Ralf Gerkmann

Mathematisches Institut
Ludwig-Maximilians-Universität München

Vorlesung im Wintersemester 2021-22

Analysis mehrerer Variablen

(Studiengang Mathematik für das gymnasialen Lehramt)

Version vom 9. Februar 2022

Inhaltsverzeichnis

§ 1. Normen und Metriken 2
§ 2. Konvergenz in metrischen Räumen 9
§ 3. Stetigkeit 16
§ 4. Offene und abgeschlossene Teilmengen 27
§ 5. Kompaktheit 36
§ 6. Zusammenhang 42
§ 7. Partielle Ableitungen und Richtungsableitungen 46
§ 8. Totale Differenzierbarkeit 55
§ 9. Höhere Ableitungen und lokale Extremstellen 67
§ 10. Lokale Umkehrbarkeit und implizit definierte Funktionen 79
§ 11. Untermannigfaltigkeiten 87
§ 12. Definition des mehrdimensionalen Riemann-Integrals 90
§ 13. Der Satz von Fubini 97
§ 14. Nullmengen und Lebesguesches Integrabilitätskriterium 101

Literaturverzeichnis 107

§ 1. Normen und Metriken

Zusammenfassung. Den Begriff der *Norm* als „verallgemeinerte Längenfunktion“ auf einem $\mathbb{R}$-Vektorraum haben wir schon in § 16 der Linearen Algebra kennengelernt. In diesem Kapitel führen wir mit den *p-Normen* eine wichtige Klasse von Beispielen ein. Außerdem definieren wir den Begriff der *Äquivalenz* von Normen und zeigen, dass je zwei Normen auf dem $\mathbb{R}^n$ äquivalent sind. Die Nützlichkeit dieser Aussage wird sich später herausstellen, wenn wir damit nachweisen, dass es bei der Definition der Stetigkeit und Differenzierbarkeit mehrdimensionaler Funktionen auf die Wahl der Norm nicht ankommt. Man kann sich dann immer die Norm aussuchen, mit der am einfachsten gerechnet werden kann.

Im Gegensatz zur Norm können *Metriken* auf beliebigen Mengen definiert werden. Hierbei handelt es sich um eine „verallgemeinerte Abstandsfunktion“, also eine Abbildung, die je zwei Punkten der Menge einen Abstand zuordnet. Auch dieser Begriff ist im Hinblick auf mehrdimensionale Funktionen von Interesse. Denn der Definitionsbereich einer solchen Funktion kann eine beliebige Teilmenge vom $\mathbb{R}^n$ sein, auf der im Allgemeinen keine Vektorraumstruktur gegeben ist.

Wichtige Begriffe und Sätze in diesem Kapitel

- Definition der p-Norm auf $\mathbb{R}^n$ für $p \in [1, +\infty]$.
  (Der Nachweis der Dreiecksungleichung erfordert die sog. Höldersche Ungleichung.)
- Definition der Äquivalenz von Normen
- Je zwei Normen auf $\mathbb{R}^n$ sind äquivalent.
- Definition von Metriken auf beliebigen Mengen
- Jede Norm auf einem $\mathbb{R}$-Vektorraum V induziert eine Metrik auf V.

(1.1) Satz. Auf dem $\mathbb{R}$-Vektorraum $V = \mathbb{R}^n$ ist für jedes $p \in \mathbb{R}$, $p \geq 1$ durch

$$\|x\|_p \quad = \quad \sqrt[p]{\sum_{i=1}^n |x_i|^p} \quad , \quad x = (x_1, ..., x_n) \in V$$

eine Norm definiert, die sogenannte ***p-Norm***. Eine weitere Norm erhält man durch

$$\|x\|_\infty \quad = \quad \max\{|x_1|, ..., |x_n|\} \quad , \quad \text{die } \textbf{\textit{Supremumsnorm}}.$$

Beweis: Wir überprüfen zunächst die Normeigenschaften der Supremumsnorm. Offenbar gilt $\|x\|_\infty = 0$ genau dann, wenn $|x_1| = ... = |x_n| = 0$ gilt, und dies wiederum genau dann der Fall, wenn $x = 0_{\mathbb{R}^n}$ ist. Sei nun $x \in V$ und $\lambda \in \mathbb{R}$. Für den Beweis der Gleichung $\|\lambda x\|_\infty = |\lambda| \|x\|_\infty$ sei $i \in \{1, ..., n\}$ so gewählt, dass $|x_i| \geq |x_j|$ für $1 \leq j \leq n$ erfüllt ist. Dann gilt auch $|\lambda x_i| \geq |\lambda x_j|$ für alle j, und es folgt $\|\lambda x\|_\infty = |\lambda x_i| = |\lambda| |x_i| = |\lambda| \|x\|_\infty$. Zum Beweis der Dreiecksungleichung seien $x, y \in \mathbb{R}^n$ vorgegeben. Für $1 \leq i \leq n$ gilt jeweils

$$|x_i + y_i| \quad \leq \quad |x_i| + |y_i| \quad \leq \quad \|x\|_\infty + \|y\|_\infty \quad ,$$

also auch $\|x + y\|_\infty \leq \|x\|_\infty + \|y\|_\infty$. Damit ist $\|\cdot\|_\infty$ tatsächlich eine Norm auf V.

Wenden wir uns nun dem Beweis der Norm-Eigenschaften für die p-Norm zu, wobei $p \in \mathbb{R}$ und $p \geq 1$ ist. Für jedes $x \in V$ gilt auch hier $\|x\|_p = 0$ genau dann, wenn die Beträge $|x_i|$ der Koordinaten alle gleich Null sind, und dies ist

wiederum äquivalent zu $x = 0_{\mathbb{R}^n}$. Für alle $x \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$ gilt

$$\|\lambda x\|_p \quad = \quad \sqrt[p]{\textstyle\sum_{i=1}^n |\lambda x_i|^p} \quad = \quad |\lambda| \sqrt[p]{\textstyle\sum_{i=1}^n |x_i|^p} \quad = \quad |\lambda| \|x\|_p. \tag{1}$$

Also ist auch die zweite Bedingung erfüllt. Der Beweis der Dreiecksungleichung ist leider nur für $p = 1$ einfach. Hier folgt sie durch die Rechung

$$\|x+y\|_1 \quad = \quad \sum_{i=1}^n |x_i + y_i| \quad \leq \quad \sum_{i=1}^n |x_i| + \sum_{i=1}^n |y_i| \qquad = \qquad \|x\|_1 + \|y\|_1. \qquad \square$$

Für den Beweis der Dreiecksungleichung im Fall $p > 1$ benötigen wir als zusätzliches Hilfsmittel die Höldersche Ungleichung. Der Beweis dieser Ungleichung erfordert allerdings ein wenig Vorbereitung.

(1.2) Lemma. Seien $p, q \in \mathbb{R}$, $p, q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$ und $x, y \in \mathbb{R}_+$. Dann gilt

$$x^{1/p} y^{1/q} \quad \leq \quad \frac{x}{p} + \frac{y}{q}.$$

Beweis: Wir können voraussetzen, dass $x, y \neq 0$ sind, denn im Fall $x = 0$ oder $y = 0$ ist die Ungleichung offensichtlich erfüllt. Aus Symmetriegründen können wir außerdem $x \geq y$ annehmen, da wir ansonsten die Aussage durch Vertauschung von x und y bzw. p und q auf diesen Fall zurückführen können. Schließlich können wir auch noch $x > y$ voraussetzen, denn im Fall $x = y$ reduziert sich die Aussage auf die offensichtliche Ungleichung $x \leq x$.

Dividieren wir die Ungleichung durch y und setzen wir $\xi = \frac{x}{y}$, dann erhalten wir auf der rechten Seite den Term

$$\tfrac{1}{p}\xi + \tfrac{1}{q} \quad = \quad \tfrac{1}{p}\xi + 1 - \tfrac{1}{p} \quad = \quad \tfrac{1}{p}(\xi - 1) + 1 \quad ,$$

auf der linken Seite $x^{1/p} y^{1/q-1} = (\frac{x}{y})^{1/p} y^{1/p+1/q-1} = \xi^{1/p}$. Es genügt also, für jedes reelle $\xi > 1$ die Ungleichung $\xi^{1/p} \leq \frac{1}{p}(\xi - 1) + 1$ zu beweisen. Setzen wir $\eta = \xi - 1$, so erhalten wir nach Subtraktion von 1 auf beiden Seiten die äquivalente Ungleichung

$$(\eta + 1)^{1/p} - 1 \quad \leq \quad \tfrac{1}{p}\eta \qquad \text{für} \quad \eta > 0$$

Diese kann nun durch den Mittelwertsatz der Differentialrechnung bewiesen werden. Dazu betrachten wir die Funktion $\phi(t) = (t+1)^{1/p}$ auf dem abgeschlossenen Intervall $[0, \eta]$. Auf Grund des Mittelwertsatzes gibt es ein $t_0 \in]0, \eta[$ mit $\phi(\eta) - \phi(0) = \eta \phi'(t_0)$. Die linke Seite dieser Gleichung ist gleich $(\eta+1)^{1/p} - 1$, die rechte Seite ist wegen $\phi'(t) = \frac{1}{p}(t+1)^{1/p-1}$ gleich $\eta \cdot \frac{1}{p}(t_0+1)^{1/p-1}$. Wegen $t_0 + 1 > 1$ und $\frac{1}{p} - 1 < 0$ kann die rechte Seite durch $\frac{1}{p}(t_0+1)^{1/p-1} \leq \frac{1}{p}\eta$ abgeschätzt werden. $\square$

Die Cauchy-Schwarzsche Ungleichung aus § 15 der Linearen Algebra lässt sich nun verallgemeinern zu

(1.3) Proposition. *(Höldersche Ungleichung)*
Seien $x, y \in \mathbb{R}^n$, $x = (x_1, ..., x_n)$ und $y = (y_1, ..., y_n)$, und seien $p, q \in \mathbb{R}$ mit $p, q > 1$ und $\frac{1}{p} + \frac{1}{q} = 1$ vorgegeben. Dann gilt

$$\sum_{i=1}^n |x_i| \cdot |y_i| \quad \leq \quad \|x\|_p \|y\|_q$$

Beweis: Nach (1.2), angewendet auf die nicht-negativen Zahlen $\frac{|x_i|^p}{\|x_i\|_p^p}$ und $\frac{|y_i|^q}{\|y_i\|_q^q}$ gilt

$$\frac{|x_i|}{\|x_i\|_p}\frac{|y_i|}{\|y_i\|_q} \quad \leq \quad \frac{1}{p}\frac{|x_i|^p}{\|x_i\|_p^p}+\frac{1}{q}\frac{|y_i|^q}{\|y_i\|_q^q} \qquad \text{für} \quad 1\leq i\leq n.$$

Daraus folgt

$$\sum_{i=1}^n \frac{|x_i|}{\|x\|_p}\cdot\frac{|y_i|}{\|y\|_q} \quad \leq \quad \sum_{i=1}^n\left(\frac{1}{p}\frac{|x_i|^p}{\|x\|_p^p}+\frac{1}{q}\frac{|y_i|^q}{\|y\|_q^q}\right) \quad \leq$$

$$\frac{1}{p}\frac{1}{\|x\|_p^p}\sum_{i=1}^n|x_i|^p+\frac{1}{q}\frac{1}{\|y\|_q^q}\sum_{i=1}^n|y_i|^p \quad = \quad \frac{1}{p}\frac{1}{\|x\|_p^p}\|x\|_p^p+\frac{1}{q}\frac{1}{\|y\|_q^q}\|y\|_q^q \quad = \quad \frac{1}{p}+\frac{1}{q} \quad = \quad 1.$$

Durch Multiplikation dieser Ungleichung mit $\|x\|_p\|y\|_q$ erhalten wir die Höldersche Ungleichung. □

Beweis der Dreiecksungleichung für die p-Norm, für $p>1$:

Seien $x,y\in\mathbb{R}^n$ vorgegeben. Wir können $x+y\neq 0_{\mathbb{R}^n}$ annehmen, weil die Dreiecksungleichung ansonsten offensichtlich sogar mit Gleichheit erfüllt ist. Sei $q\in\mathbb{R}$ die eindeutig bestimmte (positive) reelle Zahl mit $\frac{1}{p}+\frac{1}{q}=1$, nämlich $q=\frac{p}{p-1}$. Außerdem sei $z\in\mathbb{R}^n$ der Vektor mit den Komponenten $z_i=(x_i+y_i)^{p-1}$ für $1\leq i\leq n$. Es gilt dann

$$\|x+y\|_p^p \quad = \quad \sum_{i=1}^n|x_i+y_i|^p \quad \leq \quad \sum_{i=1}^n|x_i||x_i+y_i|^{p-1}+\sum_{i=1}^n|y_i||x_i+y_i|^{p-1} \quad = \quad \sum_{i=1}^n|x_i||z_i|+\sum_{i=1}^n|y_i||z_i|$$

$$\leq \quad \|x\|_p\|z\|_q+\|y\|_p\|z\|_q \quad = \quad \|x\|_p\left(\sum_{i=1}^n|x_i+y_i|^{(p-1)q}\right)^{1/q}+\ \|y\|_p\left(\sum_{i=1}^n|x_i+y_i|^{(p-1)q}\right)^{1/q}$$

$$= \quad \|x\|_p\left(\sum_{i=1}^n|x_i+y_i|^p\right)^{(p-1)/p}+\ \|y\|_p\left(\sum_{i=1}^n|x_i+y_i|^p\right)^{(p-1)/p} \quad = \quad \|x\|_p\cdot\|x+y\|_p^{p-1}+\ \|y\|_p\cdot\|x+y\|_p^{p-1}$$

wobei im vierten Schritt die Höldersche Ungleichung und im sechsten Schritt die Gleichungen $q=\frac{p}{p-1}$ und $\frac{1}{q}=\frac{p-1}{p}$ verwendet wurden. Division durch $\|x+y\|_p^{p-1}$ auf beiden Seiten liefert das gewünschte Ergebnis.

