrix} \sum_{k=1}^n a_{1k} x_k \\ \vdots \\ \sum_{k=1}^n a_{mk} x_k, \end{pmatrix} \qquad \text{für } x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

und somit

$$\|Tx\|_\infty = \max_{1 \leq j \leq m} \left| \sum_{k=1}^n a_{jk} x_k \right| \leq n \max_{\substack{1 \leq j \leq m \\ 1 \leq k \leq n}} |a_{jk}| \max_{1 \leq k \leq n} |x_k|,$$

d.h.,

$$\|Tx\|_\infty \leq M\|x\|_\infty$$

für $M := n \max_{\substack{1 \leq j \leq m \\ 1 \leq k \leq n}} |a_{jk}|$. □

Lemma 4.5. *Für alle $n, m \in \mathbb{N}$ ist der normierte Vektorraum $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ isomorph zum normierten Vektorraum $M_{m \times n}(\mathbb{R})$ der $m \times n$-Matrizen mit reellen Einträgen. Insbesondere ist $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ ein Banachraum.*

Beweis. Das Beispiel 87 zeigt, dass die Abbildung $A \mapsto T_A$ eine Bijektion von der Menge der *linearen* Abbildungen nach $M_{m \times n}(\mathbb{R})$, und wir können nun Dank Lemma 4.4 ergänzen, sie ist eigentlich sogar eine Bijektion von der Menge der *beschränkten linearen* Abbildungen nach $M_{m \times n}(\mathbb{R})$.

Wiederum ist $M_{m\times n}(\mathbb{R})$ offensichtlich isomorph zu $\mathbb{R}^{m\cdot n}$ vermöge der Identifikation

$$A = (a_{jk})_{\substack{1\le j\le m\\ 1\le k\le n}} \mapsto (a_{11}, \dots a_{m1}, a_{12}, \dots a_{m2}, \dots, a_{1n}, \dots a_{mn}).$$

Um die Vollständigkeit vom $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ zu beweisen reicht es zu beachten, dass $\mathbb{R}^{m\cdot n}$ aufgrund des Satzes von Bolzano–Weierstraß vollständig ist und dass darüber hinaus jeder einem Banachraum isomorphe normierter Raum selber vollständig ist: Denn ist $(x_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge, so ist ihr Bild im Banachraum unter dem Isomorphismus wieder Cauchy, und somit konvergent, und nun folgt die Konvergenz der ursprünglichen Folge Dank der Linearität und somit der Stetigkeit der Inverse des Isomorphismus. $\square$

Wir wollen nun den Spezialfall einer Abbildung zwischen zwei endlichdimensionalen Vektorräumen betrachten, also o.B.d.A. von $\mathbb{R}^n$ nach $\mathbb{R}^m$. Wir können natürlich nach wie vor den Begriff der Fréchet-Differenzierbarkeit untersuchen, wobei man im endlichdimensionalen Kontext oft auch von *totaler Differenzierbarkeit* spricht – die Definition bleibt aber gleich, wobei das Lemma 4.5 nun erlaubt, die Definition 93 leicht anders zu formulieren.

KOROLLAR 4.6. *Es sei $x_0 \in A$, $A \subset \mathbb{R}^n$. Eine Abbildung $f : A \to \mathbb{R}^m$ ist genau dann an der Stelle x_0 Fréchet-differenzierbar, wenn es ein $A_{x_0} \in M_{m\times n}(\mathbb{R})$ existiert, für die*

$$\lim_{x\to x_0} \frac{f(x) - f(x_0) - A_{x_0}\cdot(x - x_0)}{\|x - x_0\|_X} = 0 \qquad \text{(bzgl. einer beliebigen Norm auf } \mathbb{R}^n\text{)}. \tag{4.11}$$

Die Notation haben wir bewusst so gewählt, dass die Bedingungen in (4.5) und (4.11) sehr ähnlich aussehen, denn wir möchten gerade dazu motivieren, eine beschränkte lineare Abbildung mit ihrer Darstellungsmatrix zu identifizieren.

DEFINITION 98. *Unter den Voraussetzungen und mit der Notation vom Korollar 4.6 heißt dann $A_{x_0} \in M_{m\times n}(\mathbb{R})$ die* Jacobi-Matrix *von f an der Stelle x_0 und wird manchmal mit $J_f(x_0)$ bezeichnet.*

Die Jacobi-Matrix wird nach Carl Gustav Jacob Jacobi genannt.

ANMERKUNG 99. Angesichts vom Satz 4.3 stimmt dann $J_f(x_0)$ mit der Matrix

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \dots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \dots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix}$$

überein, wobei

$$\frac{\partial f}{\partial x_j} := \frac{\partial f}{\partial e_j},$$

die Richtungsableitung in Richtung des *Basisvektors* e_j ist, d.h. des Vektors

$$e_j = (0, \dots, 0, 1, 0, \dots, 0)$$

von $\mathbb{R}^n$, der eine 1 an der j-ten Stelle hat, und sonst laute 0.

Alle bekannte Rechenregel für differenzierbare Funktionen können nun als Gleichungen für Jacobi-Matrizen dargestellt werden. So gilt z.B.

$$J_{f\circ h}(y_0) = J_f(h(y_0)) \cdot J_h(y_0), \tag{4.12}$$

an jeder Stelle y_0, an der h differenzierbar ist und für die auch f an der Stelle $h(y_0)$) differenzierbar ist.

Dabei wird ein noch speziellerer Fall eine Sonderrolle spielen. Wir bezeichnen wir für einen normierten Raum Y auf $\mathbb{K}$ mit Y' den *Dualraum* von Y, d.h., den Raum aller linearen Abbildungen von Y nach $\mathbb{K}$.

Definition 100. *Es seien $X = \mathbb{R}$ und $A \subset Y = \mathbb{R}^m$. Ist $f : A \to \mathbb{R}$ Fréchet-differenzierbar, so heißt das Differenzial Df von f das* Gradient *von f. Es wird aus historischen Gründen oft*

$$\mathrm{grad} f \qquad \textit{oder auch} \qquad \nabla f$$

bezeichnet[2].

Definitionsgemäß ist das Gradient einer Funktion $f : A \to \mathbb{R}$ selber eine Abbildung von A nach Y'.

Anmerkung 101. Mit $Y = \mathbb{R}^m$ gilt formal eigentlich $f'(x_0) \in Y'$ während $\mathrm{grad} f(x_0) \equiv \left(\frac{\partial f}{\partial e_1}(x_0), \ldots, \frac{\partial f}{\partial e_m}(x_0)\right) \in Y$. Angesichts vom Lemma 4.5 ist aber dieser Unterschied kaum relevant. Die letzte Identifikation ist vermöge

$$\mathbb{R}^n \ni \xi \mapsto (\xi|\cdot)_{\mathbb{R}^n} \in \mathbb{R}$$

zu verstehen, wobei $(\xi|\eta)_{\mathbb{R}^n}$ das *Skalarprodukt* von ξ und η darstellt,

$$(\xi|\eta)_{\mathbb{R}^n} := \xi^T \cdot \eta = \sum_{k=1}^n \xi_n \eta_n.$$

Der Grund für die letzte Notation in (4.6) sollte nun klar sein: Denn sind $X = Y = \mathbb{R}$, so ist die Ableitung einer Funktion f an einer Stelle x_0 (eben $f'(x_0)$) eine lineare Abbildung von $\mathbb{R}$ nach $\mathbb{R}$ – die man als 1×1-Matrix, also als Skalar, darstellen kann: Das ist gerade die übliche Ableitung, welche man in der Vorlesung zu Analysis 1 kennengelernt hat.

Somit hat man gesehen, dass falls eine Funktion von $\mathbb{R}^m$ nach $\mathbb{R}^n$ Fréchet-differenzierbar ist, so ist ihre Ableitung an der Stelle x_0 eine $n \times m$-Matrix $f'(x_0)$. So ist die Ableitung eine $1 \times m$-Matrix – also eine reelle Zahl, falls $n = 1$. Dies zeigt, dass der Begriff von Fréchet-Ableitung eine echte Verallgemeinerung des in der Vorlesung zu Analysis 1 eingeführten Begriffs ist.

Der Fall von $Y = \mathbb{R}^n$ ist nicht wirklich verschieden vom allgemeinen Fall, doch nun werden einige Rechnungen vereinfacht und man kann neue Beispiele betrachten.

Beispiel 102. Es sei $A \in M_{n\times n}(\mathbb{R})$ symmetrisch und betrachte die Funktion

$$f : \mathbb{R}^n \ni x \mapsto x^T A x = \sum_{j=1}^n a_{ij} x_i x_j.$$

Dann ist f Fréchet-differenzierbar und man hat

$$f'(x) = 2x^T A, \qquad x \in \mathbb{R}^n :$$

Dieser soll als Operator

$$x^T A : y \mapsto x^T A y = \sum_{j=1}^n a_{ij} x_i y_j$$

verstanden werden. □

[2] Das Symbol ∇ spricht man "Nabla" aus.

Anmerkung 103. Richtungsdifferenzierbarkeit wird aus historischen Gründen im Fall $A \subset \mathbb{R}^n$ anders genannt, falls die Richtung ein Basisvektor e_i ist – man spricht üblicherweise von partiellen Differenzierbarkeit.

Genauer: Es seien $A \subset \mathbb{R}^n$ eine offene Menge, $f : A \to X$ eine Funktion und $x_0 \in A$. Existiert die Richtungsableitung von f an der Stelle x_0 in Richtung e_i, d.h. der Grenzwert

$$\frac{\partial f}{\partial e_i}(x_0) := \lim_{t\to 0} \frac{f(x_0 + te_i) - f(x_0)}{t},$$

für ein $i \in \{1, \dots, n\}$, so nennt man ihn die *partielle Ableitung von f im Punkt x_0 in Richtung e_i* und man schreibt dann meistens

$$\frac{\partial f}{\partial x_i}(x_0) := \frac{\partial f}{\partial e_i}(x_0).$$

Die Funktion f heißt dann *in x_0 in Richtung e_i*, oder *in der i-ten Koordinate, partiell differenzierbar*. Ist f in x_0 in Richtung e_i für jedes $i = 1, \dots, n$ differenzierbar, so heißt f *in x_0 partiell differenzierbar*. Ist f in jedem $x_0 \in A$ partiell differenzierbar, so heißt f schlicht *partiell differenzierbar*. Ist f partiell differenzierbar und ist jede partielle Ableitung stetig, so heißt f *stetig partiell differenzierbar*.

Anders gesagt, partielle Differenzierbarkeit einer Funktion f an einer Stelle x_0 bedeutet, dass jedes ϕ_ℓ an der Stelle 0 (als vektorwertige Funktion einer reellen Variabel) differenzierbar ist, wobei

$$\phi_\ell : t \mapsto f(x_0 + te_\ell) \qquad \ell = 1, \dots, m. \tag{4.13}$$

Wir bekräftigen, dass sich die Notation von der Notation im 1-dimensionalen Fall unterscheidet. Die Einführung des Symbols ∂ geht auf Adrien-Marie Legendre zurück, sie wurde üblich nachdem 1841 Carl Gustav Jacob Jacobi sie wieder einführte.

Ist f auf einer Teilmenge von $\mathbb{R}^2$ oder $\mathbb{R}^3$ definiert, so benutzt man oft die Bezeichnung (x, y) bzw. (x, y, z) statt (x_1, x_2, x_3) für ihre Variablen. Dementsprechend werden oft die partielle Ableitungen mit $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$ bezeichnet. Vorsicht! In diesem Fall bezeichnet x nicht mehr einen Vektor aus $\mathbb{R}^n$ ist, sondern eine reelle Zahl.

Anmerkung 104. So ist z.B. eine *partielle Differenzialgleichung* eine Gleichung, deren Unbekannte eine Funktion mehreren Variablen ist, deren partiellen Ableitung eine gewisse Relation erfüllen müssen. Ein einfacher Fall ist die *Transportgleichung*

$$\frac{\partial u}{\partial t}(t, x) = \frac{\partial u}{\partial x}(t, x), \qquad t, x \in \mathbb{R},$$

deren Lösungen immer der Gestalt

$$u(t, x) := f(t + x), \qquad t, x \in \mathbb{R},$$

für eine differenzierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ sind. Die Funktion f wird eindeutig von $u(0, \dot{)}$ bestimmt. Intepretieren wir t als *Zeitvariabel*, so stellt $u(0, \cdot)$ das Profil der Funktion am Zeitpunkt 0 – so zu sagen am Anfang: und tatsächlich wird

$$u(0, x) = f(x), \qquad x \in \mathbb{R},$$

als *Anfangsbedingung* betrachtet. Anders gesagt: Die Transportgleichung hat eine Lösung, die von der Anfangsbedingung eindeutig bestimmt ist.

Die Grundidee der partiellen Differentiation in Richtung e_i ist also, die anderen Koordinaten der Argumente von f "einzufrieren", und bzgl. der i-te Koordinate so abzuleiten, als ob f Funktion *einer* Variable wäre: anders gesagt, eine übliche Ableitung der i-ten partielle Funktion zu berechnen.