Die Verwendung vom Index ∞ bei der Supremumsnorm ist auf Grund der folgenden Beziehung zwischen p-Norm und Supremumsnorm gerechtfertigt.

(1.4) Proposition. Für alle $x\in\mathbb{R}^n$ gilt $\lim\limits_{p\to\infty}\|x\|_p=\|x\|_\infty$.

Beweis: Für $x=0$ ist die Gleichung offensichtlich erfüllt, denn dann gilt $\|x\|_\infty=0$ und $\|x\|_p=0$ für alle $p\in\mathbb{N}$. Sei nun $x\neq 0$ und x_k die Komponente von x mit dem größten Betrag. Dann können wir $\|x\|_p$ auch in der Form

$$|x_k|\left(\sum_{i=1}^n\frac{|x_i|^p}{|x_k|^p}\right)^{1/p}$$

schreiben. Alle Summanden mit $|x_i|<|x_k|$ laufen für $p\to\infty$ gegen Null, wärend die Summanden mit $|x_i|=|x_k|$ konstant gleich 1 sind. Deshalb läuft der Term in der Klammer für $p\to\infty$ gegen die Anzahl ℓ der Summanden mit $|x_i|=|x_k|$. Insbesondere gilt $\ell\leq\sum_{i=1}^n\frac{|x_i|^p}{|x_k|^p}\leq 2\ell$ für hinreichend großes p. Aus $\lim\limits_{p\to\infty}\sqrt[p]{\ell}=\lim\limits_{p\to\infty}\sqrt[p]{2\ell}=1$ folgt nun

mit dem Sandwich-Lemma aus der Analysis einer Variablen

$$\lim_{p\to\infty} \|x\|_p \quad = \quad |x_k| \cdot \lim_{p\to\infty} \left(\sum_{i=1}^{n} \frac{|x_i|^p}{|x_k|^p} \right)^{1/p} \quad = \quad |x_k| \cdot 1 \quad = \quad \|x\|_\infty . \qquad \square$$

Häufig lassen sich Normen durch eine Konstante gegeneinander abschätzen. So gilt für alle $x \in \mathbb{R}^n$ zum Beispiel

$$\|x\|_2 \quad = \quad \left(\sum_{k=1}^{n} |x_k|^2 \right)^{1/2} \quad \leq \quad (n|x_i|^2)^{1/2} \quad = \quad \sqrt{n}\|x\|_\infty \quad ,$$

wobei x_i wieder eine Komponente von x mit maximalem Betrag $|x_i|$ bezeichnet. Eine ebenso einfache Rechnung zeigt $\|x\|_\infty \leq \|x\|_2$. Zwischen der 1- und der Supremumsnorm hat man die Abschätzungen $\|x\|_1 \leq n\|x\|_\infty$ und $\|x\|_\infty \leq \|x\|_1$.

(1.5) Definition. Zwei Normen $\|\cdot\|$ und $\|\cdot\|'$ auf einem $\mathbb{R}$-Vektorraum V werden als ***äquivalent*** bezeichnet, wenn reelle Konstanten $\gamma_1, \gamma_2 > 0$ mit der Eigenschaft

$$\gamma_1\|x\| \quad \leq \quad \|x\|' \quad \leq \quad \gamma_2\|x\| \qquad \text{für alle } x \in V \text{ existieren.}$$

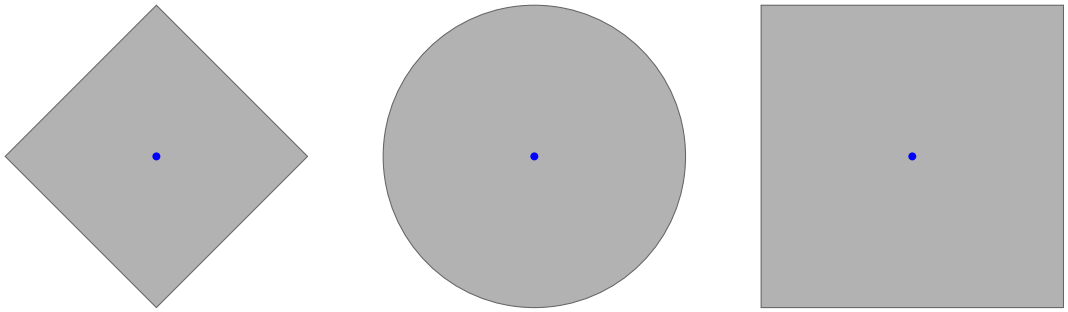
Offenbar gleichwertig mit dieser Bedingung ist die Existenz von reellen Konstanten $\delta_1, \delta_2 > 0$ mit $\|x\| \leq \delta_1\|x\|'$ und $\|x\|' \leq \delta_2\|x\|$ für alle $x \in V$.

Es ist nicht schwer zu sehen, dass jede Norm auf einem $\mathbb{R}$-Vektorraum V zu sich selbst äquivalent ist. Ist eine Norm $\|\cdot\|$ auf V äquivalent zu $\|\cdot\|'$, dann ist auch $\|\cdot\|'$ äquivalent zu $\|\cdot\|$. Sind $\|\cdot\|$, $\|\cdot\|'$, $\|\cdot\|''$ drei Normen auf V, wobei $\|\cdot\|$ äquivalent zu $\|\cdot\|'$ und $\|\cdot\|'$ äquivalent zu $\|\cdot\|''$, dann ist auch $\|\cdot\|$ äquivalent zu $\|\cdot\|''$. Die Ausarbeitung der Details ist eine leichte Übungsaufgabe. Man fast die drei Aussagen zusammen in der Feststellung, dass durch den Begriff der Äquivalenz auf der Menge der Normen von V eine ***Äquivalenzrelation*** gegeben ist.

Dem Begriff der Äquivalenz lässt sich folgendermaßen eine anschauliche Bedeutung geben. Für alle $r \in \mathbb{R}^+$ und $a \in V$ bezeichnen wir mit

$$B_{\|\cdot\|,r}(a) \quad = \quad \{x \in V \mid \|x-a\| < r\} \qquad \text{bzw.} \qquad \bar{B}_{\|\cdot\|,r}(a) \quad = \quad \{x \in V \mid \|x-a\| \leq r\}$$

den ***offenen*** bzw. ***abgeschlossenen Ball*** vom Radius r um den Punkt a bezüglich der Norm $\|\cdot\|$. Ebenso definieren wir $B_{\|\cdot\|',r}$ und $\bar{B}_{\|\cdot\|',r}$ für die Norm $\|\cdot\|'$. Die folgende Graphik zeigt die abgeschlossenen Bälle einiger Normen auf dem $\mathbb{R}^2$, wobei der blaue Punkt jeweils den Koordinatenursprung kennzeichnet.



$\{x \in \mathbb{R}^2 \mid \|x\|_1 \leq 1\}$ $\{x \in \mathbb{R}^2 \mid \|x\|_2 \leq 1\}$ $\{x \in \mathbb{R}^2 \mid \|x\|_\infty \leq 1\}$

(1.6) Proposition. Sei $\delta \in \mathbb{R}^+$. Dann ist die Ungleichung $\|x\|' \leq \delta\|x\|$ für alle $x \in V$ gleichbedeutend mit $B_{\|\cdot\|,r}(a) \subseteq B_{\|\cdot\|',\delta r}(a)$ für $a \in V$ und $r \in \mathbb{R}^+$. Eine entsprechende Aussage gilt auch für die abgeschlossenen Bälle.

Beweis: Wir beschränken uns darauf, die Äquivalenzaussage für die offenen Bälle zu beweisen.

„⇒“ Setzen wir $\|x\|' \leq \delta\|x\|$ für alle $x \in V$ voraus. Ist nun $x \in B_{\|\cdot\|,r}(a)$ vorgegeben, dann gilt $\|x-a\| < r$, damit $\delta\|x-a\| < \delta r$ und $\|x-a\|' < \delta r$. Es folgt $x \in B_{\|\cdot\|',\delta r}(a)$.

„⇐“ Nehmen wir an, dass $B_{\|\cdot\|,r}(a) \subseteq B_{\|\cdot\|',\delta r}(a)$ für alle $r \in \mathbb{R}^+$ und $a \in V$ erfüllt ist, zugleich aber ein $x \in V$ mit $\|x\|' > \delta\|x\|$ existiert. Setzen wir $r = \|x\|'$, dann gilt $\|\delta x\| = \delta\|x\| < r$ und somit $\delta x \in B_{\|\cdot\|,r}(0_V)$. Andererseits ist $\|x\|' = r$, also $\|\delta x\|' = \delta r$ und damit $\delta x \notin \mathrm{B}_{\|\cdot\|',\delta r}(0_V)$. Dies steht zur angenommenen Inklusion im Widerspruch. □

Dem folgenden Resultat hat für die Analysis endlich-dimensionaler Vektorräume eine zentrale Bedeutung.

(1.7) Satz. Auf jedem endlich-dimensionalen $\mathbb{R}$-Vektorraum V sind je zwei Normen äquivalent.

Wir schicken dem Beweis von (1.7) ein einfaches Lemma voraus.

(1.8) Lemma. Seien V, W zwei $\mathbb{R}$-Vektorräume, $\phi : W \to V$ eine injektive lineare Abbildung und $\|\cdot\|$ eine Norm auf V. Dann ist durch $\|w\|' = \|\phi(w)\|$ eine Norm auf W definiert.

Beweis: Wir überprüfen, dass die Abbildung $\|\cdot\|'$ die Eigenschaften einer Norm besitzt, indem wir diese auf die Norm-Eigenschaften von $\|\cdot\|$ zurückführen. Zunächst gilt $\|0_W\|' = \|\phi(0_W)\| = \|0_V\| = 0$. Ist $w \in W$ ein Vektor mit $\|w\|' = 0$, dann folgt $\|\phi(w)\| = 0 \Rightarrow \phi(w) = 0_V$ und somit $w = 0_W$ auf Grund der Injektivität von ϕ. Für alle $w \in V$ und $\lambda \in \mathbb{R}$ gilt $\|\lambda w\|' = \|\phi(\lambda w)\| = \|\lambda\phi(w)\| = |\lambda|\|\phi(w)\| = |\lambda|\|w\|'$, und für vorgegebene $w_1, w_2 \in V$ ist wegen $\|w_1 + w_2\|' = \|\phi(w_1 + w_2)\| = \|\phi(w_1) + \phi(w_2)\| \leq \|\phi(w_1)\| + \|\phi(w_2)\| = \|w_1\|' + \|w_2\|'$ die Dreiecks-Ungleichung erfüllt. □

Beweis von (1.7):

Seien $\|\cdot\|$ und $\|\cdot\|'$ zwei Normen auf V. Beim Nachweis der Äquivalenz beschränken wir uns zunächst auf den Fall $V = \mathbb{R}^d$, für beliebiges $d \in \mathbb{N}$. Es genügt zu zeigen, dass $\|\cdot\|$ äquivalent zur 1-Norm $\|\cdot\|_1$ ist. Weil $\|\cdot\|$ nämliche eine beliebig gewählte Norm ist, folgt aus dem Beweis dann auch die Äquivalenz von $\|\cdot\|'$ und $\|\cdot\|_1$, und auf Grund der Bemerkungen von oben erhalten wir somit insgesamt die Äquivalenz von $\|\cdot\|$ und $\|\cdot\|'$.

Sei $\delta_1 = \max\{\|e_1\|, ..., \|e_d\|\}$, wobei $e_i \in \mathbb{R}^d$ jeweils den i-ten Einheitsvektor bezeichnet. Für einen beliebigen Vektor $v \in \mathbb{R}^n$ mit $v = (v_1, ..., v_n)$ gilt dann

$$\|v\| \quad = \quad \left\| \sum_{i=1}^{d} v_i e_i \right\| \quad \leq \quad \sum_{i=1}^{d} |v_i| \|e_i\| \quad \leq \quad \delta_1 \sum_{i=1}^{d} |v_i| \quad = \quad \delta_1 \|v\|_1.$$

Um zu zeigen, dass auch eine Konstante δ_2 mit der Eigenschaft $\|v\|_1 \leq \delta_2\|v\|$ existiert, betrachten wir die Menge $S = \{w \in \mathbb{R}^d \mid \|w\|_1 = 1\}$ und definieren $\gamma = \inf\{\|w\| \mid w \in S\}$. Angenommen, wir können zeigen, dass $\gamma > 0$ ist. Für einen Vektor $v \in \mathbb{R}^d$, $v \neq 0_{\mathbb{R}^d}$ ist $w = \|v\|_1^{-1} v$ in S enthalten, es gilt also $\|v\|_1^{-1}\|v\| = \|w\| \geq \gamma$ und somit

$\|v\|_1 \leq \gamma^{-1}\|v\|$. Im Fall $v = 0_{\mathbb{R}^d}$ ist die Ungleichung $\|v\|_1 \leq \gamma^{-1}\|v\|$ offenbar auch erfüllt. Setzen wir also $\delta_2 = \gamma^{-1}$, dann gilt $\|v\|_1 \leq \delta_2\|v\|$ für alle $v \in \mathbb{R}^d$, und die Äquivalenz der beiden Normen ist damit bewiesen.

Die Ungleichung $\gamma > 0$ erhalten wir durch den folgenden Widerspruchsbeweis. Angenommen, es ist $\gamma = 0$. Dann gibt es eine Folge $(w^{(n)})_{n\in\mathbb{N}}$ von Vektoren $w^{(n)} \in S$ mit $\lim_n \|w^{(n)}\| = 0$. Wegen $\|w^{(n)}\|_1 = 1$ gilt jeweils $|w_i^{(n)}| \leq 1$ für die Komponenten von $w^{(n)}$ (für alle $n \in \mathbb{N}$, $1 \leq i \leq d$). Insbesondere die Folge $(w_1^{(n)})_{n\in\mathbb{N}}$ ist also ganz im Intervall $[-1,1]$ enthalten. Nach dem Satz von Bolzano-Weierstraß aus der Analysis einer Variablen gibt es eine Teilfolge von $(w_1^{(n)})_{n\in\mathbb{N}}$, die gegen eine Zahl $a_1 \in \mathbb{R}$ konvergiert. Durch Übergang zu einer Teilfolge von $(w_n)_{n\in\mathbb{N}}$ können wir also erreichen, dass $\lim_n w_1^{(n)} = a_1$ erfüllt ist. Indem wir die Folge noch weiter ausdünnen, können wir sogar

$$\lim_{n\to\infty} w_i^{(n)} \quad = \quad a_i \qquad \text{für} \quad 1 \leq i \leq d \quad \text{annehmen.}$$

Setzen wir $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$, so folgt $\lim_n \|a - w^{(n)}\|_1 = \lim_n \sum_{i=1}^d |a_i - w_i^{(n)}| = 0$. Aus $w^{(n)} \in S$ folgt nun einerseits $\|w^{(n)}\| = 1$ für alle $n \in \mathbb{N}$ und somit auch

$$\|a\|_1 \quad = \quad \sum_{i=1}^d |a_i| \quad = \quad \lim_{n\to\infty} \sum_{i=1}^d |w_i^{(n)}| \quad = \quad \lim_{n\to\infty} \|w^{(n)}\|_1 \quad = \quad 1.$$

Also ist auch a in S enthalten. Betrachten wir andererseits für jedes $n \in \mathbb{N}$ die Ungleichung $\|a\| \leq \|a - w^{(n)}\| + \|w^{(n)}\| \leq \delta_1\|a - w^{(n)}\|_1 + \|w^{(n)}\|$ und lassen wir n gegen Unendlich laufen, so folgt $\|a\| = 0$. Auf Grund der Norm-Eigenschaft von $\|\cdot\|$ müsste $a = 0_{\mathbb{R}^d}$ gelten. Aber in diesem Fall kann a kein Element von S sein, denn es gilt $\|0_{\mathbb{R}^d}\|_1 = 0$. Wir haben die Annahme $\gamma = 0$ auf einen Widerspruch geführt. Damit ist der Beweis für $V = \mathbb{R}^d$ insgesamt abgeschlossen.

Sei nun V ein beliebiger d-dimensionaler $\mathbb{R}$-Vektorraum, und seien $\|\cdot\|$ und $\|\cdot\|'$ zwei Normen auf V. Aus der Linearen Algebra ist bekannt, dass je zwei $\mathbb{R}$-Vektorräume derselben Dimension isomorph sind. Es gibt also einen Isomorphismus $\phi : \mathbb{R}^d \to V$ von $\mathbb{R}$-Vektorräumen. Nach (1.8) sind durch $\|v\|_* = \|\phi(v)\|$ und $\|v\|'_* = \|\phi(v)\|'$ Normen auf $\mathbb{R}^d$ definiert. Wie wir bereits gezeigt haben, sind zwei Normen auf $\mathbb{R}^d$ äquivalent, also gibt es Konstanten $\delta_1, \delta_2 \in \mathbb{R}^+$ mit $\|v\|'_* \leq \delta_1\|v\|_*$ und $\|v\|_* \leq \delta_2\|v\|'_*$ für alle $v \in \mathbb{R}^d$. Für ein beliebig vorgegebenes $w \in V$ setzen wir $v = \phi^{-1}(w)$. Es gilt dann $\phi(v) = w$ und folglich

$$\|w\|' \quad = \quad \|\phi(v)\|' \quad = \quad \|v\|'_* \quad \leq \quad \delta_1\|v\|_* \quad = \quad \delta_1\|\phi(v)\| \quad = \quad \delta_1\|w\|.$$

Genauso beweist man die Abschätzung $\|w\| \leq \delta_2\|w\|'$. □

Wie bereits erwähnt, liefert eine Norm $\|\cdot\|$ auf einem $\mathbb{R}$-Vektorraum V einen Abstandsbegriff: Sind $x, y \in V$, dann kann $\|y - x\|$ als Abstand der Punkte x und y interpretiert werden. Für die folgende Theorie ist es aber wünschenswert, auch auf allgemeineren Mengen einen Abstandsbegriff zur Verfügung zu haben. Eine solche allgemeinere Fassung erhält man durch den Begriff der Metrik.

(1.9) Definition. Sei X eine Menge. Eine ***Metrik*** auf X ist eine Abbildung $d : X \times X \to \mathbb{R}_+$ mit den folgenden Eigenschaften.

(i) Für alle $x, y \in X$ gilt $d(x,y) = 0$ genau dann, wenn $x = y$ ist.

(ii) Es gilt $d(x,y) = d(y,x)$ für alle $x, y \in X$.

(iii) Für alle x, y, z gilt $d(x,z) \leq d(x,y) + d(y,z)$. *(Dreiecksungleichung)*

Das Paar (X,d) bezeichnet man als ***metrischen Raum***.

Genau wie bei den normierten Vektorräumen definieren wir

(1.10) Definition. Sei (X,d) ein metrischer Raum. Für jeden Punkt $x \in X$ und jede Zahl $r \in \mathbb{R}^+$ bezeichnet man $B_r(x) = \{y \in X \mid d(x,y) < r\}$ bzw. $\bar{B}_r(x) = \{y \in X \mid d(x,y) \leq r\}$ als den ***offenen*** bzw. den ***abgeschlossenen Ball*** vom Radius r um den Punkt x.

Durch die folgende Bemerkung können die normierten Vektorräume den metrischen Räumen untergeordnet werden.

(1.11) Proposition. Sei $(V, \|\cdot\|)$ ein normierter Raum und $X \subseteq V$ eine Teilmenge. Dann ist durch die Definition $d(x,y) = \|x-y\|$ für $x,y \in X$ eine Metrik auf X gegeben. Man nennt sie die von der Norm $\|\cdot\|$ ***induzierte*** Metrik.

Beweis: Wir überprüfen die Bedingungen (i) bis (iii) aus der Definition. Für alle $x \in X$ gilt $d(x,x) = \|x-x\| = \|0_V\| = 0$. Sind umgekehrt $x,y \in X$ mit $d(x,y) = \|x-y\| = 0$, dann folgt $x-y = 0_V$ aus der Normdefinition und somit $x = y$. Damit ist (i) bewiesen. Für alle $x,y \in X$ gilt

$$d(x,y) \quad = \quad \|x-y\| \quad = \quad \|(-1)(y-x)\| \quad = \quad |-1|\|y-x\| \quad = \quad \|y-x\| \quad = \quad d(y,x) \quad ,$$

also ist auch Bedingung (ii) gültig. Seien schließlich $x,y,z \in X$ vorgegeben. Aus der Dreiecksungleichung für die Norm folgt dann $d(x,z) = \|x-z\| = \|(x-y)+(y-z)\| \leq \|x-y\| + \|y-z\| = d(x,y) + d(y,z)$. □

Ist speziell $V = \mathbb{R}^d$ für ein $d \in \mathbb{N}$ und $\|\cdot\| = \|\cdot\|_p$ für ein $p \in \mathbb{R}$ mit $p \geq 1$ oder $p = \infty$, dann bezeichnen wir die induzierte Metrik mit d_p. Ein wichtiger Spezialfall ist die von jeder p-Norm auf $\mathbb{R}^1 = \mathbb{R}$ induzierte Metrik gegeben durch $d(x,y) = |x-y|$ für $x,y \in \mathbb{R}$, die wir auch als ***Standardmetrik*** auf $\mathbb{R}$ bezeichnen.

Allgemein wird nicht jede Metrik von einer Norm induziert, selbst dann nicht, wenn die unterliegende Menge X Teilmenge eines $\mathbb{R}$-Vektorraums ist. Wir betrachten dazu das folgende Beispiel.

(1.12) Definition. Auf jeder Menge X ist die ***diskrete Metrik*** δ_X folgendermaßen definiert: Für alle $x \in X$ ist $\delta_X(x,x) = 0$, und für alle $x,y \in X$ mit $x \neq y$ setzt man $\delta_X(x,y) = 1$.

Die Überprüfung der Metrik-Eigenschaften von δ_X ist dem Leser als Übungsaufgabe überlassen. Ist nun V ein mindestens eindimensionaler $\mathbb{R}$-Vektorraum, dann wird δ_V nicht durch eine Norm $\|\cdot\|$ auf V induziert. Denn nehmen wir an, dies wäre doch der Fall. Für einen beliebigen Vektor $v \neq 0_V$ gilt dann $\delta_V(v,0_V) = \|v\|$ und $\delta_V(2v,0_V) = \|2v\| = 2\|v\|$. Aus der ersten Gleichung und der Definition der diskreten Metrik folgt dann $\|v\| = \delta_V(v,0_V) = 1$. Durch erneute Anwendung der Definition erhalten wir dann aber den Widerspruch $1 = \delta_V(2v,0_V) = \|2v\| = 2\|v\| = 2$.

Für das Verständnis ist es auch hilfreich sich zu überlegen, wie die offenen bzw. abgeschlossenen Bälle bezüglich der diskreten Metrik auf einer Menge X aussehen: Für alle $x \in X$ und $r < 1$ ist $B_r(x) = \bar{B}_r(x) = \{x\}$. Für $r = 1$ gilt $B_r(x) = \{x\}$ und $\bar{B}_r(x) = X$. Im Fall $r > 1$ gilt schließlich $B_r(x) = \bar{B}_r(x) = X$.

§ 2. Konvergenz in metrischen Räumen

Zusammenfassung. In diesem Abschnitt verallgemeinern wir die Definition der *Konvergenz*, die wir aus dem ersten Semester für Folgen reeller Zahlen definiert haben, auf Folgen in beliebigen metrischen Räumen. Wie dort ist auch hier das Ziel, der ungenauen Aussage, dass eine Folge $(x^{(n)})_{n\in\mathbb{N}}$ in einem metrischen Raum (X, d_X) sich einem Punkt $a \in X$ „nähert", oder dass der Abstand zwischen $x^{(n)}$ und a „immer kleiner wird", eine präzise Bedeutung zu geben. Während der Abstand in $\mathbb{R}$ durch den Absolutbetrag $|x^{(n)} - a|$ gemessen wird, übernimmt diese Aufgabe nun mit dem Ausdruck $d_X(x^{(n)}, a)$ die Abstandsfunktion d_X des metrischen Raums.

In der Analysis einer Variablen haben wir festgestellt, dass die Konvergenz von Folgen auch mit Hilfe der *Cauchyfolgen-Eigenschaft* überprüft werden kann, was den Vorteil hat, dass man für dieses Kriterium den Grenzwert a nicht zu kennen braucht. Wir werden sehen, dass dies in vielen Fällen auch in metrischen Räumen möglich ist. Zum Schluss beweisen wir den sog. *Banachschen Fixpunktsatz*, ein wichtiges Hilfsmittel zur praktischen Bestimmung von Grenzwerten. Daneben spielt dieser Satz auch für die Theorie eine wichtige Rolle; unter anderem werden wir mit ihm sim nächsten Semester die eindeutige Lösbarkeit einer großen Klasse von Differenzialgleichungen beweisen.

Wichtige Begriffe und Sätze in diesem Kapitel

- Grenzwerte in metrischen Räumen, konvergente und divergente Folgen
- Vektorräume mit äquivalenten Normen haben die gleichen konvergenten Folgen.
- Cauchyfolgen und Vollständigkeit metrischer Räume
- Banachscher Fixpunktsatz

In der Analysis einer Variablen haben wir die Schreibweise $(x_n)_{n\in\mathbb{N}}$ für Folgen reeller Zahlen verwendet. Als mathematisches Objekt ist eine solche Folge lediglich eine Abbildung $\mathbb{N} \to \mathbb{R}$. Unter einer Folge in einem metrischen Raum (X, d) verstehen wir nun entsprechend eine Abbildung $\mathbb{N} \to X$ und verwenden für solche Folgen Bezeichnungen der Form $(x^{(n)})_{n\in\mathbb{N}}$. Den Index n des Folgengliedes setzen wir nach oben, da es sich bei unseren metrischen Räumen oft um Teilmengen des $\mathbb{R}^n$ handelt und wir den unteren Index zur Bezeichnung der Komponenten des Vektors benötigen. Die Komponenten eines Folgenglieds $x^{(n)}$ im Vektorraum $\mathbb{R}^m$ bezeichnen wir üblicherweise mit $x_k^{(n)}$ für $1 \le k \le m$. In dieser Schreibeweise gilt dann $x^{(n)} = (x_1^{(n)}, \ldots, x_m^{(n)})$.

(2.1) Definition. Sei (X, d) ein metrischer Raum, $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge in X und $a \in X$ ein Punkt. Man sagt, die Folge ***konvergiert*** in (X, d) gegen a und schreibt

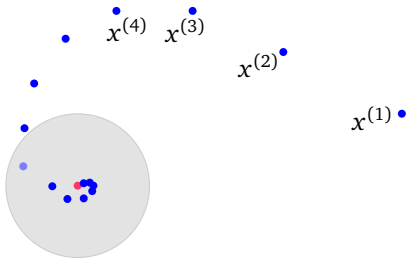
$$\lim_{n\to\infty} x^{(n)} \quad = \quad a \quad ,$$

wenn für jedes $\varepsilon \in \mathbb{R}^+$ ein $N \in \mathbb{N}$ existiert, so dass $x \in B_\varepsilon(a)$ für alle $n \ge N$ gilt. Der Punkt a wird in diesem Fall ein ***Grenzwert*** der Folge genannt. Eine Folge, die gegen keinen Punkt von X konvergiert, bezeichnet man als ***divergent***.