Beispiel 105. Betrachte die Funktion aus Beispiel 84. Wir zeigen, dass ihre beide Komponenten partiell differenzierbar sind. Für die erste Komponente x gilt also

$$\frac{\partial x}{\partial v} = \lim_{h\to 0} \frac{(v+h)t\cos(\alpha) - vt\cos(\alpha)}{h} = t\cos(\alpha) \lim_{h\to 0} \frac{(v+h)-v}{h} = t\cos(\alpha),$$

da für den betrachteten Grenzwert t und $\cos(\alpha)$ die Rolle von Konstanten spielen. ähnlich gilt

$$\frac{\partial x}{\partial \alpha} = \lim_{h\to 0} \frac{vt\cos(\alpha+h) - vt\cos(\alpha)}{h} = vt \lim_{h\to 0} \frac{\cos(\alpha+h)-\cos(\alpha)}{h} = -vt\sin(\alpha).$$

Weiter gilt

$$\frac{\partial x}{\partial g} = 0,$$

weil die Funktion x nicht von g abhängt, und schließlich

$$\frac{\partial x}{\partial t} = \lim_{h\to 0} \frac{v(t+h)\cos(\alpha) - vt\cos(\alpha)}{h} = v\cos(\alpha) \lim_{h\to 0} \frac{(t+h)-t}{h} = v\cos(\alpha).$$

ähnlich kann man sehen, dass die partielle Ableitungen der Funktion y

$$\frac{\partial y}{\partial v} = t\sin(\alpha), \qquad \frac{\partial y}{\partial \alpha} = vt\cos(\alpha), \qquad \frac{\partial y}{\partial g} = -\frac{t^2}{2}, \qquad \frac{\partial y}{\partial t} = -gt + v\sin(\alpha) \tag{4.14}$$

betragen. □

Anmerkung 106. Es sei $f : A \to \mathbb{R}^n$ eine an der Stelle $x_0 \in A$ partiell differenzierbare Funktion. Bezeichne $f \equiv (f_1, \ldots, f_n)$, mit $f_i : A \to \mathbb{R}$ für alle i. Dann heißt

$$\nabla \cdot f(x_0) := \mathrm{div} f(x_0) := \frac{\partial f_1}{\partial x_1}(x_0) + \ldots + \frac{\partial f_n}{\partial x_n}(x_0)$$

die *Divergenz* von f an der Stelle x_0.

Beispiel 107. Anders als die Differenzierbarkeit von Funktionen einer Variablen gewährleistet die partielle Differenzierbarkeit einer Funktionen mehrerer Variablen f die Stetigkeit von f nicht. Betrachte z.B.

$$f(x,y) := \begin{cases} \frac{xy}{x^2+y^2} & \text{falls } (x,y) \neq (0,0), \\ 0 & \text{sonst.} \end{cases}$$

Die Funktion f ist dann in 0 partiell differenzierbar mit

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} = 0 = \lim_{h\to 0} \frac{f(0,h) - f(0,0)}{h} = \frac{\partial f}{\partial y}(0,0).$$

Doch ist sie dort unstetig, wie man sieht, indem dass man die Folgen

$$\left(\frac{1}{n}, \frac{1}{n}\right)_{n\in\mathbb{N}} \qquad \text{und} \qquad \left(\frac{1}{n}, -\frac{1}{n}\right)_{n\in\mathbb{N}}$$

betrachtet. Denn beide sind Nullfolgen in $\mathbb{R}^2$, doch ist

$$\lim_{n\to\infty} f\left(\frac{1}{n}, \frac{1}{n}\right) = \lim_{n\to\infty} \frac{\frac{1}{n^2}}{\frac{2}{n^2}} = \frac{1}{2}$$

während

$$\lim_{n\to\infty} f\left(\frac{1}{n}, -\frac{1}{n}\right) = \lim_{n\to\infty} \frac{\frac{-1}{n^2}}{\frac{2}{n^2}} = -\frac{1}{2}$$

□

Während die partiellen Ableitungen eine Funktion (wenn sie überhaupt existieren) zu Bestimmen normalerweise eine leichte Aufgabe ist, ist das direkte Überprüfen der Fréchet-Differenzierbarkeit in der Regel mühsam. Wie wir schon wissen stellt grundsätzlich die partielle Differenzierbarkeit eine Funktion leider keine ausreichende Bedingung für die Fréchet-Differenzierbarkeit dar.

Allerdings kann man wohl ein Kriterium für Fréchet-Differnzierbarkeit anhand von partieller Differenzierbarkeit finden, wenn man die Voraussetzungen etwas verstärkt.

Satz 4.7. *Es sei $A \subset \mathbb{R}^m$ offen. Genau dann ist eine Funktion $f : A \to \mathbb{R}^n$ Fréchet-differenzierbar, wenn sie stetig partiell differenzierbar ist.*

Beweis. Den Beweis der ersten Implikation ist eine unmittelbare Folgerung des Satzes 4.3.

Zum Beweis der zweiten Implikation betrachte ein $x_0 \in A$ und führe eine Matrix $A_{x_0} \in M_{n\times m}(\mathbb{R})$ durch

$$(A_{x_0})_{ij} := \frac{\partial f_j}{\partial x_i}(x_0), \qquad h \equiv (h_1, \dots, h_m) \in \mathbb{R}^m,$$

ein, wobei f_j die j-te Komponente der vektorwertigen Funktion f ist. Aufgrund seiner Linearität und angesichts vom Lemma 4.4 ist A_{x_0} sicherlich beschränkt. Wir wollen nun zeigen, dass

$$\lim_{h\to 0} \frac{f(x+h) - f(h) - A_{x_0} \cdot h}{|h|} = 0.$$

d.h., dass A_{x_0} die Fréchet-Ableitung von f ist. Die Idee ist dabei, dass man die Fréchet-Ableitung so zu sagen als lineare Kombination von Fréchet-Ableitungen von Funktionen einer Variabel – also als Ableitungen im Sinne der Analysis 1! – darstellen kann, woran man die bekannten Sätze anwenden kann: in diesem Fall den Satz von Lagrange.

Es sei $\varepsilon > 0$ mit $B_\varepsilon(x_0) \subset A$ und betrachte $h = (h_1, \dots, h_m) \in \mathbb{R}^m$ und eine Menge $\{x_1, \dots, x_m\}$ mit

$$x_k := x_0 + \sum_{j=1}^{k} h_j e_j, \quad k = 1 \dots, m, \qquad \text{und} \qquad |h| < \varepsilon.$$

Dies gewährleistet, dass

- $x_j \in U_\varepsilon(x_0)$ für alle $j = 1, \dots, m$,
- jedes $x_j - x_{j-1}$ parallel zur h-ten Achse ist für alle $j = 1, \dots, m$ und somit
- $(x_0, x_1), (x_1, x_2), \dots, (x_{m-1}, x_m)$ einen Kantenzug von x_0 nach $x_0 + h$ bildet, dessen Kanten jeweils parallel zu einer der Achsen laufen.

Nun gilt

$$\begin{aligned} f(x_0+h)-f(x_0) &= \sum_{j=1}^m \big(f(x_j)-f(x_{j-1})\big) \\ &= \sum_{j=1}^m \big(f(x_{j-1}+h_je_j)-f(x_{j-1})\big) \\ &= \sum_{j=1}^m \big(\phi_j(1)-\phi_j(0)\big) \end{aligned}$$

wobei ϕ_j – ähnlich wie in (4.13) – durch

$$\phi_j : [0,1] \ni t \mapsto f(x_{j-1}+th_je_j) \in \mathbb{R}$$

definiert ist. Dank unseren Voraussetzungen können wir den Hauptsatz der Differenzial- und Integralrechnung (in der Form vom Korollar 2.14 anwenden und schließen, dass

$$\phi_j(1)-\phi_j(0) \stackrel{!}{=} \int_0^1 \phi_j'(\xi_j)d\xi_j = \int_0^1 h_j \frac{\partial f}{\partial x_j}(\xi_j)d\xi_j.$$

Deshalb

$$\begin{aligned} f(x_0+h)-f(x_0)-A_{x_0}\cdot h &= \sum_{j=1}^m \int_0^1 h_j \frac{\partial f}{\partial x_j}(\xi_j)d\xi_j - \sum_{j=1}^m h_j \frac{\partial f_j}{\partial x_i}(x_0) \\ &= \sum_{j=1}^m \int_0^1 h_j \left(\frac{\partial f}{\partial x_j}(\xi_j) - \frac{\partial f}{\partial x_j}(x_0)\right) d\xi_j, \end{aligned}$$

wobei die letzte Identität gilt, weil die Terme $h_j \frac{\partial f_j}{\partial x_i}(x_0)$ bzgl. der Variabel ξ_j konstant sind. Daher bzgl. einer beliebigen Norm von $\mathbb{R}^n$ (beachte die Übungsaufgabe 97!)

$$\|f(x_0+h)-f(x_0)-A_{x_0}\cdot h\| \le \max_{1\le j\le m} |h_j| \sum_{j=1}^m \int_0^1 \left\|\frac{\partial f}{\partial x_j}(\xi_j) - \frac{\partial f}{\partial x_j}(x_0)\right\| d\xi_j.$$

Im Grenzwert für $h\to 0$ konvergieren alle ξ_j gegen x_0, und somit in Anbetracht der Stetigkeiten aller partiellen Ableitungen

$$\lim_{h\to 0} \left\|\frac{\partial f}{\partial x_j}(\xi_j) - \frac{\partial f}{\partial x_j}(x_0)\right\| = 0 \qquad \text{für alle } j=1,\dots,m.$$

Wir folgern, dass

$$\lim_{h\to 0} \frac{f(x_0+h)-f(x_0)-A_{x_0}}{\|h\|} = 0$$

(angesichts der Übungsaufgabe 97 spielt natürlich keine Rolle, welche Norm wir betrachten). Dies vervollständigt den Beweis. □

Anmerkung 108. Im Beispiel 21 war der Definitionsbereich von f das kartesische Produkt dreier offener Intervalle. Man kann zeigen, dass Produkte offener Intervalle immer offen im Produktraum (in diesem Fall, $\mathbb{R}^3$) ist. Doch ist nicht jede in $\mathbb{R}^n$ offene Menge ein kartesisches Produkt: ein Gegenbeispiel liefern alle ε-Umgebungen

$$U_\varepsilon(x_0) = \left\{ x \in \mathbb{R}^n : \|x - x_0\| = \sqrt{\sum_{i=1}^N (x_i - x_{i0})^2} < \varepsilon. \right\}$$

Übungsaufgabe 109. (1) Es sei $\alpha \in \mathbb{N}^N$ ein *Multiindex der Länge* $|\alpha|$, d.h., $\alpha = (\alpha_1, \ldots, \alpha_N)$ mit $\alpha_1, \ldots, \alpha_N \in \mathbb{N}$ und $|\alpha| := \sum_{j=1}^N \alpha_j$. Definiere für $x = (x_1, \ldots, x_N) \in \mathbb{R}^n$ die Potenz $x^\alpha := x_1^{\alpha_1} \cdots x_N^{\alpha_N} \in \mathbb{R}$.

Es seien nun $k \in \mathbb{N}$ und $a_\alpha \in \mathbb{R}$ für alle Multiindizes α mit $|\alpha| \leq k$. Dann heißt $f(x) := \sum_{|\alpha| \leq k} a_\alpha x^\alpha$ ein *multivariates Polynom* vom Grad $k \in \mathbb{N}$ (und ist für $N \neq 1$ *kein* übliches Polynom!). Zeige, dass f und alle ihre partielle Ableitungen stetig sind.

(2) Es sei f ein *homogenes* multivariates Polynom vom Grad $k \in \mathbb{N}$, d.h., $f(x) = \sum_{|\alpha|=k} x^\alpha$. Zeige, dass $f(\lambda x) = \lambda^k f(x)$ für alle $x \in \mathbb{R}^n$ und alle $\lambda \geq 0$.

Übungsaufgabe 110. Bestimme die partiellen Ableitungen der Funktion f, die folgendermaßen definiert ist.

(1) $f(x, y, z) := \exp(xyz^2)$;
(2) $f(x, y) := \log(\frac{x}{y})$;
(3) $f(x, y) := \cosh(x) \cdot \sinh(y)$.

Man merke, dass die partielle Ableitung in Richtung e_i selbst eine Funktion von A nach $\mathbb{R}$ ist, falls f in ganz A in Richtung e_i partiell differenzierbar ist. Sie kann somit selbst differenzierbar sein.

Definition 111. *Es seien $A \subset \mathbb{R}^n$ eine offene Menge und $f : A \to X$ eine in $x_0 \in A$ in Richtung e_j differenzierbare Funktion. Ist $\frac{\partial f}{\partial x_j}$ selber in x_0 in Richtung e_i differenzierbar, d.h., existiert der Grenzwert*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) := \lim_{h \to 0} \frac{\frac{\partial f}{\partial x_j}(x_0 + he_i) - \frac{\partial f}{\partial x_j}(x_0)}{h}$$

für ein $i \in \{1, \ldots, N\}$ so nennt man ihn die 2. partielle Ableitung von f im Punkt x_0 in Richtung e_j und dann e_i. *Die Funktion f heißt dann* in x_0 in Richtung e_j und dann e_i partiell differenzierbar. *Ist f in x_0 in Richtung e_i und dann e_j für jede $i, j = 1, \ldots, n$ differenzierbar, so heißt f* in x_0 2-mal partiell differenzierbar. *Ist f in jedem $x_0 \in A$ 2-mal partiell differenzierbar, so heißt f schlicht* 2-mal partiell differenzierbar.

Dementsprechend definiert man den Begriff von k-mal partielle Differenzierbarkeit für $k \in \mathbb{N}$ beliebig. Ist eine Funktion k-mal partiell Differenzierbar und sind alle ihrer k-te partielle Ableitungen stetig, so heißt sie k-mal stetig partiell differenzierbar. *Den Raum der stetig k-mal partiell differenzierbaren Funktionen von $A \subset \mathbb{R}^m$ nach X bezeichnet man $C^k(A; X)$.*

Eine glatte Funktion *ist eine Funktion $f : A \to \mathbb{R}$, welche für jedes $n \in \mathbb{N}^*$ n-mal partiell differenzierbar ist.*

Die 2. partielle Ableitung in der einzigen Richtung e_i bezeichnet man mit

$$\frac{\partial^2 f}{\partial x_i^2} := \frac{\partial f}{\partial x_i \partial x_i}$$

und allgemeiner

$$\frac{\partial^n f}{\partial x_i^n} := \frac{\partial f}{\partial x_i \dots \partial x_i}$$

Man nennt die partielle Ableitung der Form $\frac{\partial^2 f}{\partial x_i \partial x_j}$ *gemischte Ableitung*, falls $i \neq j$. Falls beide existieren, müssen die partiellen Ableitungen $\frac{\partial^2 f}{\partial x_i \partial x_j}$ und $\frac{\partial^2 f}{\partial x_j \partial x_i}$ übereinstimmen? Im Allgemeinen nicht.