Nach Definition ist die Bedingung $x \in B_\varepsilon(a)$ äquivalent zu $d(x, a) < \varepsilon$. Die Konvergenz der Folge $(x^{(n)})_{n\in\mathbb{N}}$ ist also äquivalent dazu, dass

$$\lim_{n\to\infty} d(x^{(n)}, a) \quad = \quad 0 \qquad \text{gilt.} \tag{2}$$

Die folgende Abbildung zeigt, wie man sich die Konvergenz im $\mathbb{R}^2$ aussehen könnte: Egal wie klein die graue Umgebung des rot eingezeichneten Grenzpunktes a gewählt wird, es liegen immer alle bis auf endlich viele Punkte innerhalb der Umgebung und nur endlich viele außerhalb.



Im speziellen metrischen Raum $(\mathbb{R}, d_1)$ ist die Konvergenz einer Folge $(x^{(n)})_{n\in\mathbb{N}}$ gegen einen Punkt $a \in \mathbb{R}$ gleichbedeutend mit der Konvergenz, wie wir sie in der Analysis einer Variablen definiert haben. In diesem Fall hat die Bedingung (2) einfach die Form $\lim_n |x^{(n)} - a| = 0$. Wie in der Analysis einer Variablen gilt auch hier

(2.2) Proposition. Jede Folge in einem metrischen Raum hat höchstens einen Grenzwert.

Beweis: Sei (X, d) ein metrischer Raum und $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge in X. Nehmen wir an, dass die Folge in (X, d) die beiden Grenzwerte $a, b \in X$ besitzt, wobei $a \neq b$ ist. Sei $\varepsilon = d(a, b)$. Dann gibt es ein $N \in \mathbb{N}$, so dass zugleich $d(x^{(n)}, a) < \frac{1}{2}\varepsilon$ und $d(x^{(n)}, b) < \frac{1}{2}\varepsilon$ für alle $n \geq N$ erfüllt ist. Dies führt nun zu dem Widerspruch

$$\varepsilon \quad = \quad d(a,b) \quad \leq \quad d(a,x^{(N)}) + d(x^{(N)},b) \quad < \quad \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon \quad = \quad \varepsilon. \qquad \square$$

Bezüglich der diskreten Metrik lässt sich die Konvergenz einer Folge besonders einfach beschreiben.

(2.3) Proposition. Sei X eine beliebige Menge. Eine Folge $(x^{(n)})_{n\in\mathbb{N}}$ im metrischen Raum (X, δ_X) ausgestattet mit der diskreten Topologie konvergiert genau dann gegen einen Punkt $a \in X$, wenn ein $N \in \mathbb{N}$ mit $x^{(n)} = a$ für alle $n \geq N$ existiert.

Beweis: „⇐“ Sei N eine natürliche Zahl mit der angegebenen Eigenschaft und $\varepsilon > 0$ vorgegeben. Für alle $n \geq N$ gilt dann $\delta_X(x^{(n)}, a) = \delta_X(a, a) = 0 < \varepsilon$, also ist die Konvergenzbedingung erfüllt. „⇒“ Nach Voraussetzung gibt es für $\varepsilon = 1$ ein $N \in \mathbb{N}$, so dass $\delta_X(x^{(n)}, a) < 1$ für alle $n \geq N$ gilt. Weil δ_X nur die Werte 0 und 1 annimmt, ist $\delta_X(x^{(n)}, a) = 0$ für alle $n \geq N$. Nach Definition der diskreten Metrik bedeutet dies wiederum $x^{(n)} = a$ für alle $n \geq N$. $\square$

(2.4) Proposition. Sei V ein $\mathbb{R}$-Vektorraum mit zwei äquivalenten Normen $\|\cdot\|$ und $\|\cdot\|'$, und seien d, d' die beiden von den Normen induzierten Metriken. Sei $a \in V$ und $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge. Genau dann konvergiert die Folge $(x^{(n)})_{n\in\mathbb{N}}$ gegen a im metrischen Raum (V, d), wenn sie im metrischen Raum (V, d') gegen a konvergiert.

Beweis: Nach Definition der Äquivalenz gibt es Konstanten $\delta, \delta' \in \mathbb{R}^+$ mit $\|v\|' \leq \delta\|v\|$ und $\|v\| \leq \delta'\|v\|'$ für alle $v \in V$. Aus Symmetriegründen genügt es zu zeigen: Konvergiert $(x^{(n)})_{n\in\mathbb{N}}$ im Raum (V, d) gegen a, dann auch im

Raum (V,d'). Setzen wir ersteres also voraus. Dann gibt es für jedes $\varepsilon > 0$ ein $N \in \mathbb{N}$ mit $\|x^{(n)}-a\| = d(x^{(n)},a) < \delta^{-1}\varepsilon$ für alle $n \geq N$. Es folgt dann

$$d'(x^{(n)},a) \quad = \quad \|x^{(n)}-a\|' \quad \leq \quad \delta\|x^{(n)}-a\| \quad < \quad \delta\delta^{-1}\varepsilon \quad = \quad \varepsilon \qquad \text{für alle } n \geq N.$$

Dies zeigt, dass $(x^{(n)})_{n\in\mathbb{N}}$ im metrischen Raum (X,d') konvergiert. □

Nach (1.7) sind je zwei Normen auf einem endlich-dimensionalen $\mathbb{R}$-Vektorraum äquivalent. Sei V ein solcher Vektorraum, $\|\cdot\|$ eine beliebige Norm, d die induzierte Metrik und $X \subseteq V$ eine Teilmenge. Bezeichnen wir die Einschränkung der Abbildung d auf die Teilmenge $X \times X \subseteq V \times V$ ebenfalls mit d, so ist (X,d) ein metrischer Raum. Eine Folge in X bezeichnen wir als ***konvergent*** gegen einen Punkt $a \in X$, wenn sie im metrischen Raum (X,d) gegen a konvergiert. Nach (2.4) ist diese Definition von der Wahl der Norm $\|\cdot\|$ unabhängig.

(2.5) Satz. Sei $m \in \mathbb{N}$. Eine Folge $(x^{(n)})_{n\in\mathbb{N}}$ im $\mathbb{R}^m$ konvergiert genau dann gegen einen Punkt $a \in \mathbb{R}^m$, wenn $\lim\limits_{n\to\infty} x_k^{(n)} = a_k$ für $1 \leq k \leq m$ erfüllt ist.

Beweis: „⇐“ Wie im Absatz zuvor erläutert wurde, genügt es zu zeigen, dass $(x^{(n)})_{n\in\mathbb{N}}$ im metrischen Raum $(\mathbb{R}^m, d_\infty)$ gegen a konvergiert. Sei $\varepsilon \in \mathbb{R}^+$ vorgegeben. Auf Grund der Voraussetzung gibt es für jedes $k \in \{1,\ldots,m\}$ ein $N_k \in \mathbb{N}$, so dass jeweils $|x_k^{(n)} - a_k| < \varepsilon$ für alle $n \geq N_k$ erfüllt ist. Setzen wir $N = \max\{N_1,\ldots,N_m\}$, dann gilt für alle $n \geq N$ die Abschätzung

$$d_\infty(x^{(n)},a) \quad = \quad \|x^{(n)}-a\|_\infty \quad = \quad \max\{|x_1^{(n)}-a_1|,\ldots,|x_m^{(n)}-a_m|\} \quad < \quad \varepsilon.$$

Nach Definition bedeutet das, dass $(x^{(n)})_{n\in\mathbb{N}}$ im metrischen Raum $(\mathbb{R}^m, d_\infty)$ konvergiert.

„⇒“ Sei $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge, die im metrischen Raum $(\mathbb{R}^m, d_\infty)$ gegen a konvergiert. Zu zeigen ist $\lim_n x_k^{(n)} = a_k$ für $1 \leq k \leq m$. Sei dazu $\varepsilon \in \mathbb{R}^+$ vorgegeben und $N \in \mathbb{N}$ so gewählt, dass $d_\infty(x^{(n)},a) < \varepsilon$ für alle $n \geq N$ gilt. Dann folgt

$$|x_k^{(n)} - a_k| \quad \leq \quad \max\{|x_1^{(n)}-a_1|,\ldots,|x_m^{(n)}-a_m|\} \quad = \quad \|x^{(n)}-a\|_\infty \quad = \quad d_\infty(x^{(n)},a) \quad < \quad \varepsilon$$

für alle $n \geq N$ und $1 \leq k \leq m$. Damit ist die Behauptung bewiesen. □

Beispielsweise ist die Folge $(a_n)_{n\in\mathbb{N}}$ im $\mathbb{R}^2$ gegeben durch $a_n = (\frac{1}{n},(-1)^n)$ divergent, weil die zweite Komponente $(-1)^n$ der Folge divergiert. Die Folge $(b_n)_{n\in\mathbb{N}}$ gegeben durch

$$b_n = \left(1 + \frac{2}{n}, \frac{3n^2-7n+6}{4n^2-5}\right)$$

ist dagegen konvergent, es gilt

$$\lim_{n\to\infty} b_n \quad = \quad \left(\lim_{n\to\infty}\left(1+\frac{2}{n}\right), \lim_{n\to\infty}\frac{3n^2-7n+6}{4n^2-5}\right) \quad = \quad (1,\tfrac{3}{4}).$$

(2.6) Definition. Eine Folge $(x^{(n)})_{n\in\mathbb{N}}$ ein einem metrischen Raum (X,d) wird ***Cauchyfolge*** genannt, wenn für jedes $\varepsilon \in \mathbb{R}^+$ ein $N \in \mathbb{N}$ existiert, so dass $d(x^{(m)},x^{(n)}) < \varepsilon$ für alle $m,n \in \mathbb{N}$ mit $m,n \geq N$ gilt.

(2.7) Proposition. Sei V ein $\mathbb{R}$-Vektorraum mit zwei äquivalenten Normen $\|\cdot\|$ und $\|\cdot\|'$, und seien d, d' die beiden von den Normen induzierten Metriken. Sei $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge. Genau dann ist die Folge $(x^{(n)})_{n\in\mathbb{N}}$ eine Cauchyfolge in (V,d), wenn sie eine Cauchyfolge in (V,d') ist.

Beweis: Dieser Beweis ist dem Beweis von (2.4) über die Konvergenz sehr ähnlich. Die Ausführung der Details ist eine leichte Übungsaufgabe. □

Da je zwei Normen auf einem endlich-dimensionalen $\mathbb{R}$-Vektorraum V äquivalent sind, können wir in Teilmengen $X \subseteq V$ von Cauchyfolgen schlechthin sprechen, ohne Festlegung einer Metrik. Gemeint ist dann immer die Cauchyfolgen-Eigenschaft bezüglich der von einer (beliebig gewählten) Norm induzierten Metrik.

Jede konvergente Folge $(x^{(n)})_{n\in\mathbb{N}}$ in einem metrischen Raum (X,d) mit einem Grenzwert $a \in X$ ist auch eine Cauchyfolge. Sei nämlich $\varepsilon \in \mathbb{R}^+$ vorgegeben und $N \in \mathbb{N}$ so gewählt, dass $d(x^{(n)}, a) < \frac{1}{2}\varepsilon$ für alle $n \geq N$ erfüllt ist. Dann folgt $d(x^{(m)}, x^{(n)}) \leq d(x^{(m)}, a) + d(a, x^{(n)}) < \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon$ für alle $m, n \geq N$.

Andererseits gibt es in einigen metrischen Räumen durchaus Cauchyfolgen, die nicht konvergieren. Als Beispiel betrachten wir $(\mathbb{Q}, d)$ mit der Metrix $d(a,b) = |a-b|$. Aus der Analysis einer Variablen ist bekannt, dass jede (rationale oder irrationale) Zahl als Grenzwert einer Folge rationaler Zahlen dargestellt werden kann. Zum Beispiel gibt es eine Folge $(x^{(n)})_{n\in\mathbb{N}}$ in $\mathbb{Q}$, die gegen $\sqrt{2}$ (im gewöhnlichen Sinn) konvergiert. Diese Folge ist in $(\mathbb{Q}, d)$ eine Cauchyfolge, denn für vorgegebenes $\varepsilon \in \mathbb{R}^+$ können wir ein $N \in \mathbb{N}$ mit $|x^{(n)} - \sqrt{2}| < \frac{1}{2}\varepsilon$ für alle $n \geq N$ finden, und es gilt dann $d(x^{(m)}, x^{(n)}) = |x^{(m)} - x^{(n)}| \leq |x^{(m)} - \sqrt{2}| + |\sqrt{2} - x^{(n)}| < \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon$ für alle $m, n \geq N$. Aber andererseits ist $(x^{(n)})_{n\in\mathbb{N}}$ im metrischen Raum $(\mathbb{Q}, d)$ nicht konvergent, denn das würde bedeutet, dass $(x^{(n)})_{n\in\mathbb{N}}$ zwei verschiedene reelle Grenzwerte im gewöhnlichen Sinn besitzen würde, nämlich neben $\sqrt{2}$ noch einen rationalen. Dies ist, wie wir bereits aus der Analysis einer Variablen wissen, unmöglich.

(2.8) Definition. Ein metrischer Raum (X,d) heißt ***vollständig***, wenn jede Cauchyfolge in (X,d) konvergiert. Ein normierter $\mathbb{R}$-Vektorraum, der vollständig bezüglich der induzierten Metrik ist, wird ***Banachraum*** genannt.

In endlich-dimensionalen $\mathbb{R}$-Vektorräumen ist diese Bedingung immer erfüllt.

(2.9) Satz. Jeder normierte, endlich-dimensionale $\mathbb{R}$-Vektorraum $(V, \|\cdot\|)$ ist ein Banachraum.