Beispiel 112. Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch

$$f(x,y) := \begin{cases} xy\frac{x^2-y^2}{x^2+y^2} & \text{falls } (x,y) \neq (0,0), \\ 0 & \text{sonst.} \end{cases}$$

Dann ist f stetig. Um das zu überprüfen reicht es zu zeigen, dass f an der Stelle 0 stetig ist: Denn anderswo ist sie offensichtlich stetig als Summe und Quotient (mit nichtverschwindendem Nenner) stetiger Funktionen. Es sei dazu $\varepsilon > 0$ und setze $\delta := \sqrt{\varepsilon}$. Dann gilt für alle $(x,y) \in \mathbb{R}^2$ mit $\sqrt{x^2+y^2} \leq \delta$ auch $|f(x,y)| \leq |xy|$ (weil aus $-y^2 \leq y^2$ folgt $x^2 - y^2 \leq x^2 + y^2$) und somit

$$|f(x,y)| \leq |xy| \leq x^2 + y^2 \leq \delta^2 = \varepsilon.$$

Dabei haben wir benutzt, dass

$$0 \leq (x \pm y)^2 = x^2 + y^2 \pm 2xy,$$

und somit

$$|xy| \leq \frac{x^2}{2} + \frac{y^2}{2} \leq x^2 + y^2.$$

Man kann aber ausrechnen (Übungsaufgabe!), dass $\frac{\partial^2 f}{\partial x \partial y}(0,0) \neq \frac{\partial^2 f}{\partial y \partial x}(0,0)$ □

Übungsaufgabe 113. Zeige, dass ein multivariates Polynom eine glatte Funktion ist.

Der folgende Satz wird üblicherweise auf Hermann Amandus Schwarz zurückgeführt. Er wurde allerdings erst 1743 von Alexis Claude Clairaut in einem Essay über Erdvermessung bewiesen, und ist gelegentlich auch als "Satz von Clairaut" bekannt.

Satz 4.8. *Es sei $f : A \to \mathbb{R}$ eine 2-mal stetig partiell differenzierbare Funktion. Dann stimmen die* gemischten *2. partiellen Ableitungen $\frac{\partial^2 f}{\partial x_j \partial x_i}$ und $\frac{\partial^2 f}{\partial x_i \partial x_j}$ auf ganz A für alle $1 \leq i,j \leq n$ überein.*

Beweis. Es reicht, den Fall von $A \subset \mathbb{R}^2$ zu betrachten, da so wie so nur zwei Richtungen eine Rolle spielen.

Es sei $x_0 := (x_1, x_2) \in \mathbb{R}^2$, $\varepsilon > 0$ mit $B_\varepsilon(x_0)) \subset A$ und $h := (h_1, h_2) \in \mathbb{R}^2$ mit $|h| < \varepsilon$, so dass insbesondere

$$(x_1 + sh_1, x_2 + th_2) \in A \qquad \text{für alle } t, s \in [0,1]$$

gilt. Unter unseren Voraussetzungen können wir den Satz von Lagrange auf den beiden Funktionen

$$\phi : [0,1] \ni s \mapsto f(x_1 + sh_1, x_2 + h_2) - f(x_1 + sh_1, x_2) \in \mathbb{R}$$

und

$$\psi : [0,1] \ni t \mapsto f(x_1 + h_1, x_2 + th_2) - f(x_1, x_2 + th_2) \in \mathbb{R}$$

anwenden. Erstens erhalten wir die Existenz eines $\xi_1 \in (0,1)$ mit

$$\begin{aligned} f(x_1+h_1,x_2+h_2) - f(x_1,x_2+h_2) & \\ -f(x_1+h_1,x_2) + f(x_1,x_2) &= \phi(x_1+h_1) - \phi(x_1) \\ &\stackrel{!}{=} h_1\phi'(x_1+\xi_1 h_1) \\ &\equiv h_1\left(\frac{\partial f}{\partial x_1}(x_1+\xi_1 h_1, x_2+h_2) - \frac{\partial f}{\partial x_1}(x_1+\xi_1 h_1, x_2)\right) \end{aligned}$$

und entsprechend auch die Existenz eines $\xi_2 \in (0,1)$ mit

$$\begin{aligned} f(x_1+h_1,x_2+h_2) - f(x_1,x_2+h_2) & \\ -f(x_1+h_1,x_2) + f(x_1,x_2) &= \psi(x_2+h_2) - \psi(x_2) \\ &\stackrel{!}{=} h_2\psi'(x_2+\xi_2 h_2) \\ &\equiv h_2\left(\frac{\partial f}{\partial x_2}(x_1+h_1, x_2+\xi_2 h_2) - \frac{\partial f}{\partial x_1}(x_1, x_2+\xi_2 h_2)\right). \end{aligned}$$

Somit

$$\begin{aligned} \tfrac{1}{h_1}\big(f(x_1+h_1,x_2+h_2) - f(x_1,x_2+h_2) & \\ -f(x_1+h_1,x_2) + f(x_1,x_2)\big) &= \tfrac{\partial f}{\partial x_1}(x_1+\xi_1 h_1, x_2+h_2) - \tfrac{\partial f}{\partial x_1}(x_1+\xi_1 h_1, x_2) \end{aligned} \tag{4.15}$$

und

$$\begin{aligned} \tfrac{1}{h_1}\big(f(x_1+h_1,x_2+h_2) - f(x_1,x_2+h_2) & \\ -f(x_1+h_1,x_2) + f(x_1,x_2)\big) &= \tfrac{\partial f}{\partial x_2}(x_1+h_1, x_2+\xi_2 h_2) - \tfrac{\partial f}{\partial x_2}(x_1, x_2+\xi_2 h_2). \end{aligned} \tag{4.16}$$

Wendet man wieder den Satz von Lagrange an, diesmal an die rechte Seite von (4.15), so erhält man die Existenz eines $\xi_3 \in (0,1)$ mit

$$\frac{\partial f}{\partial x_1}(x_1+\xi_1 h_1, x_2+h_2) - \frac{\partial f}{\partial x_1}(x_1+\xi_1 h_1, x_2) = h_2\frac{\partial^2 f}{\partial x_2\partial x_1}(x_1+\xi_1 h_1, x_2+\xi_3 h_2)$$

und entsprechend für (4.16) erhält man die Existenz eines $\xi_4 \in (0,1)$ mit

$$\frac{\partial f}{\partial x_2}(x_1+h_1, x_2+\xi_2 h_2) - \frac{\partial f}{\partial x_2}(x_1, x_2+\xi_2 h_2) = h_1\frac{\partial^2 f}{\partial x_1\partial x_2}(x_1+\xi_4 h_1, x_2+\xi_2 h_2).$$

Vergleicht man diese letzten beiden Gleichungen mit (4.15), (4.16), so findet man

$$\begin{aligned} \frac{\partial^2 f}{\partial x_1\partial x_2}(x_1+\xi_4 h_1, x_2+\xi_2 h_2) &= \frac{1}{h_1 h_2}\big(f(x_1+h_1,x_2+h_2) - f(x_1,x_2+h_2) \\ &\qquad -f(x_1+h_1,x_2) + f(x_1,x_2)\big) \\ &= \frac{\partial^2 f}{\partial x_2\partial x_1}(x_1+\xi_1 h_1, x_2+\xi_3 h_2). \end{aligned}$$

Nun reicht es zu beachten, dass sowohl $(x_1+\xi_4 h_1, x_2+\xi_2 h_2)$ als auch $(x_1+\xi_4 h_1, x_2+\xi_2 h_2)$ für $h \to 0$ gegen $x_0 = (x_1, x_2)$ konvergieren. Da aber auch die zweiten gemischten partiellen Ableitungen stetig sind, folgert man, dass schließlich

$$\begin{aligned}\frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) &= \lim_{(h_1,h_2)\to 0} \frac{1}{h_1 h_2}\big(f(x_1+h_1, x_2+h_2) - f(x_1, x_2+h_2) \\ &\qquad - f(x_1+h_1, x_2) + f(x_1, x_2)\big) \\ &= \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_1, x_2),\end{aligned}$$

wie man zeigen wollte. □

ÜBUNGSAUFGABE 114. Es sei $f : \mathbb{R}^m \to \mathbb{R}$ eine 2-mal differenzierbare Funktion, deren 2. partielle Ableitungen stetig sind. Man bezeichnet mit Δf die stetige Funktion von $\mathbb{R}^m$ nach $\mathbb{R}$, die durch

$$\Delta f := \sum_{i=1}^{m} \frac{\partial^2 f}{\partial x_i^2}$$

definiert ist[3]. Löst f die *Laplacesche Gleichung* $\Delta f = 0$, so heißt f *harmonisch.* Zeige: für $d = 2$ bzw. $d = 3$ sind die Funktionen $f_{(2)}$ bzw. $f_{(3)}$ harmonisch, wobei

$$f_{(2)}(x, y) := -\frac{\log\sqrt{x^2+y^2}}{2\pi}, \qquad f_{(3)}(x, y, z) := -\frac{1}{4\pi\sqrt{x^2+y^2+z^2}}.$$

In der Theorie des Elektromagnetismus bzw. der Gravitation wird $f_{(3)}$ *Coulombsches* (nach Charles Augustin de Coulomb) bzw. *Newtonsches Potential* genannt.

ÜBUNGSAUFGABE 115. Es seien f, g 2-mal differenzierbare Funktionen auf einer offenen Menge $A \subset \mathbb{R}^m$. Zeige, dass

$$\Delta(f \cdot g) = \Delta f \cdot g + 2(\mathrm{grad} f, \mathrm{grad} g) + f \cdot \Delta g.$$

ÜBUNGSAUFGABE 116. Folgere aus den Rechenregeln für differenzierbare Funktionen einer Variablen entsprechende Regeln für die partielle Ableitungen einer partiell differenzierbaren Funktion. Schließe auf ähnliche Regeln für die Jacobi-Matrix einer differenzierbaren Funktion.

DEFINITION 117. *Ist $A \subset \mathbb{R}^m$ und $f : A \to \mathbb{R}$, so kann man die $m \times m$-Matrix betrachten, die durch*

$$H_f(x_0) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_0) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_m}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_m \partial x_1}(x_0) & \dots & \frac{\partial^2 f}{\partial x_m^2}(x_0) \end{pmatrix}$$

für alle $x_0 \in A$ definiert wird. Sie ist als Hesse-Matrix *bekannt.*

Das folgende Resultat leitet man unmittelbar aus dem Satz von Schwarz her.

KOROLLAR 4.9. *Unter den Voraussetzung des Satzes von Schwarz ist die Hesse-Matrix $H_f(x_0)$ an jeder Stelle $x_0 \in A$ symmetrisch.*

[3] Die Abbildung $f \mapsto \Delta f$ heißt *Laplace-Operator.* Aus den Rechenregeln für differenzierbare Funktionen folgt, dass Δ linear ist. Vorsicht! Der Operator Δ ist *nicht* beschränkt.

Den Begriff von Maximum und Minimum einer Funktion mehreren Variablen haben wir schon in [**6**, Definition 7.80] eingeführt. Jetzt möchten wir Kriterien zur Bestimmung von Extremstellen vorstellen, welche die ähnlichen Kriterien für Funktionen einer Variable (vgl. das Korollar 1.17) ergänzen.

Definition 118. *Es seien $A \subset \mathbb{R}^m$ eine offene Menge und $f : A \to \mathbb{R}$ eine Fréchet-differenzierbare Funktion. Ist für $x_0 \in A$ $f'(x_0) = 0$, so heißt x_0* kritischer Punkt *von f.*

Zur Erinnerung: Unter diesen Voraussetzungen spricht man oft vom *Gradient von f*. Ein kritischer Punkt für f ist also eine Stelle, an der das Gradient von f verschwindet.

Satz 4.10. *Sei $A \subset \mathbb{R}^m$ eine offene Menge und $f : A \to \mathbb{R}$ eine Fréchet-differenzierbare Funktion. Ist $x_0 \in A$ eine lokale Maximum- oder Minimustelle von f, so ist x_0 ein kritischer Punkt von f.*

Beweis. Ist x_0 eine lokale Maximum- bzw. Minimumstelle für f, so ist x_0 insbesondere eine Maximum- bzw. Minimumstelle für alle Funktionen

$$\phi_i : h \mapsto f(x_0 + he_i), \qquad i = 1, \dots, m,$$

welche – aufgrund der Fréchet-Differenzierbarkeit von f, und deshalb von seiner partiellen Differenzierbarkeit – selber differenzierbar sind. Somit haben alle ϕ_i auch verschwindende Ableitung in $h = 0$. Folglich verschwinden alle partielle Ableitungen von f in x_0. Somit folgt aus dem Satz 4.3, dass $f'(x_0) = 0$. □

Beispiel 119. Wir betrachten wieder die Funktion y aus Beispiel 105. Wir fassen sie aber nur als Funktion der Zeit t und des Abschusswinkels α auf, also $y : (0, \infty) \times (0, \frac{\pi}{2})$. Hat eine solche Funktion zweier Variablen ein Maximum (t_0, α_0), so muß er ein kritischer Punkt sein. Vermöge von (4.14) ist

$$\nabla f(t, \alpha) = \begin{pmatrix} -gt + v\sin(\alpha) \\ vt\cos(\alpha) \end{pmatrix}.$$

Damit $\nabla f(t, \alpha) = 0$ muss insbesondere $t\cos(\alpha) = 0$ sein, und somit $t = 0$ oder $\cos(\alpha) = 0$, also $\alpha = \frac{\pi}{2} + k\pi$ für $k \in \mathbb{Z}$. Da für alle $k \in \mathbb{Z}$ $(0, \frac{\pi}{2} + k\pi)$ nicht im Definitionsbereich der Funktion y liegt, hat die betrachtete Funktion y kein Extremum. □

Die Umkehrung vom Satz 4.10 gilt nicht: es gibt kritische Punkte, welche weder Maximum- noch Minimumstellen sind. Das ist typischerweise der Fall, wenn eine Funktion *in einer Richtung wachsend und in einer anderen Richtung fallend ist.*

Beispiel 120. Betrachte die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, die durch

$$f(x, y) := x^2y^3$$

definiert ist. Diese Funktion ist Fréchet differenzierbar mit

$$\frac{\partial f}{\partial x}(x, y) = 2xy^3, \qquad \frac{\partial f}{\partial y}(x, y) = 3x^2y^2 :$$

da diese beide Funktionen stetig sind, ist f auch Fréchet-differenzierbar. Also verschwindet $\nabla f(x_0, y_0)$ genau dann, wenn $2x_0y_0^3 = 3x_0^2y_0^2 = 0$, also genau dann, wenn $x_0 = 0$ oder $y_0 = 0$. Somit ist $(0, 0)$ ein kritischer Punkt, doch weder eine Maximum- noch eine Minimustelle, denn z.B. entlang der Linie $x = y$ nimmt die Funktion Werte $f(x, x) = x^5$ an, sodass $f(x_1, x_1) < f(0, 0) < f(x_2, x_2)$ für alle x_1, x_2 mit $x_1 < 0 < x_2$. □

Man benötigt also ein weiteres Kriterium, welche die Bestimmung von Extremstellen erlaubt. Dies wird im Folgenden vorgestellt.