Beweis: Zunächst betrachten wir den Fall $V = \mathbb{R}^d$ für ein $d \in \mathbb{N}$. Sei $(x^{(n)})_{n\in\mathbb{N}}$ eine Cauchyfolge in V. Dann ist $(x^{(n)})_{n\in\mathbb{N}}$ insbesondere eine Cauchyfolge im metrischen Raum $(\mathbb{R}^d, d_\infty)$. Wir zeigen, dass die Folgen $(x_k^{(n)})_{n\in\mathbb{N}}$ für $1 \leq k \leq d$ Cauchyfolgen im Sinne der Analysis einer Variablen sind. Sei dazu $k \in \{1, \ldots, d\}$ beliebig gewählt und $\varepsilon \in \mathbb{R}^+$ vorgegeben. Weil $(x^{(n)})_{n\in\mathbb{N}}$ eine Cauchyfolge ist, gibt es ein $N \in \mathbb{N}$, so dass $d_\infty(x^{(m)}, x^{(n)}) < \varepsilon$ für alle $m, n \geq N$ erfüllt ist. Es folgt

$$|x_k^{(m)} - x_k^{(n)}| \quad \leq \quad \max\{|x_1^{(m)} - x_1^{(n)}|, \ldots, |x_d^{(m)} - x_d^{(n)}|\} \quad = \quad \|x^{(m)} - x^{(n)}\|_\infty \quad = \quad d_\infty(x^{(m)}, x^{(n)}) \quad < \quad \varepsilon$$

für alle $m, n \geq N$. Also ist jede Folge $(x_k^{(n)})_{n\in\mathbb{N}})$ tatsächlich eine Cauchyfolge in $\mathbb{R}$. Aus der Analysis einer Variablen ist bekannt, dass jede Cauchyfolge in $\mathbb{R}$ konvergiert. Es gibt also $a_1, ..., a_d \in \mathbb{R}$, so dass

$$\lim_{n\to\infty} x_k^{(n)} \quad = \quad a_k \qquad \text{für} \quad 1 \leq k \leq d$$

erfüllt ist. Nach (2.5) konvergiert die Folge $(x^{(n)})_{n\in\mathbb{N}}$ im metrischen Raum $(\mathbb{R}^d, d_\infty)$ gegen den Punkt $a = (a_1, ..., a_d)$.

Sei nun V ein beliebiger endlich-dimensionaler $\mathbb{R}$-Vektorraum. Wegen (2.4) und (2.7) genügt es zu zeigen, dass V bezüglich irgendeiner Norm vollständig ist, die wir frei wählen können. Weil $d = \dim V$ endlich ist, gibt es einen Isomorphismus $\phi : \mathbb{R}^d \to V$ von $\mathbb{R}$-Vektorräumen. Nach (1.8) ist auf V durch $\|v\| = \|\phi^{-1}(v)\|_\infty$ auf V eine Norm definiert. Sei nun $(y^{(n)})_{n\in\mathbb{N}}$ eine Cauchyfolge in V bezüglich der durch $\|\cdot\|$ induzierten Metrik, die wir mit d_V bezeichnen. Für jedes $n \in \mathbb{N}$ sei $x^{(n)} = \phi^{-1}(y^{(n)})$. Für alle $m, n \in \mathbb{N}$ gilt

$$d_\infty(x^{(m)}, x^{(n)}) \quad = \quad d_\infty(\phi^{-1}(y^{(m)}), \phi^{-1}(y^{(n)})) \quad = \quad \|\phi^{-1}(y^{(m)}) - \phi^{-1}(y^{(n)})\|_\infty \quad =$$
$$\|y^{(m)} - y^{(n)}\| \quad = \quad d_V(y^{(m)}, y^{(n)})$$

nach Definition, also ergibt sich aus der Cauchyfolgen-Eigenschaft von $(y^{(n)})_{n\in\mathbb{N}}$ in (V, d_V) dieselbe Eigenschaft für die Folge $(x^{(n)})_{n\in\mathbb{N}}$ in $(\mathbb{R}^d, d_\infty)$. Wie bereits gezeigt, konvergiert die Cauchyfolge $(x^{(n)})_{n\in\mathbb{N}}$ in $(\mathbb{R}^d, d_\infty)$ gegen einen Grenzwert $a \in \mathbb{R}^d$. Sei nun $b = \phi(a) \in V$. Für alle $n \in \mathbb{N}$ gilt dann

$$d_\infty(x^{(n)}, a) \quad = \quad d_\infty(\phi^{-1}(y^{(n)}), \phi^{-1}(b)) \quad = \quad \|\phi^{-1}(y^{(n)}) - \phi^{-1}(b)\|_\infty \quad = \quad \|y^{(n)} - b\| \quad = \quad d_V(y^{(n)}, b).$$

Aus der Konvergenz von $(x^{(n)})_{n\in\mathbb{N}}$ gegen a im metrischen Raum $(\mathbb{R}^d, d_\infty)$ ergibt sich also die Konvergenz von $(y^{(n)})_{n\in\mathbb{N}}$ gegen b in (V, d_V). □

Wir betrachten nun eine wichtige Situation, in der die Vollständigkeit eines metrischen Raumes verwendet wird.

(2.10) Definition. Sei (X, d) ein metrischer Raum. Eine Abbildung $\phi : X \to X$ wird ***Kontraktion*** genannt, wenn eine Konstante $\gamma \in]0, 1[$ existiert, so dass $d(\phi(x), \phi(y)) \leq \gamma d(x, y)$ für alle $x, y \in X$ erfüllt ist.

Diese Eigenschaft einer Abbildung ist entscheidend für den

(2.11) Satz. *(Banachscher Fixpunktsatz)*

Sei (X, d) ein vollständiger metrischer Raum. Dann besitzt jede Kontraktion $\phi : X \to X$ genau einen Fixpunkt. Es gibt also ein eindeutig bestimmtes $z \in X$ mit $\phi(z) = z$.

Beweis: Existenz: Sei $\gamma \in \mathbb{R}^+$ eine Konstante mit $\gamma < 1$ und $d(\phi(x), \phi(y)) \leq \gamma d(x, y)$ für alle $x, y \in X$. Wir wählen $x^{(0)} \in X$ beliebig und definieren rekursiv $x^{(n+1)} = \phi(x^{(n)})$ für alle $n \in \mathbb{N}$. Als erstes schätzen wir nun die Abstände $d(x^{(n)}, x^{(n+p)})$ für beliebige $n, p \in \mathbb{N}$ ab. Für jedes $n \in \mathbb{N}$ gilt

$$d(x^{(n+1)}, x^{(n+2)}) \quad = \quad d(\phi(x^{(n)}), \phi(x^{(n+1)})) \quad \leq \quad \gamma d(x^{(n)}, x^{(n+1)}) \quad ,$$

und durch vollständige Induktion über p zeigt man leicht

$$d(x^{(n+p)}, x^{(n+p+1)}) \quad \leq \quad \gamma^p d(x^{(n)}, x^{(n+1)}).$$

Mit der Summenformel $\sum_{k=0}^{p-1} \gamma^k = (1-\gamma^p)/(1-\gamma)$ aus der Analysis einer Variablen und der Dreiecksungleichung erhalten wir für alle $p \in \mathbb{N}$ jeweils

$$d(x^{(n)}, x^{(n+p)}) \quad \leq \quad \sum_{k=0}^{p-1} d(x^{(n+k)}, x^{(n+k+1)}) \quad \leq \quad \sum_{k=0}^{p-1} \gamma^k d(x^{(n)}, x^{(n+1)}) \quad =$$

$$\frac{1-\gamma^p}{1-\gamma} d(x^{(n)}, x^{(n+1)}) \quad \leq \quad \frac{1}{1-\gamma} d(x^{(n)}, x^{(n+1)}) \quad \leq \quad \frac{\gamma^n}{1-\gamma} d(x^{(0)}, x^{(1)}).$$

Wegen $\lim_n \gamma^n = 0$ ist $(x^{(n)})_{n\in\mathbb{N}}$ somit eine Cauchyfolge in X. Auf Grund der Vollständigkeit von (X,d) existiert der Grenzwert $z = \lim\limits_{n\to\infty} x^{(n)}$.

Um zu zeigen, dass z ein Fixpunkt von ϕ ist, beweisen wir, dass die Folge $(x^{(n)})_{n\in\mathbb{N}}$ auch gegen $\phi(z)$ konvergiert. Ist $\varepsilon \in \mathbb{R}^+$ vorgegeben, dann existiert ein $N \in \mathbb{N}$ mit $d(x^{(n)}, z) < \varepsilon$ für alle $n \geq N$. Daraus folgt

$$d(x^{(n+1)}, \phi(z)) \quad = \quad d(\phi(x^{(n)}), \phi(z)) \quad \leq \quad \gamma d(x^{(n)}, z) \quad < \quad \gamma\varepsilon \quad < \quad \varepsilon$$

für alle $n \geq N$. Es folgt $\phi(z) = \lim_n x^{(n+1)} = \lim_n x^{(n)} = z$.

Eindeutigkeit: Angenommen, $z' \in X$ ist ein weiterer Fixpunkt von ϕ. Aus $\phi(z) = z$ und $\phi(z') = z'$ folgt dann $d(z,z') = d(\phi(z), \phi(z')) \leq \gamma d(z,z')$. Wegen $\gamma < 1$ ist dies nur für $d(z,z') = 0$, also $z = z'$, möglich. □

(2.12) Proposition. Für den Abstand der Folgenglieder zum Fixpunkt z hat man die „a priori"-Abschätzung

$$d(x^{(n)}, z) \quad \leq \quad \frac{\gamma^n}{1-\gamma} d(x^{(0)}, x^{(1)}).$$

Beweis: Es gilt die Abschätzung

$$\begin{aligned} d(x^{(n)}, z) \quad \leq \quad d(x^{(n)}, x^{(n+1)}) + d(x^{(n+1)}, z) \quad &= \quad d(x^{(n)}, x^{(n+1)}) + d(\phi(x^{(n)}), \phi(z)) \\ &\leq \quad \gamma^n d(x^{(0)}, x^{(1)}) + \gamma d(x^{(n)}, z) \quad , \end{aligned}$$

was zu $(1-\gamma) d(x^{(n)}, z) \leq \gamma^n d(x^{(0)}, x^{(1)}) \Leftrightarrow d(x^{(n)}, z) \leq \frac{\gamma^n}{1-\gamma} d(x^{(0)}, x^{(1)})$ umgeformt werden kann. □

Als Anwendungsbeispiel setzen wir uns zum Ziel, den Wert von $\sqrt{2}$ numerisch durch Anwendung des Banachschen Fixpunktsatzes mit hoher Genauigkeit zu bestimmen. Der naheliegende Ansatz, $\sqrt{2}$ als Fixpunkt der Abbildung $\phi(x) = \frac{2}{x}$ zu betrachten, führt nicht zum Ziel, weil diese Abbildung in einer Umgebung von $\sqrt{2}$ keine Kontraktion ist. Statt dessen definieren wir $\phi : \mathbb{R}^+ \to \mathbb{R}$ durch $\phi(x) = \frac{1}{2}(x + \frac{2}{x})$. Die Zahl $\sqrt{2}$ ist auch ein Fixpunkt dieser Abbildung, wie die Rechnung $\phi(\sqrt{2}) = \frac{1}{2}(\sqrt{2} + \frac{2}{\sqrt{2}}) = \frac{1}{2}(\sqrt{2} + \sqrt{2}) = \frac{1}{2}(2\sqrt{2}) = \sqrt{2}$ zeigt.

Zunächst weisen wir nach, dass ϕ auf dem Intervall $X = [1, \frac{3}{2}]$ eine Kontraktion ist. Es gilt $\phi'(x) = \frac{1}{2} - \frac{1}{x^2}$ für alle $x \in \mathbb{R}^+$. Auf dem offenen Intervall $]1, \frac{3}{2}[$ gilt die Abschätzung

$$-\tfrac{1}{2} \quad = \quad \phi'(1) \quad < \quad \phi'(x) \quad < \quad \tfrac{1}{18} \quad = \quad \phi'(\tfrac{3}{2})$$

und somit $|\phi'(x)| < \frac{1}{2}$ für alle $x \in]1, \frac{3}{2}[$. Seien nun $x, y \in [1, \frac{3}{2}]$ vorgegeben. Nach dem Mittelwertsatz der Differentialrechung gibt es ein $x_0 \in]1, \frac{3}{2}[$ mit

$$\phi'(x_0) = \frac{\phi(x) - \phi(y)}{x - y} \quad \Leftrightarrow \quad \phi(x) - \phi(y) = \phi'(x_0)(x - y).$$

Es folgt $|\phi(x) - \phi(y)| = |\phi'(x_0)||x - y| \leq \frac{1}{2}|x - y|$. Also ist ϕ auf X tatsächlich eine Kontraktion, und mit der Konstanten $\gamma = \frac{1}{2}$ ist die Abschätzung $|\phi(x) - \phi(y)| \leq \gamma|x - y|$ erfüllt. Definieren wir nun $x^{(0)} = 1{,}5$ und $x^{(n+1)} = \phi(x^{(n)})$ für alle $n \in \mathbb{N}$, dann erhalten wir die Werte

n	$x^{(n)}$
0	1,5
1	1,41666666666666667
2	1,41421568627450980
3	1,41421356237468991
4	1,41421356237309505

Die Abstand $|x^{(0)} - x^{(1)}|$ kann durch den Wert 0,084 nach oben abgeschätzt werden. Auf Grund der Fehlerabschätzung (2.12) gilt nach vier Schritten $|x^{(4)} - \sqrt{2}| < 0{,}105$, nach zwanzig Schritten $|x^{(20)} - \sqrt{2}| < 1{,}6 \cdot 10^{-7}$. Die Tabelle zeigt aber, dass die Approximation in der Praxis deutlich besser ist: Es gilt $\sqrt{2} \approx 1{,}41421356237309504880$, also sind schon für das Folgenglied $x^{(4)}$ alle 17 angegebenen Nachkommastellen korrekt. Statt in der Größenordnung 0,1 beträgt der tatsächliche Fehler also höchstens $0{,}5 \cdot 10^{-17}$.