Zur Erinnerung: Eine Teilmenge A von einem Vektorraum Y heißt *konvex*, falls für je zwei Elemente $x, y \in A$ auch die Linie, die sie verbindet, in A liegt, d.h. falls

$$sx + (1-s)y \in A \qquad \text{für alle } s \in [0,1].$$

Satz 4.11. *Es sei $A \subset \mathbb{R}^m$ eine offene, konvexe Menge und $f : A \to \mathbb{R}$ eine 2-mal stetig partiell differenzierbare Funktion. Es seien $y, x_0 \in A$. Dann gilt die* Taylorformel

$$f(x_0 + x) = f(x_0) + f'(x_0)(x) + \frac{1}{2} x^T \cdot H_f(x_0) \cdot x + o(\|x\|^2) \qquad (x \to 0).$$

Beweis. Es seien $\varepsilon > 0$ und $i, j = 1, \ldots, m$. Da die zweiten partiellen Ableitungen stetig sind, gibt es $\delta > 0$ mit

$$\left| \frac{\partial^2 f}{\partial x_i \partial x_j} f(x_0 + h) - \frac{\partial^2 f}{\partial x_i \partial x_j} f(x_0) \right| \leq \frac{\varepsilon}{m^2} \qquad \text{für alle } h \text{ mit } \|h\| < \delta. \tag{4.17}$$

Betrachte nun für ein festes $x \in A$ die Funktion

$$\phi : [0,1] \ni t \mapsto f(x_0 + t(x - x_0)) \in \mathbb{R}.$$

Beachte: die Funktion ist überhaupt wohldefiniert, weil $x_0 + t(x - x_0) \in A$ für alle $t \in [0,1]$ ist, d.h., weil A konvex ist. Für f gilt offensichtlich

$$\phi(0) = f(x_0), \qquad \phi(1) = f(x). \tag{4.18}$$

Aufgrund der Konvexität von A ist diese Funktion wohl definiert und, angesichts unserer Voraussetzungen, 2-mal stetig partiell differenzierbar (warum?). Da ϕ aber eine herkömmliche reellwertige Funktion einer reellen Variabel, kann man an ϕ laut dem Satz 1.20 nach Taylor um 0 mit einem Fehlerterm $R_1^{\phi,0}$ entwickeln. Somit erhält man

$$\begin{aligned} \phi(1) &= \phi(0) + \phi'(0) + \tfrac{1}{2}\phi''(\xi) \\ &= \phi(0) + \phi'(0) + \tfrac{1}{2}\phi''(0) + \tfrac{1}{2}\left(\phi''(\xi) - \phi''(0)\right) \end{aligned} \tag{4.19}$$

für ein passendes $\xi \in (0,1)$. Um diesen Ausdruck für f umzuschreiben beachten wir (4.18) sowie, nach der Kettenregel,

$$\phi'(t) = f'(x_0 + t(x - x_0))(x - x_0) = \sum_{j=1}^{m} \frac{\partial f}{\partial x_j}(x_0 + t(x - x_0))(x_j - x_{0j}),$$

wobei die letzte Identität aus dem Satz 4.3 folgt. Diese Summe kann summandenweise partiell differenziert werden, um einen geschickteren Ausdruck für ϕ'' zu finden. Man erhält somit

$$\phi''(t) = \sum_{i,j=1}^{m} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0 + t(x - x_0))(x_j - x_{0j})(x_i - x_{0i}) = (x - x_0)^T \cdot H_f(x_0 + t(x - x_0)) \cdot (x - x_0), \qquad t \in [0,1].$$

Insbesondere kann man nun (4.19) umschreiben:

$$f(x) = f(x_0) + f'(x_0 + t(x - x_0))(x - x_0) + (x - x_0)^T \cdot H_f(x_0) \cdot (x - x_0) + R_2^{f,0}(x),$$

wobei

$$R_2^{f,0}(x) := (x - x_0)^T \cdot (H_f(x_0 + \theta(x - x_0)) - \cdot H_f(x_0)) \cdot (x - x_0).$$

Nun folgt unmittelbar aus (4.17), dass $R_2^{f,0}(x) = o(\|x - x_0\|^2)$. $\square$

Beispiel 121. Betrachte die Funktion $\sqrt{x}\sin(y)$. Um die Taylorentwitcklung von f in $(1, \frac{\pi}{2})$ zu bestimmen berechne die 1. sowie die 2. partiellen Ableitungen:

$$\frac{\partial f}{\partial x}(x,y) = \frac{\sin(y)}{2\sqrt{x}}, \qquad \frac{\partial f}{\partial y}(x,y) = \sqrt{x}\cos(y),$$

und (auch dank dem Satz von Schwarz)

$$\frac{\partial^2 f}{\partial x^2}(x,y) = -\frac{\sin(y)}{4\sqrt{x^3}}, \qquad \frac{\partial^2 f}{\partial x \partial y}(x,y) = \frac{\partial^2 f}{\partial y \partial x}(x,y) = \frac{\cos(y)}{2\sqrt{x}}, \qquad \frac{\partial^2 f}{\partial y^2}(x,y) = -\sqrt{x}\sin(y),$$

also

$$\nabla f\left(1, \frac{\pi}{2}\right) = \left(\frac{1}{2}, 0\right)$$

sowie

$$Hf\left(1, \frac{\pi}{2}\right) = -\begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 1 \end{pmatrix}.$$

Also wird f in $\left(1, \frac{\pi}{2}\right)$ durch

$$f(x,y) = 1 + \frac{x-1}{2} - \frac{1}{8}(x-1)^2 - \frac{1}{2}\left(y - \frac{\pi}{2}\right) + O\left(|(x,y) - \left(1, \frac{\pi}{2}\right)|^3\right)$$

entwickelt. $\square$

Übungsaufgabe 122. Man bestimme die Taylorentwicklung der Funktionen $f, g : \mathbb{R}^2 \to \mathbb{R}$, welche durch

(1) $f(x,y) := \frac{2x-y}{2x+y}$ und
(2) $g(x,y,z) := \sin(xyz)$

definiert sind, in einer Umgebung der Stelle $(\frac{1}{2}, 1)$ bzw. an der Stelle $(0, \pi, 0)$.

Zur Erinnerung: Eine $n \times n$ symmetrische Matrix A heißt *positiv semidefinit*, falls $x^T Ax \geq 0$ für alle $x \in \mathbb{C}^n$. Sie heißt *negativ semidefinit*, falls $-A$ positiv semidefinit heißt. Äquivalent heißt eine Matrix positiv bzw. negativ semidefinit, falls alle ihre Eigenwerte[4] ≥ 0 bzw. ≤ 0 sind. Sie heißt *positiv definit*, falls $x^T Ax > 0$ für alle $x \in \mathbb{C}^n \setminus \{0\}$. Sie heißt *negativ definit*, falls $-A$ positiv definit heißt. Äquivalent heißt eine Matrix positiv bzw. negativ definit, falls alle ihre Eigenwerte > 0 bzw. < 0 sind. Schließlich heißt A *indefinit*, falls A weder positiv noch negativ semidefinit ist.

Satz 4.12. *Es seien $A \subset \mathbb{R}^m$ eine offene Menge, $f : A \to \mathbb{R}$ eine 2-mal stetig partiell differenzierbare Funktion und $x_0 \in A$ ein kritischer Punkt. Dann gelten folgende Aussagen.*

(1) Ist $H_f(x_0)$ negativ bzw. positiv definit, so ist x_0 eine lokale Maximum- bzw. Minimumstelle.
(2) Ist $H_f(x_0)$ indefinit, so ist x_0 keine lokale Maximum- bzw. Minimumstelle.
(3) Ist $x_0 \in A$ eine lokale Maximum- bzw. Minimustelle von f, so ist die Matrix $H_f(x_0)$ negativ bzw. positiv semidefinit.

[4] Die elementaren Eigenschaften von *Eigenwerten* und *Eigenvektoren* einer quadratischen Matrix können in jedem Lehbuch über lineare Algebra gefunden werden, auch der Eintrag `de.wikipedia.org/wiki/Eigenwertproblem` ist vollständig genug für unsere Zwecke. Die Eigenwerte einer Matrix können leicht durch z.B. `maple` oder `mathematica` berechnet werden.

Beweis. Beachte zuerst, dass an der kritischen Stelle x_0 die Taylorformel aus dem Satz 4.11 angesichts vom Satz 4.10 einfach

$$f(x_0 + x) = f(x_0) + \frac{1}{2} x^T \cdot H_f(x_0) \cdot x + R_2^{f,x_0}(x) \tag{4.20}$$

wird, mit $R_2^{f,x_0}(x) = o(\|x\|^2)$ für $x \to x_0$.

(1) Wir betrachten nur dern negativ definiten Fall – der positiv definiter Fall ist ähnlich. Ist $H_f(x_0)$ negativ definit, so ist insbesondere $h^T \cdot H_f(x_0) \cdot h < 0$ für alle $h \in \mathbb{R}^m$ mit $\|h\| = 1$. Diese Teilmenge von $\mathbb{R}^m$ – seine *Einheitssphäre* S_m – ist abgeschlossen und beschränkt, also kompakt. Wir können also die quadratische Form

$$S_m \ni h \mapsto h^T \cdot H_f(x_0) \cdot h \in (-\infty, 0)$$

betrachten, welche – wie jeder auf einer kompakten Menge definierte Funktion – ein Maximum λ_0 annimmt, welches notwendigerweise strikt negativ ist. Somit gilt

$$\frac{1}{2} \frac{x^T}{\|x\|} \cdot H_f(x_0) \cdot \frac{x}{\|x\|} < \lambda_0 < 0 \qquad \text{für alle } x \in \mathbb{R}^m$$

und deshalb wegen (4.20)

$$\frac{1}{\|x\|^2}(f(x_0 + x) - f(x_0)) = \frac{1}{2} \frac{x^T}{\|x\|} \cdot H_f(x_0) \cdot \frac{x}{\|x\|} + \frac{1}{\|x\|^2} R_2^{f,x_0}(x) < \frac{\lambda_0}{4} < 0$$

für alle $x \in \mathbb{R}^n$ mit $0 < \|x\|$ klein genug: das ist möglich, weil in einer Umgebung von x_0 $R_2^{f,x_0}(x)$ schneller als $\|x\|^2$ abfällt. Daher ist x_0 ein lokales Maximum.

(2) Da die Hesse-Matrix $H_f(x_0)$ indefinit ist, gibt es (mindestens) zwei Eigenwerte $\lambda_\pm$ mit zugehörigen Eigenvektoren $h_\pm$ so, dass $h_+^T \cdot H_f(x_0) \cdot h_+ > 0$ bzw. $h_-^T \cdot H_f(x_0) \cdot h_- < 0$. Somit gilt für $t \in \mathbb{R}$ mit $0 < |t|$ klein

$$\frac{1}{t^2}(f(x_0 + th_\pm) - f(x_0)) = \frac{1}{2}\lambda_\pm + \frac{1}{t^2} R_2(th_\pm).$$

Wir folgern, dass $f(x_0 + h) - f(x_0)$ in jeder Umgebung von x_0 sowohl positive als auch negative Werte annimmt, d.h., f kann an der Stelle x_0 weder ein Maximum noch ein Minimum haben.

(3) Ist x_0 eine lokale Minimumstelle und wäre $H_f(x_0)$ dennoch nicht positiv semidefinit, so gäbe es $v \in \mathbb{R}^n$ mit $v^T \cdot H_f(x_0) \cdot v < 0$. Für die Abbildung

$$\phi : t \mapsto f(x_0 + tv)$$

würde dann

$$\frac{\partial^2 f}{\partial v^2}(x_0) = \frac{1}{v_j^2} \phi''(0)$$

gelten. Da aber

$$\frac{\partial^2 f}{\partial v^2}(x_0) = v^T \cdot H_f(x_0) \cdot v < 0$$

können wir anhand vom Korollar 1.17 schließlich folgern, dass ϕ an der Stelle 0 ein Maximum hat – was die Tatsache widerspricht, dass x_0 eine Minimumstelle von f ist. □

Übungsaufgabe 123. (1) Betrachte die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, die durch $f(x, y) = x^2 - xy$ definiert wird. Hat die Funktion lokale Maxima oder Minima?