§ 3. Stetigkeit

Zusammenfassung. In der Analysis einer Variablen haben wir die Stetigkeit einer Funktion $f : \mathbb{R} \to \mathbb{R}$ in einem Punkt $a \in \mathbb{R}$ zunächst mit Hilfe von Folgen definiert: Für jede Folge $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ mit $\lim_n x_n = a$ muss auch $\lim_n f(x_n) = f(a)$ gelten. Genauso definieren wir auch die Stetigkeit einer Abbildung $f : X \to Y$ zwischen metrischen Räume (X, d_X) und (Y, d_Y) in einem Punkt $a \in X$. Sind (X, d_X) und (Y, d_Y) beide $\mathbb{R}$ mit der Standardmetrik $d(a, b) = |b - a|$, dann stimmt die neue Definition der Stetigkeit mit der alten überein. Auch das bereits bekannte ε-δ-Kriterium lässt sich für metrische Räume formulieren.

Ein *Homömorphismus* ist eine stetige bijektive Abbildung zwischen metrischen Räumen, deren Umkehrabbildung ebenfalls stetig ist. Metrische Räume, zwischen denen eine solche Abbildung existiert, nennt man *homöomorph*. Diese Beziehung spielt eine wichtige Rolle, wenn man die Gesamtheit der stetigen Funktionen auf einem metrischen Raum studieren möchte, weil man dafür zu einem beliebigen, eventuell leichter handhabbaren Raum wechseln kann. Für Anwendungen in der Physik besonders nützliche Homömorphismen erhält man durch diverse Koordinatenabbildungen (Polar-, Zylinder- und Kugelkoordinaten).

Am Ende des Kapitels befassen wir uns noch mit der Stetigkeit linearer Abbildungen. Lineare Abbildungen zwischen endlich-dimensionalen $\mathbb{R}$-Vektorräumen sind immer stetig. Bei Anwendungen in der Theorie der Differenzialgleichungen und in der Funktionalanalysis (und auch in der Physik) stößt man aber häufig auch auf unendlich-dimensionale Räume, bei denen die Stetigkeit einer linearen Abbildungen für viele Anwendungen relevant ist. Die in diesem Zusammenhang autretende *Operatornorm* wird häufig auch bei endlich-dimensionalen Problemen verwendet.

Wichtige Begriffe und Sätze in diesem Kapitel

- Definition der Stetigkeit und der Funktionsgrenzwerte für Abbildungen zwischen metrischen Räumen
- ε-δ-Kriterium für Stetigkeit
- Stetigkeit bleibt erhalten unter punktweiser Addition, Multiplikation, Division und unter Komposition
- Homöomorphismen zwischen metrischen Räumen
- Polar-, Zylinder- und Kugelkoordinaten
- Stetigkeit linearer Abbildungen, Operatornorm

(3.1) Definition. Seien (X, d_X) und (Y, d_Y) metrische Räume. Eine Abbildung $f : X \to Y$ wird ***stetig*** in einem Punkt $a \in X$ bezüglich der Metriken d_X und d_Y genannt, wenn für jede Folge $(x^{(n)})_{n\in\mathbb{N}}$ die Implikation

$$\lim_{n\to\infty} x^{(n)} = a \quad \text{in} \quad (X, d_X) \qquad \Rightarrow \qquad \lim_{n\to\infty} f(x^{(n)}) = f(a) \quad \text{in} \quad (Y, d_Y) \qquad \text{gilt.}$$

Wir bezeichnen f insgesamt als stetig, wenn f in jedem Punkt $x \in X$ stetig ist.

Eine Funktion $f : X \to \mathbb{R}$ auf einem metrischen Raum (X, d_X) bezeichnen wir als *stetig*, wenn sie bezüglich d_X und der von der 1-Norm induzierten Metrik $d_1(a, b) = |a - b|$ auf $\mathbb{R}$ stetig ist. Ist auch X eine Teilmenge von $\mathbb{R}$ und $d_X = d_1$, dann stimmt der Stetigkeitsbegriff mit dem Begriff aus der Analysis einer Variablen überein.

(3.2) Proposition. Seien (X, d_X), (Y, d_Y) und (Z, d_Z) metrische Räume.

(i) Jede konstante Funktion auf X ist stetig.

(ii) Seien $f : X \to Y$ und $g : Y \to Z$ Abbildungen und $a \in X$ ein Punkt, so dass f in a und g in $f(a)$ stetig ist. Dann ist auch $g \circ f$ in a stetig.

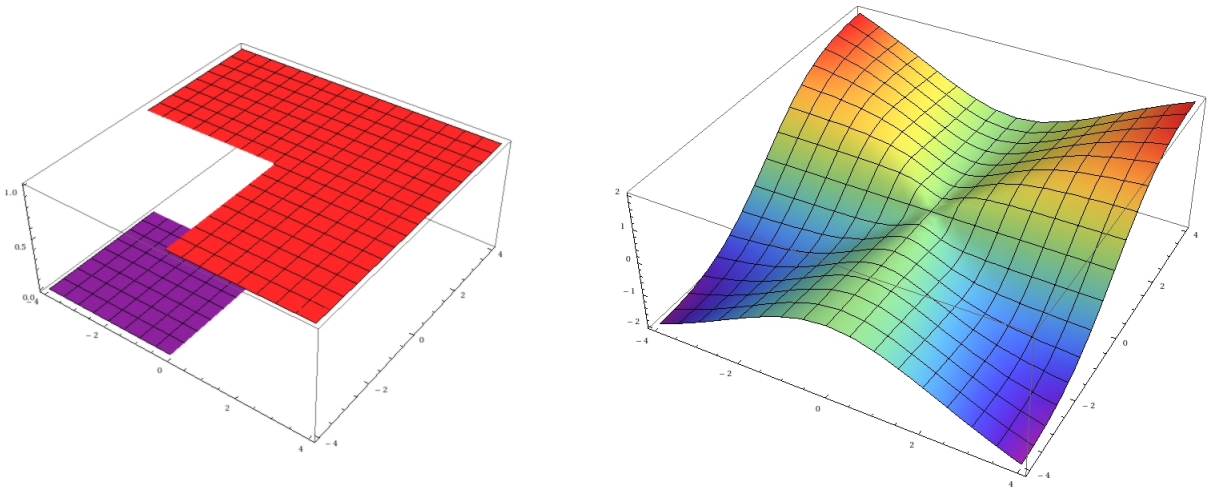
Beweis: zu (i) Sei $c \in \mathbb{R}$ und $f : X \to \mathbb{R}$ gegeben durch $f(x) = c$ für alle $x \in X$. Sei außerdem $a \in X$ ein beliebig gewählter Punkt und $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge mit $\lim_n x^{(n)} = a$. Dann gilt

$$\lim_{n\to\infty} f(x^{(n)}) \quad = \quad \lim_{n\to\infty} c \quad = \quad c \quad = \quad f(a).$$

zu (ii) Sei $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge in X mit $\lim_n x^{(n)} = a$. Weil f in a stetig ist, gilt $f(a) = \lim_n f(x^{(n)})$, und auf Grund der Stetigkeit von g im Punkt $f(a)$ erhalten wir

$$(g \circ f)(a) \quad = \quad g(f(a)) \quad = \quad \lim_{n\to\infty} g(f(x^{(n)})) \quad = \quad \lim_{n\to\infty} (g \circ f)(x^{(n)}).$$

Damit ist die Stetigkeit von $g \circ f$ in a bewiesen. □



Funktionsgraph einer unstetigen und eine stetigen Funktion auf dem $\mathbb{R}^2$.
Die unstetige Funktion besitzt unendlich viele Unstetigkeitsstellen, nämlich alle $(x, 0)$ mit $x \leq 0$ und alle $(0, y)$ mit $y \leq 0$

Auch der Grenzwertbegriff für Funktionen lässt sich auf beliebige metrische Räume übertragen.

(3.3) Definition. Seien (X, d_X) und (Y, d_Y) metrische Räume, $f : D \to Y$ eine Abbildung auf einer Teilmenge $D \subseteq X$ und a ein Punkt in $X \setminus D$. Wir bezeichnen $b \in Y$ als ***Grenzwert*** von f für $x \to a$, wenn eine Folge $(x^{(n)})$ in D existiert, so dass $\lim_n x^{(n)} = a$ in (X, d_X) gilt und außerdem für jede solche Folge jeweils

$$\lim_{n\to\infty} f(x^{(n)}) = b \qquad \text{in } (Y, d_Y) \text{ erfüllt ist.}$$

Sind X und Y Teilmengen von endlich-dimensionalen $\mathbb{R}$-Vektorräumen, dann bezeichnen wir eine Funktion $f : X \to Y$ als *stetig* in einem Punkt $a \in X$, wenn sie bezüglich der induzierten Metriken von beliebig gewählten Normen auf X und Y stetig ist. Auf Grund der in (2.4) formulierten Unabhängigkeit der Konvergenz von der gewählten Norm hat die Wahl der Norm keinen Einfluss auf die Stetigkeitseigenschaft. Wichtig ist nur, dass die Metriken d_X und d_Y überhaupt von einer Norm induziert werden.

(3.4) Proposition. Die Abbildung $\pi_i : \mathbb{R}^m \to \mathbb{R}$, $(x_1, ..., x_m) \mapsto x_i$ ist stetig, für $1 \le i \le m$.

Beweis: Sei $a \in \mathbb{R}^m$ und $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge mit $\lim_{n\to\infty} x^{(n)} = a$. Nach (2.5) gilt dann insbesondere $\lim\limits_{n\to\infty} x_i^{(n)} = a_i$, also insgesamt $\lim_n \pi_i(x^{(n)}) = \lim_n x_i^{(n)} = a_i = \pi_i(a)$. □

(3.5) Satz. *(ε-δ-Kriterium)*

Seien (X, d_X) und (Y, d_Y) metrische Räume. Eine Abbildung $f : X \to Y$ ist genau dann stetig im Punkt a bezüglich d_X und d_Y, wenn für jedes $\varepsilon \in \mathbb{R}^+$ ein $\delta \in \mathbb{R}^+$ existiert, so dass die Implikation

$$d_X(a, x) < \delta \quad \Rightarrow \quad d_Y(f(a), f(x)) < \varepsilon \qquad \text{für alle } x \in X \text{ erfüllt ist.}$$

Beweis: „⇐“ Sei $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge mit $\lim_n x^{(n)} = a$. Zu zeigen ist $\lim_n f(x^{(n)}) = f(a)$. Sei $\varepsilon \in \mathbb{R}^+$ vorgegeben und $\delta \in \mathbb{R}^+$ so gewählt, dass die Implikation $d_X(a, x) < \delta \Rightarrow d_Y(f(a), f(x)) < \varepsilon$ erfüllt ist. Auf Grund der Konvergenz von $(x^{(n)})_{n\in\mathbb{N}}$ gibt es ein $N \in \mathbb{N}$, so dass $d_X(a, x_n) < \delta$ für alle $n \ge N$ erfüllt ist. Es folgt dann $d_Y(f(a), f(x_n)) < \varepsilon$ für alle $n \ge N$.

„⇒“ Nehmen wir an, dass die Funktion f in a stetig, aber das ε-δ-Kriterium nicht erfüllt ist. Dann gibt es ein $\varepsilon \in \mathbb{R}^+$ mit der Eigenschaft, dass für jedes $\delta \in \mathbb{R}^+$ ein $x \in X$ mit $d_X(a, x) < \delta$ aber $d_Y(f(a), f(x)) \ge \varepsilon$ existiert. Insbesondere gibt es für $n \in \mathbb{N}$ ein $x^{(n)} \in X$ mit $d_X(a, x^{(n)}) < \frac{1}{n}$ und $d_Y(f(a), f(x^{(n)})) \ge \varepsilon$. Es ist dann $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge mit $\lim_n x^{(n)} = a$, aber die Folge $(f(x^{(n)}))_{n\in\mathbb{N}}$ konvergiert nicht gegen $f(a)$, im Widerspruch zur Stetigkeit. □

(3.6) Proposition. Sei (X, d_X) ein metrischer Raum, $r \in \mathbb{N}$ und $f : X \to \mathbb{R}^d$ eine Funktion mit den Komponenten $f_1, ..., f_d : X \to \mathbb{R}$, so dass also $f(x) = (f_1(x), ..., f_d(x))$ für alle $x \in X$ gilt. Genau dann ist f in einem Punkt $a \in X$ stetig, wenn die Funktionen $f_1, ..., f_d$ alle in a stetig sind.

Beweis: „⇒“ Sei $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge, die in (X, d_X) gegen a konvergiert. Weil die Funktion f in a stetig ist, gilt $\lim_n f(x^{(n)}) = f(a)$. Nach (2.5) folgt daraus $\lim_n f_k(x^{(n)}) = f_k(a)$ für $1 \le k \le d$. Dies wiederum bedeutet, dass die Funktionen $f_1, ..., f_d$ in a stetig sind. „⇐“ Sei wieder $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge mit $\lim_n x^{(n)} = a$. Nach Voraussetzung sind die Funktionen $f_1, ..., f_d$ in a stetig, es gilt also $\lim_n f_k(x^{(n)}) = f_k(a)$ für $1 \le k \le n$. Nach (2.5) folgt daraus $\lim_n f(x^{(n)}) = a$. Also ist f im Punkt a stetig. □

(3.7) Proposition. Die folgenden Abbildungen $\alpha : \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto x + y$, $\mu : \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto xy$ und $\delta : \mathbb{R} \times \mathbb{R}^\times \to \mathbb{R}$, $(x, y) \mapsto \frac{x}{y}$ sind stetig.