(2) Betrachte die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, die durch $f(x, y) = e^x(1 - \cos y)$ definiert wird. Berechne die Hesse-Matrix von f und untersuche f auf Extrema an den Stellen $(0, 0)$, $(\pi, 0)$ und $(0, \pi)$.

(3) Betrachte die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, die durch $f(x, y) = -\cos(x^2 y^2)$ definiert wird. Berechne die Hesse-Matrix von f und untersuche f auf Extrema.

Der Folgende ist eine mehrdimensionale Verallgemeinerung des vektorwertigen Satzes von Lagrange (vgl. [**6**, Satz 8.39] und wird gelegentlich *Schrankensatz* genannt.

Satz 4.13. *Es sei $A \subset \mathbb{R}^m$ und $f : A \to \mathbb{R}^n$ Fréchet-differenzierbar. Dann gilt für alle kompakte und konvexe $B \subset A$*

$$\|f(x) - f(y)\|_{\mathbb{R}^n} \leq \sup_{\xi \in A} \|f'(\xi)\| \|x - y\|_{\mathbb{R}^m}, \qquad x, y \in B,$$

wobei $\|f'(\xi)\|$ die Norm von $f'(\xi)$ als Operator von $\mathbb{R}^m$ nach $\mathbb{R}^n$ bezeichnet.

Insbesondere ist die Einschränkung von f auf B Lipschitz-stetig.

Übungsaufgabe 124. Beweise den Satz 4.13. *(Hinweis: Stelle f als Funktion einer Variabel dar und verwende dann* [**6**, Satz 8.39]*).*

KAPITEL 5

Einige Anwendungen der Differenzialrechnung

Dieses Kapitel ist zwei der wichtigsten Sätzen der Analysis 2 gewidmet. Wir folgen den Zugang in [**4**] und erwähnen nur, dass man in [**2**, § VII.8] eine interessante aber fortgeschrittenere darstellung des unendlichdimensionalen Falls finden kann.

Als erste benötigen wir ein Hilfsresultat, das 1922 von Stefan Banach bewiesen wurde und als *Banachscher Fixpunktsatz* bekannt ist.

Definition 125. *Es seien X, Y Banachräume. Dann heißt $T : Y \to X$ eine* Kontraktion*, falls es $\alpha \in (0,1)$ gibt mit*

$$\|Tx - Ty\|_X \leq \alpha \|x - y\|_Y \qquad \forall x, y \in X.$$

Lemma 5.1. *Es seien X ein Banachraum, A eine abgeschlossene Teilmenge von X und $T : A \to A$ eine Kontraktion. Dann hat T genau einen Fixpunkt, d.h., es gibt genau ein $x \in A$ mit $Tx = x$.*

Beweis. Es sei $x_0 \in A$ (beliebig) und betrachte die Folge $(x_k)_{k\in\mathbb{N}}$, welche durch

$$x_k := T^k x_0, \qquad k \in \mathbb{N}$$

definiert ist. Dann gilt für alle $m, n \in \mathbb{N}$ $x_k \in A$ und darüber hinaus

$$\|x_n - x_m\| = \|T^n x_0 - T^m x_0\| \leq \alpha \|T^{n-1} x_0 - T^{m-1} x_0\| \leq \ldots \leq \alpha^n \|x_0 - T^{m-n} x_0\|$$

und somit

$$\begin{aligned} \|x_0 - T^{m-n} x_0\| &\leq \|x_0 - Tx_0\| + \|Tx_0 - T^2 x_0\| + \ldots + \|T^{m-n-1} x_0 - T^{m-n} x_0\| \\ &\leq \sum_{k=0}^{m-n-1} \alpha^k \|x_0 - Tx_0\| \\ &< \frac{1}{1-\alpha} \|x_0 - Tx_0\|. \end{aligned}$$

Daher

$$\|x_n - x_m\| \leq \frac{\alpha^n}{1-\alpha} \|x_0 - Tx_0\|.$$

Da der rechte Term gegen 0 strebt, ist also $(x_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge – und somit gegen ein $x \in X$ konvergent, da X vollständig ist. Ist aber jedes $x_n \in A$ und ist $(x_n)_{n\in\mathbb{N}} \subset A$, so ist auch $x \in A$, da A abgeschlossen ist. Weiter gilt für alle gegen x konvergente Folgen $(x_n)_{n\in\mathbb{N}}$

$$Tx = \lim_{n\to\infty} Tx_n = im_{n\to\infty} x_{n+1} = x.$$

Schließlich zeigen wir die Eindeutigkeit des Fixpunktes. Gäbe es ein weiterer Fixpunkt y, d.h., gelte es $Ty = y$ für einen $y \in X$, so wäre

$$\|x - y\| = \|Tx - Ty\| \leq \alpha \|x - y\|,$$

d.h. $\|x-y\| = 0$ und daher $x = y$. □

Das folgende fundamentales Resultat wurde 1907 von Ulisse Dini bewiesen.

SATZ 5.2. *Es seien $A_1 \subset \mathbb{R}^m$ und $A_2 \subset \mathbb{R}^n$ offen, $a = (a_1, a_2) \in A_1 \times A_2$ und $f : A_1 \times A_2 \to \mathbb{R}^n$ eine stetig differenzierbare Funktion*[1]*. Gilt*

- *$f(a) = 0$ und*
- *$J_{f_2}(a)$ ist invertierbar,*

Dann gibt es offene Umgebungen B_1 und B_2 jeweils von a_1 und a_2 mit $B_1 \subset A_1$ und $B_2 \subset A_2$ und eine stetig differenzierbare Funktion $g : B_1 \to B_2$ mit

- *$g(a_1) = a_2$,*
- *$f(x, g(x)) = 0$ für alle $x \in B_1$ und umgekehrt*
- *$y = g(x)$ für alle $(x, y) \in V_1 \times V_2$ mit $f(x, y) = 0$ und schließlich*
- *$J_g(x) = -J_{f_2}(x, g(x))^{-1} J_{f_1}(x, g(x))$ für alle $x \in B_1$.*

Dieser ist als *Satz über die implizite Funktion* bekannt: Da g in der obigen Aussage nur als (lokale) Lösung der Gleichung

$$f(x, y) = 0$$

ist, spricht man von "implizit definierten Funktion". Im Beweis werden wir durchgehend die Aussage in der Übungsaufgabe 97 verwenden, welche uns erlaubt, uns auf keine Norm auf $\mathbb{R}^n$ oder $\mathbb{R}^m$ festzulegen.

BEWEIS. Nach unserer Voraussetzungen ist die Matrix $J_{f_2}(a)$ invertierbar und somit können wir eine neue Abbildung

$$G : A_1 \times A_2 \ni (x, y) \mapsto y - J_{f_2}(a)^{-1} f(x, y) \in \mathbb{R}^n,$$

welche offensichtlich folgende Eigenschaften erfüllt:

(E1) für alle $(x, y) \in A_1 \times A_2$ gilt (Dank der Invertierbarkeit von $J_{f_2}(a)^{-1}$) genau dann $y = G(x, y)$, wenn $f(x, y) = 0$;

(E2) Die vektorwertige Abbildung G ist als Summe und Verkettung stetiger Funktionen selber stetig;

(E3) Dank der Kettenregel und dem Beispiel 94 ist G stetig partiell (bezüglich jeder ihrer Koordinate, die gemeinsam mit y bezeichnet wurden) differenzierbar mit

$$\frac{\partial G}{\partial y}(x, y) = \mathbf{1} - J_{f_2}(a)^{-1} \frac{\partial f}{\partial y}(x, y);$$

(E4) Insbesondere gilt

$$\frac{\partial G}{\partial y}(0, 0) = \mathbf{1} - J_{f_2}(a)^{-1} \frac{\partial f}{\partial y}(0, 0) = \mathbf{1} - J_{f_2}(a)^{-1} J_{f_2}(a) = 0.$$

Dabei bezeichnet $\mathbf{1}$ die identische $n \times n$-Matrix. Aufgrund von (E3) kann man eine offene Umgebung $C_1 \times C_2 \subset A_1 \times A_2$ von $(0, 0)$ so, dass

$$\left\| \frac{\partial G}{\partial y}(x, y) \right\| \le \frac{1}{2} \qquad \text{für alle } (x, y) \in C_1 \times C_2. \tag{5.1}$$

[1] Beachte das Format der Funktion: die Jacobi-Matrix vof f ist, an der Stellen wo f partiell Differenzierbar ist, eine $(m+n) \times n$-Matrix. Wir wollen diese Jacobi-Matrix darstellen als die Nebeneinanderstellung zweier Matrizen, die als Jacobi-Matrizen der Einschränkungen von f auf der 1. bzw. 2. Koordinate seines Arguments angesehen werden können, die jeweils Funktionen von A_1 bzw. A_2 nach $\mathbb{R}^n$ sind. Diese Beide Jacobi-Matrizen bezeichnen wir mit J_{f_1} bzw. J_{f_2}

Es wird behilflich sein, dass wir $C_1 \times C_2$ klein genug wählen, dass zusätzlich gilt:

$$(5.2) \qquad J_{f_2}(x,y) \text{ ist invertierbar,} \qquad \text{für alle } (x,y) \in C_1 \times C_2.$$

Es sei nun $r > 0$ klein genug, dass die Kugel $B_2 := \{y \in \mathbb{R}^n : \|y\| \leq r\} \subset \mathbb{R}^n$ mit Radius r in C_2 enthalten ist, und verwende dieses r und (E2), um eine offene Umgebung B_1 zu finden, so dass

$$(5.3) \qquad \max_{x \in \overline{B_1}} \|G(x,0)\| =: \varepsilon \leq \frac{r}{2}.$$

Da G stetig differenzierbar ist, ist es nach dem Satz 4.13 auch Lipschitzstetig mit Lipschitzkonstante, die kleiner als die Norm ihrer Fréchet-Ableitung ist: Nach (5.1) gilt somit

$$(5.4) \qquad \|G(x,y) - G(x,z)\| \leq \frac{1}{2}\|y - z\| \qquad \text{für alle } (x,y),(x,z) \in C_1 \times C_2.$$

Da C_2 eine Umgebung der 0 ist, kann man $z = 0$ wählen und schließlich auch angesichts von (5.3)

$$(5.5) \quad \|G(x,y)\| \leq \|G(x,y) - G(x,0)\| + \|G(x,0)\| \leq \frac{1}{2}\|y - 0\| + \frac{r}{2} \leq r \qquad \text{für alle } (x,y) \in B_1 \times B_2,$$

da B_2 genau die Kugel mit Radius r ist.

Das heißt insbesondere, dass für alle $x \in B_1$ fest kann man G – als Funktion allein ihrer zweiten Koordinate – als Kontraktion von B_2 nach B_2 ansehen: Wir verwenden dabei (5.4) und (5.5). Der Banachscher Fixpunktsatz liefert nun die Existenz eines Fixpunktes, die i.A. von x abhängen wird, d.h.: für alle $x \in B_1$ gibt es $g(x) := y \in B_2$ mit $G(x,y) = y$, was nach (E1) heißt, dass $f(x, g(x)) = 0$. Dieses $g : B_1 \to B_2$ ist die von uns gesuchte Abbildung. Wir wollen jetzt überprüfen, dass g alle Eigenschaften hat, die wir behaupten. Um anzufangen zeigen wir, dass g stetig ist.

Dies wird durch eine weitere Anwendung des Banachschen Fixpunktsatzes bewiesen. Wir werden ihn auf $C_b(B_1; \mathbb{R}^n)$ anwenden, der nach dem Korollar 1.2 wohl ein Banachraum ist. Wir brauchen dazu eine Kontraktion Φ auf $C_b(B_1; \mathbb{R}^n)$ zu definieren. Die Konstruktion von Φ erfolgt so: Ist ϕ Element der Kugel von $\mathbb{R}^m$ von Radius r sowie auch von B_1 (ihr Durschnitt ist nichtleer, da z.B. beide die 0 enthalten), so kann man dazu die Funktion

$$\psi : x \to G(x, \phi(x))$$

zuordnen, d.h., wir wollen einen Φ durch $\Phi(\phi) := \psi$ definieren. Wir wollen noch zeigen, dass Φ eine Kontraktion auf der Kugel A vom Radius r in $C_b(B_1; \mathbb{R}^n)$ ist: Denn haben wir es bewiesen, so gibt es nach dem Banaschschen Fixpunktsatz genau ein $g \in A$ mit $\Phi(g) = g$, und aufgrund der Definition von Φ gilt somit

$$G(x; g(x)) = g(x) \qquad \text{und somit} \qquad F(x; g(x)) = 0, \qquad \text{für alle } x \in V_1.$$

Da auch das oben eingeführte g die selbe Eigenschaft hat, müssen die beiden g tatsächlich übereinstimmen. Insbesondere ist die oben definierte Funktion Element von $C_b(B_1; \mathbb{R}^n)$, was zum Beweis unseres Beweises ausreicht. Lass uns somit schließlich nur zeigen, dass Φ eine Kontraktion von A nach A bildet. Sind nämlich $f_1, f_2 \in A$, so gilt definitionsgemäß

$$\|\Phi(\phi_1) - \Phi(\phi_2)\|_\infty = \sup_{x \in B_1} \|\psi_1(x) - \psi_2(x)\|_\infty = \sup_{x \in B_1} \|G(x, \phi_1(x)) - G(x, \phi_2(x))\|_\infty \leq \frac{1}{2}\|\phi_1(x) - \phi_2(x)\|_\infty,$$

angesichts von (5.4), wie wir zeigen wollten.