Beweis: Sei $(a, b) \in \mathbb{R} \times \mathbb{R}$ und $((x^{(n)}, y^{(n)}))_{n\in\mathbb{N}}$ eine Folge in $\mathbb{R} \times \mathbb{R}$ mit $\lim_{n\to\infty}(x^{(n)}, y^{(n)}) = (a, b)$. Nach (2.5) gilt dann $\lim_n x^{(n)} = a$ und $\lim_n y^{(n)} = b$. Auf Grund der Grenzwertsätze aus der Analysis einer Variablen gilt

$$\lim_{n\to\infty} \alpha(x^{(n)}, y^{(n)}) \quad = \quad \lim_{n\to\infty} (x^{(n)} + y^{(n)}) \quad = \quad \lim_{n\to\infty} x^{(n)} + \lim_{n\to\infty} y^{(n)} \quad = \quad a + b \quad = \quad \alpha(a, b)$$

und ebenso

$$\lim_{n\to\infty} \mu(x^{(n)}, y^{(n)}) \quad = \quad \lim_{n\to\infty} x^{(n)} y^{(n)} \quad = \quad \left(\lim_{n\to\infty} x^{(n)}\right) \cdot \left(\lim_{n\to\infty} y^{(n)}\right) \quad = \quad ab \quad = \quad \mu(a, b)$$

Genauso leitet man die Stetigkeit von δ aus dem Grenzwertsatz für Quotientenfolgen ab. □

(3.8) Folgerung. Sei (X, d_X) ein metrischer Raum, und seien $f, g : X \to \mathbb{R}$ stetige Funktionen. Dann sind auch die Funktionen

$$(f+g)(x) = f(x) + g(x) \qquad \text{und} \qquad (fg)(x) = f(x)g(x) \qquad \text{stetig.}$$

Gilt zusätzlich $g(x) \neq 0$ für alle $x \in X$, dann ist auch $(f/g)(x) = f(x)/g(x)$ stetig.

Beweis: Nach (3.6) ist die Abbildung $(f, g) : X \to \mathbb{R}^2$ gegeben durch $(f, g)(x) = (f(x), g(x))$ stetig. Nach Definition gilt $f + g = \alpha \circ (f, g)$, $fg = \mu \circ (f, g)$ und $f/g = \delta \circ (f, g)$. Weil die Komposition stetiger Abbildungen nach (3.2) (ii) stetig ist, sind also $f + g$, fg und f/g stetige Funktionen. □

Als Anwendungsbeispiel zeigen wir

(3.9) Proposition. Die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto \dfrac{x^3y + 5xy}{x^2 + y^4 + 1}$ ist stetig.

Beweis: Wie wir in (3.4) gezeigt haben, sind die Abbildungen $(x, y) \mapsto x$ und $(x, y) \mapsto y$ stetig. Ebenso ist nach (3.2) (i) die konstante Abbildung $(x, y) \mapsto 5$ eine stetige Funktion. Weil die Abbildung $(x, y) \mapsto x^2$ durch punktweise Multiplikation der Abbildung $(x, y) \mapsto x$ mit sich selbst zu Stande kommt, können wir (3.8) anwenden und erhalten die Stetigkeit von $(x, y) \mapsto x^2$. Die Abbildung $(x, y) \mapsto x^3$ kommt durch punktweise Multiplikation von $(x, y) \mapsto x^2$ und $(x, y) \mapsto x$ zu Stande. Durch eine weitere Anwendung der Folgerung erhält man die Stetigkeit von $(x, y) \mapsto x^3$. Indem man weiter so vorgeht, beweist man nacheinander die Stetigkeit von $(x, y) \mapsto x^3y$, $(x, y) \mapsto 5x$, $(x, y) \mapsto 5xy$ und $(x, y) \mapsto x^3y + 5xy$. Genauso beweist man die Stetigkeit der Funktion $(x, y) \mapsto x^2 + y^4 + 1$ im Nenner. Diese ist offenbar für alle $(x, y) \in \mathbb{R}^2$ ungleich Null. Durch eine letzte Anwendung von (3.8) erhält man schließlich die Stetigkeit von f. □

(3.10) Lemma. Sei $(V, \|\cdot\|)$ ein normierter $\mathbb{R}$-Vektorraum. Dann gilt

$$|\|v\| - \|w\|| \quad \leq \quad \|v - w\| \qquad \text{für alle} \quad v, w \in V.$$

Beweis: Seien $v, w \in V$. Dann gilt auf Grund der Dreiecksungleichung einerseits $\|v\| = \|w + (v - w)\| \leq \|w\| + \|v - w\|$, also $\|v\| - \|w\| \leq \|v - w\|$. Andererseits gilt auch $\|w\| = \|v - (w - v)\| \leq \|v\| + \|w - v\| = \|v\| + \|v - w\|$ und somit $\|w\| - \|v\| \leq \|v - w\|$. Da der Betrag $|\|v\| - \|w\||$ immer mit $\|v\| - \|w\|$ oder $\|w\| - \|v\|$ übereinstimmt, erhalten wir insgesamt $|\|v\| - \|w\|| \leq \|v - w\|$. □

(3.11) Folgerung. Sei $(V, \|\cdot\|)$ ein normierter $\mathbb{R}$-Vektorraum. Dann ist $V \to \mathbb{R}$, $x \mapsto \|x\|$ eine stetige Funktion.

Beweis: Sei $v \in V$ und $(x^{(n)})$ eine Folge in V mit $\lim_n x^{(n)} = v$. Zu zeigen ist $\lim_n \|x^{(n)}\| = \|v\|$. Sei $\varepsilon \in \mathbb{R}^+$ vorgegeben. Nach Definition gibt es ein $N \in \mathbb{N}$ mit $\|x^{(n)} - v\| < \varepsilon$ für alle $n \geq N$. Durch (3.10) folgt $|\|x^{(n)}\| - \|v\|| \leq \|x^{(n)} - v\| < \varepsilon$ für alle $n \geq N$. □

(3.12) Definition. Eine Abbildung $f : X \to Y$ zwischen metrischen Räumen (X, d_X) und (Y, d_Y) wird ***Homöomorphismus*** genannt, wenn sie bijektiv, stetig und die Umkehrabbildung $f^{-1} : Y \to X$ ebenfalls stetig ist. Metrische Räume, zwischen denen ein Homöomorphismus existiert, nennt man ***homöomorph***.

Das folgende Beispiel zeigt, dass Homöomorphismen auch zwischen Räumen ganz unterschiedlicher Gestalt existieren können.

(3.13) Proposition. Sei $\|\cdot\|$ eine beliebige Norm auf $V = \mathbb{R}^n$. Dann ist der offene Ball $B_1(0_V)$ vom Radius 1 um den Ursprung homöomorph zum $\mathbb{R}^n$.

Beweis: Die Abbildungen $f : B_1(0_V) \to \mathbb{R}^n$ und $g : \mathbb{R}^n \to B_1(0_V)$ gegeben duch

$$f(x) \quad = \quad \frac{1}{1-\|x\|}x \qquad \text{und} \qquad g(x) \quad = \quad \frac{1}{1+\|x\|}x$$

sind nach (3.11) und (3.8) beide stetig. Wir überprüfen, dass sie außerdem zueinander invers sind, woraus dann die Bijektivität folgt. Für alle $x \in B_1(0_V)$ gilt $\|x\| < 1$ und $\|f(x)\| = \frac{\|x\|}{1-\|x\|}$, somit

$$\frac{1}{1+\|f(x)\|} \quad = \quad \frac{1}{1+\frac{\|x\|}{1-\|x\|}} \quad = \quad \frac{1-\|x\|}{1-\|x\|+\|x\|} \quad = \quad 1-\|x\|$$

und folglich $(g \circ f)(x) = g(f(x)) = \frac{1}{1+\|f(x)\|} f(x) = (1-\|x\|) \cdot \frac{1}{1-\|x\|} \cdot x = x$. Für jedes $x \in \mathbb{R}^n$ gilt $\|g(x)\| = \frac{\|x\|}{1+\|x\|}$, somit

$$\frac{1}{1-\|g(x)\|} \quad = \quad \frac{1}{1-\frac{\|x\|}{1+\|x\|}} \quad = \quad \frac{1+\|x\|}{1+\|x\|-\|x\|} \quad = \quad 1+\|x\|$$

und $(f \circ g)(x) = f(g(x)) = \frac{1}{1-\|g(x)\|} g(x) = (1+\|x\|) \cdot \frac{1}{1+\|x\|} \cdot x = x$. Insgesamt haben wir damit $g \circ f = \mathrm{id}_{B_1(0_V)}$ und $f \circ g = \mathrm{id}_{\mathbb{R}^n}$ nachgewiesen. □

Wir betrachten nun eine Reihe von Abbildungen, die besonders für physikalische und technische Anwendungen von Interesse sind. Im Folgenden bezeichnen wir eine Teilmenge von $\mathbb{R}$ der Form $[a, b[$ oder $]a, b]$ mit $a, b \in \mathbb{R}$ und $a < b$ als *halboffenes Intervall* der Länge $b-a$. Zugleich ist $b-a$ auch die Länge des offenen Intervalls $]a, b[$ und des abgeschlossenen Intervalls $[a, b]$.

(3.14) Satz. *(Definition der Polarkoordinaten)*

(i) Die Abbildung $\phi : \mathbb{R} \to \mathbb{R}^2$ gegeben durch $\varphi \mapsto (\cos(\varphi), \sin(\varphi))$ ist stetig, und für jedes halboffene Intervall I der Länge 2π ist die eingeschränkte Abbildung $\varphi|_I$ eine stetige Bijektion auf ihre Bildmenge.

(ii) Die Abbildung $\rho_{\mathrm{pol}} : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}^2$, $(r, \varphi) \mapsto (r\cos(\varphi), r\sin(\varphi))$ wird ***Polarkoordinaten-Abbildung*** genannt. Für jedes Intervall I wie unter (i) ist auch $\rho_{\mathrm{pol}}|_{\mathbb{R}^+ \times I}$ eine stetige Bijektion auf ihre Bildmenge.

Beweis: zu (i) Die Abbildung ϕ ist nach (3.6) stetig, weil die Komponentenfunktionen $\varphi \mapsto \cos(\varphi)$ und $\varphi \mapsto \sin(\varphi)$ stetig sind. Zu zeigen ist nun nur noch die Injektivität von $\varphi|_I$ für ein Intervall der Form $I = [\varphi_0, \varphi_0 + 2\pi[$ oder $I =]\varphi_0, \varphi_0 + 2\pi]$, für ein beliebiges $\varphi_0 \in \mathbb{R}$. Wir beschränken uns auf den ersten Fall, da der andere vollkommen analog behandelt wird. Seien also $\varphi_1, \varphi_2 \in I$ mit $\phi(\varphi_1) = \phi(\varphi_2)$; zu zeigen ist $\varphi_1 = \varphi_2$. Aus der Voraussetzung folgt $\cos(\varphi_1) = \cos(\varphi_2)$ und $\sin(\varphi_1) = \sin(\varphi_2)$.

Nun wählen wir $n_1, n_2 \in \mathbb{Z}$ so, dass $\varphi_1' = \varphi_1 - 2n_1\pi$ und $\varphi_2' = \varphi_2 - 2n_2\pi$ beide im Intervall $[-\pi, \pi[$ liegen. Es gilt dann auch $\cos(\varphi_1') = \cos(\varphi_2')$ und $\sin(\varphi_1') = \sin(\varphi_2')$. Weil die Kosinusfunktion auf $[0, \pi]$ injektiv ist und außerdem $\cos(x) = \cos(-x)$ für alle $x \in \mathbb{R}$ gilt, folgt aus der ersten Gleichung $\varphi_2' \in \{\pm\varphi_1'\}$. Weil aber $\sin(x) > 0$ für alle $x \in]0, \pi[$ und $\sin(x) < 0$ für alle $x \in]-\pi, 0[$ gilt, ist auf Grund der zweiten Gleichung nur $\varphi_2' = \varphi_1'$ möglich. Es folgt $\varphi_2 = \varphi_2' + 2n_2\pi = \varphi_1' + 2n_2\pi = \varphi_1 + 2(n_2 - n_1)\pi$. Weil aber $\varphi_1 + 2m\pi$ im Fall $m \neq 0$ außerhalb des Intervalls I liegen müsste, im Widerspruch zu $\varphi_2 \in I$, erhalten wir $n_1 = n_2$ und insgesamt $\varphi_1 = \varphi_2$.

zu (ii) Weil die Abbildungen $(r, \varphi) \mapsto r$ und $(r, \varphi) \mapsto \varphi$ nach (3.4) stetig sind, gilt dasselbe nach (3.8) auch für $(r, \varphi) \mapsto r\cos(\varphi)$ und $(r, \varphi) \mapsto r\sin(\varphi)$. Eine erneute Anwendung von (3.6) liefert dann die Stetigkeit der Abbildung ρ_{pol}. Zum Nachweis der Injektivitätsaussage sei I wieder ein Intervall der angegebenen Form, und seien (r, φ) und (r', φ') in $\mathbb{R}^+ \times I$ mit $\rho_{\text{pol}}(r, \varphi) = \rho_{\text{pol}}(r', \varphi')$ vorgegeben. Dann gilt $r\cos(\varphi) = r'\cos(\varphi')$ und $r\sin(\varphi) = r'\sin(\varphi')$. Zunächst erhalten wir

$$r^2 \quad = \quad (r\cos(\varphi))^2 + (r\sin(\varphi)^2)^2 \quad = \quad (r'\cos(\varphi'))^2 + (r'\sin(\varphi')^2)^2 \quad = \quad (r')^2$$

und somit $r = r'$. Daraus folgt $\cos(\varphi) = \cos(\varphi')$ und $\sin(\varphi) = \sin(\varphi')$, also $\phi(\varphi) = \phi(\varphi')$ für die Abbildung ϕ aus Teil (i). Auf Grund der dort bewiesenen Injektivität folgt $\varphi = \varphi'$, insgesamt also $(r, \varphi) = (r', \varphi')$ wie gewünscht. □

Es ist nicht schwer zu sehen, dass die Bildmenge von $\phi|_I$ jeweils der Einheitskreis ist, und dass die Bildmenge von $\rho_{\text{pol}}|_{\mathbb{R}^+ \times I}$ durch $\mathbb{R}^2 \setminus \{0_{\mathbb{R}^2}\}$ gegeben ist. Wir verzichten an dieser Stelle aber auf einen Beweis.