Wir führen den Beweis der Differenzierbarkeit von g punktweise wie folgt, erstmal für die Stelle $0 \in \mathbb{R}^m$. Da F an der Stelle $0 \in \mathbb{R}^{m+n}$ differenzierbar ist, gilt nach dem Lemma 4.1

$$f(x,y) = J_{f_1}(0)x + J_{f_2}(0)y + R_0(x,y), \qquad (x,y) \in A_1 \times A_2,$$

mit $R_0(x,y) = o(\|(x,y)\|_{\mathbb{R}^{m+n}})$ für $(x,y) \to 0$. Inbsesondere hat man für jedes $x_\in B_1$ $f(x,g(x)) = 0$, und somit liefert die obige Formel

$$0 = f(x,g(x)) = J_{f_1}(0)x + J_{f_2}(0)g(x) + R_0(x,g(x)), \qquad x \in B_1,$$

oder äquivalent

$$g(x) = -J_{f_2}(0)^{-1}J_{f_1}(0)x - J_{f_2}(0)^{-1}R_0(x,g(x)), \qquad x \in B_1. \tag{5.6}$$

Können wir zeigen, dass

$$J_{f_2}(0)^{-1}R_0(x,g(x)) = o(\|x\|) \qquad (x \to 0), \qquad \text{für alle } x \in B_1, \tag{5.7}$$

so haben wir – wieder Dank Lemma 4.1 – bewiesen, dass g an der Stelle 0 differenzierbar ist: Den Beweis an den anderen Stellen von B_1 kann nun ähnlich geführt werden.

Um (5.7) nun zu beweisen beachte also jetzt, dass aufgrund von $R_0(x,y) = o(\|(x,y)\|_{\mathbb{R}^{m+n}})$ für $(x,y) \to 0$ gibt es zu $\varepsilon := \frac{1}{2\|J_{f_2}(0)^{-1}\|}$ eine offene Umgebung $D \subset B_1 \times B_2$ von $0 \in \mathbb{R}^{n+m}$ mit

$$\|R_0(x,y)\| \le \varepsilon\|(x,y)\| \le \varepsilon(\|x\| + \|y\|) \qquad \text{für alle } (x,y) \in D.$$

Da g insbesondere an der Stelle $0 \in \mathbb{R}^m$ stetig ist, gibt es eine offene Umgebung $\tilde{D} \subset B_1$, so dass $(x,g(x)) \in D$ für alle $x \in \tilde{D}$, und somit

$$\|R_0(x,g(x))\| \le \frac{1}{2\|J_{f_2}(0)^{-1}\|}(\|x\| + \|g(x)\|) \qquad \text{für alle } x \in \tilde{D}.$$

Angesichts von (5.6) findet man nun für alle $x \in \tilde{D}$

$$\begin{aligned} \|g(x)\| &\le \|J_{f_2}(0)^{-1}J_{f_1}(0)\|\|x\| + \|J_{f_2}(0)^{-1}\|\|R_0(x,g(x))\| \\ &\le \|J_{f_2}(0)^{-1}J_{f_1}(0)\|\|x\| + \frac{\|J_{f_2}(0)^{-1}\|}{2\|J_{f_2}(0)^{-1}\|}(\|x\| + \|g(x)\|) \\ &\le \left(\|J_{f_2}(0)^{-1}J_{f_1}(0)\| + \frac{1}{2}\right)\|x\| + \frac{1}{2}\|g(x)\|. \end{aligned}$$

Diesen Ausdruck kann man als

$$\|g(x)\| \le 2\left(\|J_{f_2}(0)^{-1}J_{f_1}(0)\| + \frac{1}{2}\right)\|x\|, \qquad x \in \tilde{D}, \tag{5.8}$$

umschreiben, was wiederum liefert, dass

$$R_0(x,g(x)) = o(\|(x,g(x))\|) = o(\|x\| + \|g(x)\|) = o(\|x\|) \qquad (x \to 0),$$

wobei die letzte (asymptotische) Identität eben von (5.8) folgt.

Noch wollen wir zeigen, dass g *stetig* differenzierbar ist, d.h., dass die Fréchet-Ableitung von g eine stetige Abbildung. Dazu verwenden wir, dass nach Konstruktion $B_1 \times B_2 \subset A_1 \times A_2$, und dass nach (5.2) J_{f_2} somit überall auf $B_1 \times B_2$ invertierbar ist. Darüber hinaus sehen wir durch Differenzieren von $f(x,g(x)) = 0$, dass nun aus der Kettenregel

$$J_{f_1}(x,g(x)) + J_{f_2}(x,g(x))J_g(x) = 0, \qquad x \in B_1,$$

folgt, und somit

$$J_g(x) = -J_{f_2}(x, g(x))^{-1} J_{f_1}(x, g(x)), \qquad x \in B_1. \tag{5.9}$$

Daher ist J_g als Verkettung stetiger Funktionen stetig, und stimmt darüber hinaus mit der Fréchet-Ableitung von g ein Dank dem Satz 4.7. Anders gesagt: g ist tatsächlich stetig differenzierbar.

Dadurch ist der Beweis vollständig. □

ÜBUNGSAUFGABE 126. Betrachte den obigen Beweis sorgfältig und zeige, dass die Existenz einer stetigen impliziten Funktion g bereits aus der Annahme, dass schlicht f und $J_{f_2}(a)$ stetig sind, folgt; und dass die Annahme über die stetige Differenzierbarkeit von f erst verwendet wird, um die stetige Differenzierbarkeit von g zu beweisen.

BEISPIEL 127. Es sei

$$f : \mathbb{R}^2 \ni (x, y) \mapsto x^2 + y^2 - 1 \in \mathbb{R}.$$

Dann verschwindet f in den Punkten, die Euklidischen Abstand 1 von 0 haben: d.h., f ist der analytische Ausdruck des Einheitskreises. Diese Funktion is stetig partiell differenzierbar mit Jacobi-Matrix

$$J_f(a, b) \equiv (J_{f_A}(a, b), J_{f_B}(a, b))) = (2a, 2b).$$

Somit ist $J_{f_B}(a, b)) = 2b$ genau dann invertierbar $b \neq 0$. Der Satz über die implizite Funktion erlaubt nun die lokale!) Darstellung des Einheitskreises durch $y = g(x)$ für alle Nullstellen $(x, y) \in A_1 \times A_2$, für die $y = 0$, d.h., genau dann, wenn $(x, y) = (\pm 1, 0)$. □

Der Folgende stellt eine elementare und doch sehr wichte Anwendung des Satzes über die implizite Funktion dar. Er ist als *Satz über die lokale Inverse* bekannt.

SATZ 5.3. *Es sei $A \subset \mathbb{R}^n$ offen, $a \in A$ und $f : U \to \mathbb{R}^n$ eine stetig differenzierbare Funktion. Dann gelten die folgenden Aussagen.*

(1) Bezeichne mit $\tilde{A}$ das Bild von A unter f, welches notwendigerweise offen ist (warum?). Ist $f : A \to \tilde{A}$ invertierbar mit stetig differenzierbarer Inverse, so ist $J_f(x)$ invertierbar für alle $x \in A$.

(2) Es sei $J_f(a)$ invertierbar. Dann gibt es eine offene Umgebung B_1 von a und B_2 von $f(a)$, mit $B_1 \subset A$, so dass die Einschränkung $f|_{B_1}$ eine bijektion von B_1 nach B_2 ist mit stetig differenzierbarer Inverse $g : B_2 \to B_1$, welche $D_g(f(a)) = D_f(a)^{-1}$ erfüllt.

BEWEIS. (1) Die Aussage folgt aus der Kettenregel und der Tatsache, dass $f \circ f^{-1} = \mathrm{id}$, so dass

$$J_{f^{-1}}(f(x)) J_f(x) = \mathbf{1},$$

wobei $\mathbf{1}$ die identische $n \times n$-Matrix bezeichnet.

(2) Betrachte die Funktion F, welche durch

$$F : \mathbb{R}^n \times A \ni (x, y) \mapsto x - f(y) \in \mathbb{R}^n$$

definiert wird. Dann gilt (mit dem selben Notationskonventionen wie im Satz 5.2)

- $F(f(a), a) = 0$ und
- $J_{F_2})(x, y) = -J_f(y)$, und insbesondere ist $J_{F_2})(f(a), a) = -J_f(a)$ invertierbar.

Somit können wir den Satz 5.2) auf F anwenden und schließen, dass es Umgebungen $C_1 \subset A$ von a und C_2 von $f(a)$ sowie eine stetig differenzierbare Funktion $g : C_2 \to C_1$ mit folgenden Eigenschaften:

- $0 = F(x, g(x)) = x - f(g(x))$, und somit ist $f(g(x)) = x$ für alle $x \in C_1$, sowie
- $y = g(x)$ für alle $(x, y) \in C_1 \times C_2$ mit $f(x, y) = 0$.

Da aber f stetig ist, gibt es $B_1 \subset C_1$ offene Umgebung von a so, dass $f(B_1) \subset C_1$, und weiter gilt angesichts der zweiten obigen Eigenschaft

$$f(B_1) = g^{-1}(B_1)$$

Ist nun $B_2 := f(B_1)$, so ist B_2 wegen der Stetigkeit von f eine offene Umgebung von $f(a)$. Wir schließen, dass die Einschränkung $f|_{B_1} \to B_2$ eine bjjektive Funktion mit Inverse g ist. Damit ist der Beweis vollständig. □

Zum Schluss betrachten wir eine andere Fragestellung: Kann man eine Maximum- oder Minimumstelle einer Funktion $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, welche gleichzeitig eine oder mehrere weitere Bedingungen erfüllt, finden? Eine Teillösung wurde von Joseph-Louis Lagrange gefunden: Er löste damit 1788 manche konkreten Probleme der Mechanik und formuliert 1797 das folgende allgemeine Prinzip. Interessanterweise kann man den Beweis als Folgerung des Satzes über die implizite Funktion.

Definition 128. *Es seien $A \subset \mathbb{R}^m \times \mathbb{R}^n$ eine offene Menge und $f : A \to \mathbb{R}$ sowie $F : A \to \mathbb{R}^n$. Dann heißt $(x_0, y_0) \in A$ ein* lokales bedingtes Maximum *(bzw.* Minimum*) – genauer: unter den* von F gegebenen Bedingungen *–, falls es eine offene Umgebung $U_0 \subset A$ von (x_0, y_0) gibt, so dass*

$$F(x, y) = 0 \qquad \text{und} \qquad f(x, y) \leq f(x_0, y_0) \ \ (\text{bzw. } f(x, y) \geq f(x_0, y_0) \qquad \text{für alle } (x, y) \in U_0.$$

Wohl gemerkt: Die von F gegebenen Bedingungen können leer (z.B. wenn $F \equiv 0$) oder redundant (Z.B. wenn alle Komponenten von F gleich sind) sein. Diesen Fall wollen wir meiden, indem wir eine linearalgebraische Bedingung über die Fréchet-Ableitung von F stellen.

Definition 129. *Es seien $A \subset \mathbb{R}^m \times \mathbb{R}^n$ eine offene Menge und $F : A \to \mathbb{R}^n$ eine stetig differenzierbare Funktion. Die Menge, die durch*

$$M_F := \{(x, y) \in A : F(x, y) = 0\}$$

*definiert ist, heißt $m + n - k$-*dimensionale Untermannigfaltigkeit (von $\mathbb{R}^{n+m}$)*, $1 \leq k \leq n$, falls*

$$\text{Rang } J_F(x, y) = k \qquad \text{für alle } x \in M_F.$$

(Ist die Abhängigkeit von F klar, so werden wir manchmal einfach M statt M_F schreiben). Ist $a \in M$, so heißt ein $z \in \mathbb{R}^n$ dann Tangentialvektor von M an der Stelle a*, falls es für ein $\delta > 0$ eine stetig differenzierbare Funktion $\gamma : (-\delta, \delta) \to M$ gibt, die*

$$\gamma(0) = a \qquad \text{und} \qquad \gamma'(0) = z$$

erfüllt. Endlich heißt $N \in \mathbb{R}^n$ Normalenvektor an M_F an der Stelle a*, falls N* senkrecht *auf jedem Tangentialvektor steht, d.h.,*

$$(z|N)_{\mathbb{R}^n} = \sum_{i=1}^{n} z_i N_i = 0 \qquad \text{für alle Tangentialvektoren } z.$$

Die Menge aller Tangentialvektoren (bzw. aller Normalenvektoren) einer Untermannigfaltigkeit M an einer Stelle $a \in M$ heißt Tangentialraum *(bzw.* Normalenraum*) von M an der Stelle a und wird mit $T_a M$ (bzw. $N_a M$) bezeichnet.*

Beispiel 130. Es sei

$$F : \mathbb{R}^2 \ni (x, y) \mapsto x^2 + y^2 - 1 \in \mathbb{R}.$$

Dann ist F stetig Fréchet differenzierbar mit

$$J_F(x,y) = \begin{pmatrix} 2 & 2 \end{pmatrix},$$

welches Rang 1 hat. Somit ist der Einheitskreis

$$\{(x,y) \in \mathbb{R}^2 : F(x,y) = 0\}$$

eine 1-dimensional Untermannigfaltigkeit von $\mathbb{R}^2$.

Beispiel 131. Es sei

$$F : \mathbb{R}^2 \ni (x,y) \mapsto (x, y, x^2 + y^2) \in \mathbb{R}^3.$$

Dann ist F stetig Fréchet differenzierbar mit

$$J_F(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 2 \end{pmatrix},$$

welches Rang 2 hat. Somit ist

$$\{(x,y) \in \mathbb{R}^2 : F(x,y) = 0\}$$

eine 0-dimensional Untermannigfaltigkeit von $\mathbb{R}^2$. Wie sieht sie aus?

Übungsaufgabe 132. Es sei

$$F : \mathbb{R}^3 \ni (x,y,z,\theta) \mapsto (\cos\theta - x, \sin\theta - y, z - \theta) \in \mathbb{R}^3.$$

Welche Dimension hat die Untermannigfaltigkeit

$$\{x \in \mathbb{R}^3 : F(x,y,z,\theta) = 0\}$$

und wie sieht sie aus?

Übungsaufgabe 133. Es sei $A \subset \mathbb{R}^n$. Allgemeiner sagt man, dass ein Vektor $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ senkrecht auf A steht, falls

$$\sum_{j=1}^{n} x_j a_j = 0 \qquad \text{für alle } a = (a_1, \ldots, a_n) \in A.$$

Die Menge aller Vektoren von $\mathbb{R}^n$, die auf jedem Vektor von A senkrecht stehen, bezeichnet man mit $A^\perp$. Zeige: Für alle $A \subset \mathbb{R}^n$ ist $A^\perp$ ein Vektorraum.

Anders gesagt ist eine k-dimensionale Mannigfaltigkeit die höherdimensionale Verallgemeinerung einer Kurve ($k = 2$) oder einer Fläche ($k = 3$). Somit ist T ist ein Tangentialvektor, wenn es die Tangente am Graph einer Funktion γ darstellt, die eine Teilmenge von M_F ist.

Übungsaufgabe 134. Es seien $n, k \in \mathbb{N}$ mit $1 \leq k < n$, U eine offene Teilmenge des $\mathbb{R}^n$ und M eine $(n-k)-$dimensionale Untermannigfaltigkeit der Form

$$\{x \in U : F(x) = 0\}$$

für eine stetig Fréchet-differenzierbaren Funktion $F : U \to \mathbb{R}^m$, deren Jacobi-Matrix die Rangbedingung

$$\operatorname{rang} J_F(x) = k \qquad \text{für alle } x \in M$$

erfüllt. Desweiteren sei $a := (a', a'') \in M$ mit $a' \in \mathbb{R}^{n-k}$, $a'' \in \mathbb{R}^k$ und es sei o.B.d.A. angenommen, dass die Matrix $\frac{\partial F}{\partial(x_{n-k+1}, \ldots, x_n)}(a)$ der partiellen Ableitungen von F nach den Koordinaten $x_{n-k+1}, ..., x_n$ in a bereits den Rang k besitzt.

Zeige, dass es offene Umgebungen $V_1 \subset \mathbb{R}^{n-k}$ von a', $V_2 \subset \mathbb{R}^k$ von a'' und eine stetig Fréchet-differenzierbare Funktion $g : V_1 \to V_2$ mit $g(a') = a''$ existieren, so dass die Abbildung

$$X : V_1 \ni x' \mapsto \big(x', g(x')\big) \in V_1 \times V_2$$

folgende Eigenschaften erfüllt:

(1) X bildet V_1 bijektiv auf $M \cap (V_1 \times V_2)$ ab mit $X(a') = a$.

(2) X ist stetig Fréchet-differenzierbar und

$$\text{rang } \frac{\partial X}{\partial(e_1, ..., e_{n-k})}(t) = n - k \quad \text{für alle } t \in V_1, \tag{5.10}$$

wobei wir mit

$$\frac{\partial X}{\partial(e_1, ..., e_{n-k})}(t)$$

die Matrix, die aus den ersten $n-k$ Spalten der Jacobi-Matrix $J_F(t)$ besteht, bezeichnen.

(3) Beobachte insbesondere, dass die letzte Bedingung impliziert, dass es eine Funktion $\xi \equiv (\xi_1, \dots, \xi_k) : V_1 \times V_2 \to \mathbb{R}^k$ gibt, für die

$$M \cap (V_1 \times V_2) = \{x \in V_1 \times V_2 : \xi(x) = 0\}$$

und

$$\text{rang } J_\xi(x') = n - k.$$

Alle Funktionen X, welche diese Eigenschaften erfüllen, heißen dann *lokale Koordinantensysteme* von M an der Stelle a.

Unseres Ziel ist es nun, eine notwendige Bedingung für die Existenz eines bedingten Minimums oder Maximums zu finden. Im Folgenden verwenden wir die obigen Notationen.

Wir wollen zuerst die Definition von Tangential- bzw. Normalen*raum* begründen.

Lemma 5.4. *Es seien $n, k \in \mathbb{N}$, $1 \le k < n$, $U \subset \mathbb{R}^n$ offen und M_F eine $(n-k)-$dimensionale Untermannigfaltigkeit der Form*

$$\{x \in U : F(x) = 0\}$$

mit einer stetig Fréchet-differenzierbaren Funktion $F : U \to \mathbb{R}^k$. Dann sind sowohl T_aM als auch N_aM Vektorräume – der Dimension $n-k$ bzw. k.

Es sei X ein lokales Koordinatensystem von M an der Stelle a wie in der Übungsaufgabe 134 und $t \in V_1$ mit $X(t) = a$. Dann bilden die Vektoren

$$\frac{\partial X}{\partial e_1}(t), \dots, \frac{\partial X}{\partial e_{n-k}}(t)$$

(d.h.: die ersten $n-k$ Spalten der Jacobi-Matrix von X) eine Basis von T_aM.

Wiederum bilden die Vektoren

$$F_1'(a), \dots, F_k'(a)$$

eine Basis von N_aM.

Beweis. Es reicht aus, zu zeigen, dass die Vektoren

$$\frac{\partial X}{\partial e_1}(t), \dots, \frac{\partial X}{\partial e_{n-k}}(t)$$

(welche aufgrund von (5.10) linear unabhängig sind), bzw.

$$\xi_1'(a), \ldots, \xi_k'(a)$$

zwei Vektorräume der Dimensionen $n-k$ bzw. k aufspannen, etwa Z_T bzw. Z_N, welche in T_aM bzw. N_aM enthalten sind (uns insbesondere steht Z_T zu Z_N senkrecht): Denn dann gilt

$$\mathbb{R}^n = Z_T \oplus Z_N \subset T_aM \oplus N_aM$$

während $T_aM \cup N_aM$ eine Teilmenge von $\mathbb{R}^n$ ist, und somit werden alle Inklusionen sogar zu Identitäten (komponentenweise).

Es sei $z \in Z_T$, welches nach Konstruktion von der Form

$$z = \sum_{i=1}^{n-k} c_i \frac{\partial X}{\partial e_i}(t)$$

ist für geeignete Koeffizienten $c_1, \ldots, c_{n-k}$, d.h., für einen Koeffizientenvektor $c \in \mathbb{R}^k$. Definitionsgemäß wollen wir eine Umgebung $(-\delta, \delta)$ une eine Funktion $\gamma : (-\delta, \delta) \to M$ finden, für die $\gamma(0) = a$ und $\gamma'(0) = z$. Betrachte dazu $\delta > 0$ klein genug, dass die Abbildung

$$\gamma : (-\delta, \delta) \ni s \mapsto X(t + sc) \in M$$

wohl definiert ist. Aufgrund der Fréchet-Differenzierbarkeit von X und angesichts der Kettenregel finden wir, dass auch γ differenzierbar ist mit

$$\frac{d\gamma}{ds}(s) = X'(t + sc) \cdot c = \sum_{j=1}^{n-k} c_j \frac{\partial X}{\partial e_j} t + sx), \qquad s \in (-\delta, \delta),$$

und insbesondere

$$\frac{d\gamma}{ds}(0) = \sum_{j=1}^{n-k} c_j \frac{\partial X}{\partial e_j}(t) = z,$$

wie wir zeigen wollten.

Lass uns nun zeigen, dass $Z_N \subset N_aM$ – dass N_aM ein Vektorraum ist, folgt unmittelbar aus der Definition von T_aM und aus der Übungsaufgabe 133. (Da die Mannigfaltigkeit M durch die Funktion F definiert wird, spielen die Komponenten $F_1, \ldots, F_k$ von F hier die Rolle der impliziten Funktionen $\xi_1, \ldots, \xi_k$, da die Jacobi-Matrix von F an der Stelle a bereits maximalen Rang hat). Es sei $j = 1, \ldots, k$ gegeben, wir wollen zeigen, dass $\xi_j'(a)$ auf jedem Tangentialvektor $z \in T_aM$ senkrecht steht. Denn sind $\delta > 0$ und $\gamma : (-\delta, \delta) \to M$ eine stetig Fréchet-differenzierbare Funktion, für die

$$\gamma(0) = a \qquad \text{und} \qquad \gamma'(0) = z,$$

so gilt

$$\xi_j(\gamma(s)) = 0 \qquad \text{für alle } s \in (-\delta, \delta),$$

weil jedes ξ_j auf M in einer Umgebung von a nach Konstruktion identisch verschwindet. Somit ist auch ihre Fréchet-Ableitung identisch Null, d.h., an der Stelle 0 gilt

$$0 = \xi_j'(\gamma(0)) \cdot \gamma'(0) = \xi_j'(a) \cdot z.$$

Anders gesagt: $(f'(a)|z) = 0$, was zu zeigen war. $\square$

Satz 5.5. *Es seien $A \subset \mathbb{R}^m \times \mathbb{R}^n$ eine offene Menge und $f : A \to \mathbb{R}$ sowie $F : A \to \mathbb{R}^n$ stetig differenzierbare Funktionen. Es gelte*

$$\operatorname{rang} J_F(x) = k \qquad \textit{für alle } x, \textit{ für die } F(x) = 0.$$

Dies definiert eine Untermannigfaltigkeit M_F der Dimension $m+n-k$. Hat f an einer Stelle $a \in M_F$ ein lokales (durch F) bedingtes Maximum oder Minimum, so existiert $\lambda \in \mathbb{R}^n$ mit

$$\frac{\partial}{\partial x_j}\left(f - \sum_{i=1}^{n} \lambda_i F_i\right)(a) = 0, \qquad \textit{für alle } j = 1, \dots, n+m.$$

Untersucht man also f auf Extrema, so ist man darauf hingewiesen, n sogenannte *Lagrange-Multiplikator* λ_1, λ_n zu finden. Dies schränkt sehr die Menge, in der man nach Maxima oder Minima sucht, ein. .

Beweis. Unter unseren Voraussetzungen ist M_F eine n-dimensionale Untermannigfaltigkeit. Da F stetig differenzierbar ist, kann man es an jeder Stelle von M_F durch eine implizite Funktion darstellen. Statt diese „gesamte“ implizite Funktion zu betrachten, nehmen wir einen beliebigen Tangentialvektor $T \in \mathbb{R}^n$ und finden somit insbesondere ein $\delta > 0$ und eine Funktion $\gamma : (-\delta, \delta) \to M_F \cap U_0$ so, dass

$$\gamma(0) = a \qquad \text{und} \qquad \gamma'(0) = T.$$

Da f an der Stelle a ein lokales Maximum oder Minimum hat, soll auch die stetig differenzierbare Funktion $f \circ \gamma$ an der Stelle 0 ein lokales Maximum oder Minimum haben, und ihre Ableitung soll dort somit verschwinden, d.h., aus (4.12) folgern wir

$$0 = J'_{f\circ\gamma}(0) \stackrel{!}{=} J_f(\gamma(0)) J_\gamma(0) = J_f(a)T = T^T \cdot J_f(a).$$

Da dies für jeden Tangentialvektor wiederholt werden kann, zeigt das, dass $J_f(0)$ senkrecht zu jedem Tangentialvektor ist, und somit selber ein Normalenvektor ist.

Wir wissen aus dem Lemma 5.4, dass die Vektoren $J_{F_1}(a), \dots, J_{F_n}(a)$ eine Basis des Raumes der Normalenvektoren an der Stelle a (vgl. z.B. [**4**, Satz I.9.3]) bilden, und somit endlich schließen, dass $J_f(a)$ eine lineare Kombination solcher Vektoren sein muss. □

Beispiel 135. Für $\alpha > 0$ betrachte die Funktion $f : A := \{p \in \mathbb{R}^2 : p_1, p_2 > 0\} \to \mathbb{R}$, die durch

$$f(p_1, p_2) := -p_1 \log_\alpha p_1 - p_2 \log_\alpha p_2$$

definiert wird. Betrachte die Nebenbedingung $F(p_1, p_2) = 0$, wobei

$$F(p_1, p_2) := p_1 + p_2 - 1,$$

was in der Sprache der Wahrscheinlichkeitstheorie heißt, dass (p_1, p_2) eine *diskrete Wahrscheinlichkeitsverteilung* ist. Falls $\alpha = 2$ heißt dann $f(p_1, p_2)$ in der theoretischen Informatik *Informationsentropie von p_1, p_2*. Wir wollen die kritischen Punkte (und wenn möglich die Extrema) unter der gegebenen Nebenbedingung bestimmen.

Sei $(p_1, p_2) \in \mathbb{R}^n$ ein kritischer Punkt von f, so existiert $\lambda \in \mathbb{R}$, sodass die Identität $J_f(p_1, p_2) = \lambda J_F(p_1, p_2)$ gilt. Dies liefert das System

$$\begin{cases} p_1 + p_2 &= 1, \\ -\log_2 p_1 &= \lambda + \frac{1}{\log \alpha}, \\ -\log_2 p_2 &= \lambda + \frac{1}{\log \alpha}. \end{cases}$$

Es folgt, dass $p_1 = p_2$ und somit auch, aufgrund der Nebenbedingung $p_1 + p_2 = 1$, dass $p_1 = p_2 = \frac{1}{2}$. Der einzige krirische Punkt von f unter der Nebenbedingung ist also durch $p = (\frac{1}{2}, \frac{1}{2})$ gegeben. Dort ist $f(p_1, p_2) > 0$. Man schließt, dass f in $(\frac{1}{2}, \frac{1}{2})$ ein lokales Maximum annimmt. Also ist die Gleichverteilung die Wahrscheinlichkeitsverteilung mit grössten Informationsentropie. $\square$

KAPITEL 6

Metrische Räume und Kompaktheit

In Analysis 1 haben wir eine Teilmenge von $\mathbb{R}^n$ *kompakt* genannt, falls sie beschränkt und abgeschlossen ist. Diese Definition einer kompakten Menge ist durchaus legitim im Kontext von endlichdmensionalen Räume – sie wird aber ungeeignet, falls man mit unendlichdimensionalen Räumen arbeiten muss, wie uns immer wieder in der Analysis 2 passiert.

Die „korrektere“ Version beruht stattdessen auf die Idee, eine kompakte Menge sollte „klein genug“ sein, dass man sie mit *wenigen* offenen Mengen überdecken kann, wenn sie sich überhaupt mit offenen Mengen überdecken lässt. Genauer führen wir die Folgende ein.

Definition 136. *Es sei X ein normierter Raum. Man sagt, eine Teilmenge $K \subset X$ hat die* Heine–Borel-Eigenschaft*, falls Folgendes gilt: Ist I eine Menge und $O_i \subset X$ eine offene Menge für alle $i \in I$, und gilt zudem*

$$K \subset \bigcup_{i \in I} O_i, \tag{6.1}$$

so gibt es $I_0 \subset I$, I_0 endlich, so dass

$$K \subset \bigcup_{i \in I_0} O_i.$$

Eine Familie $(O_i)_{i \in I}$ offener Mengen, die (6.1) erfüllt, heißt eine *offene Überdeckung* von K. Somit kann man die obige Definition folgendermaßen äquivalent formulieren: *Für jede offene Überdeckung von K gibt es eine endliche Teilüberdeckung.*

Beispiel 137. Jede endliche Menge hat offensichtlich die Heine–Borel-Eigenschaft. □

Eine verwandte Eigenschaft ist die Folgende.

Definition 138. *Es sei X ein normierter Raum. Dann heißt $K \subset X$* totalbeschränkt*, falls es $\delta > 0$ und $x_1, \dots, x_n \in K$ gibt so, dass*

$$X \subset \bigcup_{j=1}^{n} B_\delta(x_j).$$

.

Lemma 6.1. *Es sei X ein normierter Raum und $K \subset X$. Hat K die Heine–Borel-Eigenschaft, so ist es auch totalbeschränkt.*

Beweis. Hat K die Heine–Borel-Eigenschaft, so gilt insbesondere für alle $\delta > 0$

$$K \subset \bigcup_{x \in K} B_\delta(x).$$

Somit gibt es eine *endliche* Menge von Vektoren $\{x_1, \ldots, x_n\}$ so, dass

$$X \subset \bigcup_{j=1}^{n} B_\delta(x_j),$$

was zu zeigen war. $\square$

Lemma 6.2. *Es sei X ein normierter Raum. Hat $K \subset X$ die Heine–Borel-Eigenschaft, so ist K abgeschlossen und beschränkt.*

Beweis. Hat K die Heine–Borel-Eigenschaft, so gilt insbesondere für alle $x \in K$

$$K \subset \bigcup_{n \in \mathbb{N}} B_n(x).$$

Somit gibt es eine *endliche* Menge $\{1, \ldots, N\}$ so, dass

$$X \subset \bigcup_{n=1}^{N} B_n(x) = B_N(x),$$

was die Beschränktheit von K zeigt.

Das Komplement von X ist die leere Menge, welche definitionsgemäß offen ist, und somit ist X selber abgeschlossen.

Ist nun $K \neq X$, betrachte $x \in K^c \neq \emptyset$. Können wir eine offene Umgebung A von x finden mit $A \subset K^c$, so heißt dies, dass K^c offen ist und somit $K = (K^c)^c$ abgeschlossen ist, was zu zeigen ist.

Für jedes $y \in K$ gilt $\|y - x\| > 0$, und somit gibt es disjunkte offene Umgebungen $A_y := U_{\frac{\|y-x\|}{2}}(x)$ und $B_y := U_{\frac{\|y-x\|}{2}}(y)$ von x bzw. y. Dadurch wird eine offene Überdeckung

$$(B_y)_{y \in K}$$

definiert, die wegen der Heine–Borel-Eigenschaft auch eine endliche Teilüberdeckung haben muss, d.h., es existierern $y_1, \ldots, y_n \in K$ mit

$$K \subset \bigcup_{i=1}^{n} B_{y_i} =: B.$$

Auch betrachtet man

$$A := \bigcap_{i=1}^{n} A_{y_i}.$$

Man überprüft leicht, dass A, B disjunkt sind (Übungsaufgabe!), und daher

$$K \cap A = \emptyset,$$

d.h.,

$$A \subset K^c.$$

Gleichzeitig ist A als Durchschnitt endlich vieler offener Umgebungen von x selber eine offene Umgebung von x. $\square$

Die folgende Definition wird eine Rolle spielen. Sie erinnert stark an die Aussage im Satz von Bolzano–Weierstraß.

Definition 139. *Es sei X ein normierter Raum. Man sagt, eine Teilmenge $K \subset X$ ist* folgenkompakt*, falls jede Folge $(x_n)_{n\in\mathbb{N}} \subset K$ eine konvergente Teilfolge hat.*

Wir wissen schon aus dem Lemma 6.1, dass jede Teilmenge (eines normierten Raumes) mit der Heine–Borel-Eigenschaft totalbeschränkt ist. Auch gilt folgendes.

Lemma 6.3. *Es sei X ein normierter Raum. Ist $K \subset X$ folgenkompakt, so ist es auch totalbeschränkt.*

Beweis. Wäre K nicht totalbeschränkt, so gäbe es $\delta > 0$, für das

$$K \not\subset \bigcup_{x\in K_0} B_\delta(x)$$

für alle endliche Teilmengen $K_0 \subset K$. Insbesondere gäbe es $x_1 \in K$ so, dass $K \not\subset B_\delta(x_1)$; $x_2 \in K\setminus \subset B_\delta(x_1)$ so, dass $K \not\subset B_\delta(x_1) \cup B_\delta(x_2)$, und iterativ könnte man eine Folge $(x_n)_{n\in\mathbb{N}} \subset K$ definieren, mit

$$x_n \in K \setminus \bigcup_{i=1}^{n-1} B_\delta(x_i) \qquad \text{so, dass} \qquad K \not\subset \bigcup_{i=1}^{n} B_\delta(x_i). \tag{6.2}$$

Wegen der Folgenkompaktheit von K hätte diese Folge eine konvergente Teilfolge und insbesondere einen Häufungspunkt, etwa x. Somit könnte man zwei Folgenglieder x_n, x_m (o.B.d.A. mit $m > n$) finden derart, dass

$$\|x - x_n\| < \frac{\delta}{2} \qquad \text{und} \qquad \|x - x_m\| < \frac{\delta}{2}.$$

Daher folgt

$$\|x_n - x_m\|_X \le \|x_n - x\|_X + \|x - x_m\|_X \le \delta,$$

d.h., $x_m \in B_\delta(x_n)$ – ein Widerspruch zu (6.2). □

Satz 6.4. *Es sei X ein normierter Raum und $K \subset X$. Die folgenden Aussagen sind äquivalent.*

(i) K hat die Heine–Borel-Eigenschaft.
(ii) K ist folgenkompakt.

Gelten die obigen äquivalenten Eigenschaften, so heißt K *kompakt*.

Beweis. (i)⇒(ii) Angesichts von [**6**, Lemma 5.6] reicht es zu zeigen, dass jede Folge mit Werten in K einen Häufungspunkt hat. Gäbe es tatsächlich eine Folge ohne Häufungspunkte (welche dann notwendigerweise unendlich viele verschiedene Werte annehmen muss!), so gäbe es für jedes $x \in K$ ein δ_x klein genug, dass $B_{\delta_x}(x)$ nur endlich viele Folgenglieder enthält. Dadurch hätten wir eine offene Überdeckung von K gefunden, welche eine endliche Teilüberdeckung haben muss, etwa

$$K \subset \bigcup_{i=1}^{n} B_{\delta_{x_i}}(x_i).$$

Da jede Menge in der Teilüberdeckung selber nur endlich viele Folgenglieder enthält, enthält auch K nur endlich viele Folgenglieder – ein Widersrpuch, da die Folge unendlich viele Werte annimmt, welche alle in K liegen.

(ii)$\Rightarrow$(i) Lass uns jetzt annehmen, dass K folgenkompakt ist. Aus dem Lemma 6.3 wissen wir, dass K auch totalbeschränkt ist, d.h., zu jedem $n \in \mathbb{N}$ gibt es $x_1, \ldots, x_n \in K$ mit

$$K \subset \bigcup_{k=1}^{n} B_{\frac{1}{n}}(x_k).$$

Wir müssen zeigen, dass jede offene Überdeckung von K eine endliche Teilüberdeckung hat. Nimm an, es wäre nicht so, d.h., es gäbe eine Familie $(O_i)_{i\in I}$ mit $O_i \subset X$ offen für alle $i \in I$ und $K \subset \bigcup_{i\in I} O_i$ so dass $K \not\subset \bigcup_{i\in I_0} O_i$, egal wie geschickt wir eine endliche Teilmenge $I_0 \subset I$ auswählen. Insbesondere müsste es mindestens ein $n_k \in \mathbb{N}$ geben, für das *keine* endliche Teilmenge $\tilde{I} \subset I$ derart ausgewählt werden kann, dass

$$B_{\frac{1}{n_k}}(x_{n_k}) \cap K \subset \bigcup_{i\in\tilde{I}} O_i;$$

(denn sonst ließe sich K, das ohnehin von endlich vielen offenen Kugeln $B_{\frac{1}{n}}(x_k)$ überdeckt wird, insgesamt auch von endlich vielen offenen Mengen O_i überdecken). Dadurch könnte man eine Teilfolge $(x_{n_k})_{k\in\mathbb{N}} \subset K$ definieren, welche selber einen Häufungspunkt in K haben muss, etwa $\bar{x}$. Da $\bar{x} \in K \subset \bigcup_{i\in I} O_i$, gibt es mindestens ein $i_0 \in I$ mit $\bar{x} \in O_{i_0}$, und weil O_{i_0} offen ist, findet man $\varepsilon > 0$ klein genug, dass

$$B_\varepsilon(\bar{x}) \subset O_{i_0}. \tag{6.3}$$

Zu diesem ε findet man ein Folgenglied mit Index groß genug, etwa $x_{n_{\bar{k}}}$ mit $n_{\bar{k}} > \frac{2}{\varepsilon}$, mit

$$\|\bar{x} - x_{n_{\bar{k}}}\| \le \frac{\varepsilon}{2}.$$

Somit gilt für jedes $x \in B_{\frac{1}{n_{\bar{k}}}}(x_{n_{\bar{k}}})$

$$\|x - \bar{x}\|_X \le \|x - x_{n_{\bar{k}}}\|_X + \|x_{n_{\bar{k}}} - \bar{x}\|_X \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{6.4}$$

Dies zeigt, dass

$$B_{\frac{1}{n_{\bar{k}}}}(x_{n_{\bar{k}}}) \cap K \subset B_{\frac{1}{n_{\bar{k}}}}(x_{n_{\bar{k}}}) \overset{(6.4)}{\subset} B_\varepsilon(\bar{x}) \overset{(6.3)}{\subset} O_{i_0},$$

ein Widerspruch zur Konstruktion von $n_{\bar{k}}$. □

Dass die obige Definition von Kompaktheit mit der Definition von Analysis 1 vereinbar ist, ist eine unmittelbare Folgerung.

Korollar 6.5. *Es sei $K \subset \mathbb{R}^n$. Genau dann ist K kompakt, wenn K abgeschlossen und beschränkt ist.*

Beweis. Ist K kompakt, so wissen wir aus dem Lemma 6.2, dass K beschränkt und abgeschlossen sein muss. Es sei wiederum $K \subset \mathbb{R}^n$ abgeshlossen und beschränkt, so liefert der Satz von Bolzano–Weierstraß, dass K auch folgenkompakt sein muss. □

Eine der wichtigsten Anwendungen des obigen Resultats ist der Satz von Ascoli–Arzelà. Untersucht man nämlich den Beweis des Satzes 1.8, so sieht man gleich, dass man nicht wirklich benötigt, dass die Funktionen auf einem 1-dimensionalen Intervall $[a, b]$ definiert sind. Vielmehr könnte man die folgende allgemeine Version des Satzes formulieren.

Satz 6.6. *Es sei $K \subset \mathbb{R}^d$ kompakt. Ist eine Funktionenfolge $(f_n)_{n\in\mathbb{N}} \subset C(K;\mathbb{R}^m)$ gleichmäßig beschränkt und gleichgradig stetig, so hat sie eine gleichmäßig konvergente Teilfolge.*

Beweis. Wir übernehmen den Beweis vom Satz 1.8. Nur die Überdeckbarkeit von K durch δ-Umgebungen, die in Elementen von $K \cap \mathbb{Q}^d$ zentriert sind (es waren die Intervalle $(q_j - \delta, q_j + \delta)$ im Beweis vom Satz 1.8), soll in dieser Allgemeinheit noch bewiesen werden. Das ist aber die Aussage vom Satz 6.4. $\square$

Wir wollen eine weitere wichtige Anwendung des Begriffes der Kompaktheit zeigen.

Definition 140. *Es seien X, Y normierte Räume und $A \subset Y$. Eine Funktion $f : A \to X$ heißt* gleichmäßig stetig, *falls für alle $\varepsilon > 0$ ein $\delta > 0$ existiert, so dass*

$$\|f(x) - f(y)\|_X \leq \varepsilon \qquad \text{für alle } x, y \in A \text{ mit } \|x - y\|_Y \leq \delta.$$

Korollar 6.7. *Es seien X, Y normierte Räume und $A \subset X$ eine kompakte Menge. Dann ist jede stetige Funktion $f : A \to X$ auch gleichmäßig stetig.*

Übungsaufgabe 141. Beweise das Korollar 6.7.

Literaturverzeichnis

[1] Herbert Amann, Joachim Escher, *Analysis I*, Birkhäuser, Basel, 2002.

[2] Herbert Amann, Joachim Escher, *Analysis II*, Birkhäuser, Basel, 2006.

[3] Otto Forster, *Analysis 1*, Vieweg+Teubner, Wiesbaden 2011.

[4] Otto Forster, *Analysis 2*, Vieweg+Teubner, Wiesbaden 2008.

[5] Konrad Königsberger, *Analysis 1*, Springer-Verlag, Berlin 2004.

[6] Delio Mugnolo, *Analysis 1*, Skript zur Vorlesung im Sommersemeser 2012 an der Universität Ulm.