Häufig bezeichnet man in der Physik eine Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ als Funktion in ***kartesischen Koordinaten***, während die Funktion f_{pol} auf $\mathbb{R}^+ \times \mathbb{R}$ gegeben durch $f_{\text{pol}} = f \circ \rho_{\text{pol}}$ die entsprechende Funktion „in Polarkoordinaten“ genannt wird. Der Übergang zu Polarkoordinaten ermöglicht besonders in Situationen eine Vereinfachung von Berechnungen, in denen die Funktion symmetrisch bezüglich des Koordinatenursprungs $(0, 0)$ ist, der Funktionswert $f(x, y)$ also nur vom Abstand $r = \|(x, y)\|_2$ des Punktes (x, y) vom Ursprung abhängt. Beispielsweise hat die Funktion $f(x, y) = (x^2 + y^2)^2$ in Polarkoordinaten die einfache Form

$$\begin{aligned} f_{\text{pol}}(r, \varphi) \quad &= \quad (f \circ \rho_{\text{pol}})(r, \varphi) \quad = \quad f(r\cos(\varphi), r\sin(\varphi)) \quad = \quad ((r\cos(\varphi))^2 + (r\sin(\varphi))^2)^2 \\ &= \quad (r^2(\cos(\varphi)^2 + \sin(\varphi)^2))^2 \quad = \quad (r^2)^2 \quad = \quad r^4. \end{aligned}$$

In diesem Zusammenhang ist es wichtig, auch partielle Ableitungen und diverse Differenzialoperatoren, die wir später einführen werden (Gradient, Divergenz, Rotation), in Polarkoordinaten ausdrücken zu können. Wir kommen zu einem geeigneten Zeitpunkt auf dieses Thema zurück.

Wir werden später auch zeigen: Ist I ein *offenes* Intervall mit Länge $\leq 2\pi$, dann sind $\phi|_I$ und $\rho_{\text{pol}}|_{\mathbb{R}^+ \times I}$ sogar Homömorphismen auf ihre Bildmenge. Im Kapitel über die implizite Funktionen und lokale Umkehrbarkeit werden wir sehen, dass man dies nachweisen kann, ohne die Umkehrabbildung vorher explizit auszrechnen.

Ist I jedoch halboffen mit Länge 2π, zum Beispiel $I = [0, 2\pi[$, dann ist $\phi|_I$ *kein* Homöomorphismus, denn die Umkehrabbildung $(\phi|_I)^{-1}$ ist im Punkt $(1,0)$ unstetig. Eine bijektive stetige Abbildung ist also im Allgemeinen kein Homömorphismus! Um das in dieser konkreten Situation zu sehen, betrachten wir die Folge $(\varphi^{(n)})_{n\in\mathbb{N}}$ in I gegeben durch $\varphi^{(n)} = 2\pi - \frac{1}{n}$ für alle $n \in \mathbb{N}$ und definieren $(x^{(n)}, y^{(n)}) = \phi(\varphi^{(n)})$ für $n \in \mathbb{N}$. Auf Grund der Stetigkeit von ϕ gilt $\lim_n (x^{(n)}, y^{(n)}) = \phi(2\pi) = (\cos(2\pi), \sin(2\pi)) = (1,0)$. Nun ist $(\phi|_I)^{-1}(1,0) = 0$, denn 0 ist der eindeutig bestimmte Punkt in I mit $(\phi|_I)(0) = (\cos(0), \sin(0)) = (1,0)$. Wäre auch $(\phi|_I)^{-1}$ stetig, dann müsste also $\lim_n \phi^{-1}(x^{(n)}, y^{(n)}) = (\phi|_I)^{-1}(1,0) = 0$ gelten. Tatsächlich gilt wegen $\varphi^{(n)} \in I$ und $\phi(\varphi^{(n)}) = (x^{(n)}, y^{(n)})$ für alle $n \in \mathbb{N}$ aber

$$\lim_{n\to\infty} \phi^{-1}(x^{(n)}, y^{(n)}) \quad = \quad \lim_{n\to\infty} \varphi^{(n)} \quad = \quad \lim_{n\to\infty} (2\pi - \tfrac{1}{n}) \quad = \quad 2\pi \quad \neq \quad 0.$$

Im $\mathbb{R}^3$ ist vor allem die Verwendung der folgenden Koordinatenabbildungen üblich.

(3.15) Satz. *(Definition der Zylinder- und Kugelkoordinaten)*
Sei I ein halboffenes Intervall der Länge 2π.

(i) Die Abbildung $\rho_{\text{zyl}} : \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^3$, $(r, \varphi, h) \mapsto (r\cos(\varphi), r\sin(\varphi), h)$ ist stetig, und die Einschränkung $\rho_{\text{zyl}}|_{M_I}$ auf $M_I = \mathbb{R}^+ \times I \times \mathbb{R}$ ist eine stetige Bijektion auf ihre Bildmenge.

(ii) Die Abbildung $\rho_{\text{kug}} : \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^3$ gegeben durch

$$(r, \vartheta, \varphi) \quad \mapsto \quad (r\sin(\vartheta)\cos(\varphi), r\sin(\vartheta)\sin(\varphi), r\cos(\vartheta))$$

ist ebenfalls stetig, und die Einschränkung $\rho_{\text{kug}}|_{N_I}$ auf $N_I = \mathbb{R}^+ \times]0, \pi[\times I$ ist eine stetige Bijektion auf ihre Bildmenge.

Die Abbildung ρ_{zyl} wird ***Zylinderkoordinaten-Abbildung***, die Abbildung ρ_{kug} ***Kugelkoordinaten-Abbildung*** genannt.

Betrachtet man für ein festes $r \in \mathbb{R}^+$ die Sphäre vom Radius r als Globus, dann lässt sich der Winkel φ, der auch ***Azimutwinkel*** genannt wird, als Längengrad interpretieren, sofern er im Bereich $[-\pi, \pi]$ gewählt wird. Es enspricht dann $\varphi = 0 = 0°$ dem Meridian, während der 180-te Grad „westlicher Länge“ ($\varphi = -\pi$) mit dem 180-ten Grad „östlicher Länge“ ($\varphi = \pi$) in der Nähe der Datumsgrenze zusammenfällt. Für den sog. ***Polarwinkel*** ϑ kann entsprechend $\vartheta + \frac{1}{2}\pi$ als „Breitengrad“ angesehen werden, sofern ϑ im Intervall $[-\pi, 0]$ und $\vartheta + \frac{1}{2}\pi$ somit im Intervall $[-\frac{1}{2}\pi, \frac{1}{2}\pi]$ liegt. Für $\vartheta = 0 \Leftrightarrow \vartheta + \frac{1}{2}\pi = \frac{1}{2}\pi = 90°$ „nördlicher Breite“ erhält man wegen $\sin(0) = 0$ und $\cos(0) = 1$ den Nordpol $(0, 0, r)$, für $\vartheta = -\pi \Leftrightarrow \vartheta + \frac{1}{2}\pi = -\frac{1}{2}\pi = -90°$, also 90° „südlicher Breite“ wegen $\sin(-\pi) = 0$ und $\cos(-\pi) = -1$ den Südpol $(0, 0, -r)$ der Erdkugel. Der Wert $\vartheta = -\frac{1}{2}\pi \Leftrightarrow \vartheta + \frac{1}{2}\pi = 0$ entspricht natürlich dem Äquator.

Beweis von Satz (3.15):

Der Beweis der Stetigkeit kann genau wie in Satz (3.14) geführt werden. Wir können uns also ganz auf den Nachweis der Injektivität konzentrieren.

zu (i) Seien $(r, \varphi, h), (r', \varphi', h')$ in $\mathbb{R}^+ \times I \times \mathbb{R}$ mit $\rho_{\text{zyl}}(r, \varphi, h) = \rho_{\text{zyl}}(r', \varphi', h')$ vorgegeben. Dann gilt $r\cos(\varphi) = x = r'\cos(\varphi')$, $r\sin(\varphi) = y = r'\sin(\varphi')$ und $h = h'$. Aus den ersten beiden Gleichungen folgt $\rho_{\text{pol}}(r, \varphi) = \rho_{\text{pol}}(r', \varphi')$ mit

der Polarkoordinaten-Abbildung. Weil die Einschränkung dieser Abbildung auf $\mathbb{R}^+ \times I$ nach Satz (3.14) injektiv ist, folgt $(r,\varphi) = (r',\varphi')$. Insgesamt ist damit $(r,\varphi,h) = (r',\varphi',h')$ nachgewiesen.

zu (ii) Seien $(r,\vartheta,\varphi),(r',\vartheta',\varphi') \in \mathbb{R}^+ \times]0,\pi[\times I$ mit $\rho_{\text{kug}}(r,\vartheta,\varphi) = (x,y,z) = \rho_{\text{kug}}(r',\vartheta',\varphi')$ vorgegeben. Dann gilt

$$x = r\sin(\vartheta)\cos(\varphi) = r'\sin(\vartheta')\cos(\varphi') \quad , \quad y = r\sin(\vartheta)\sin(\varphi) = r'\sin(\vartheta')\sin(\varphi')$$
$$\text{und} \quad z = r\cos(\vartheta) = r'\cos(\vartheta').$$

Zunächst erhalten wir die Gleichung $r = r'$, weil

$$x^2+y^2+z^2 \quad = \quad r^2\sin(\vartheta)^2(\cos(\varphi)^2+\sin(\varphi)^2) + r^2\cos(\vartheta)^2 \quad = \quad r^2\sin(\vartheta)^2\cdot 1 + r^2\cos(\vartheta)^2 \quad = \quad r^2$$

gilt und man durch eine ebensolche Rechnung auch $x^2+y^2+z^2 = (r')^2$ erhält. Mit Hilfe der Gleichung für die z-Koordinate folgt $\cos(\vartheta) = \cos(\vartheta')$. Weil die Kosinusfunktion auf dem Intervall $]0,\pi[$ streng monoton fallend, also insbesondere injektiv ist, folgt $\vartheta = \vartheta'$. Weil die Sinusfunktion auf dem Intervall $]0,\pi[$ nicht Null wird, dürfen wir die Gleichung für die x-Koordinate durch $r\sin(\vartheta) = r'\sin(\vartheta')$ dividieren und erhalten $\cos(\varphi) = \cos(\varphi')$. Die Gleichung für die y-Koordinate liefert ebenso $\sin(\varphi) = \sin(\varphi')$. Somit gilt $\phi(\varphi) = \phi(\varphi')$, wobei ϕ die Abbildung aus Satz (3.14) bezeichnet. Weil $\phi|_I$ laut diesem Satz injektiv ist, erhalten wir $\varphi = \varphi'$. Insgesamt ist damit $(r,\vartheta,\varphi) = (r',\vartheta',\varphi')$ nachgewiesen.

Im letzten Teil dieses Kapitels untersuchen wir die Stetigkeit von linearen Abbildungen.

(3.16) Satz. Eine lineare Abbildung $\phi : V \to W$ zwischen normierten $\mathbb{R}$-Vektorräumen ist genau dann stetig, wenn eine Konstante $\gamma \in \mathbb{R}^+$ mit $\|\phi(v)\| \le \gamma\|v\|$ für alle $v \in V$ existiert.

Beweis: „⇒“ Auf Grund der Stetigkeit von ϕ im Punkt 0_V und auf Grund des ε-δ-Kriteriums gibt es für $\varepsilon = 1$ ein $\delta \in \mathbb{R}^+$ mit $\|v\| < \delta \Rightarrow \|\phi(v)\| < 1$ für alle $v \in V$. Sei nun $\gamma = 2\delta^{-1}$ und $v \in V$ mit $v \neq 0_V$ vorgegeben. Dann ist $v' = \frac{1}{\gamma\|v\|}v$ ein Vektor mit $\|v'\| = \frac{1}{2}\delta < \delta$. Es folgt $\|\phi(v')\| < 1$, und wegen $v = \gamma\|v\|v'$ und der Linearität von ϕ erhalten wir die Ungleichung $\|\phi(v)\| = \|\phi(\gamma\|v\|v')\| = \gamma\|v\|\|\phi(v')\| < \gamma\|v\|$.

„⇐“ Sei $\gamma \in \mathbb{R}^+$ eine Konstante wie angegeben und $a \in V$ beliebig. Für alle $v \in V$ mit dann

$$\|\phi(v)-\phi(a)\| \quad \le \quad \|\phi(v-a)\| \quad \le \quad \gamma\|v-a\|.$$

Ist nun $(v^{(n)})_{n\in\mathbb{N}}$ eine Folge mit $\lim_n v^{(n)} = a$, dann gilt $\lim_n \|v^{(n)}-a\| = 0$. Aus den Ungleichungen $0 \le \|\phi(v^{(n)})-\phi(a)\| \le \gamma\|v^{(n)}-a\|$ für $n \in \mathbb{N}$ folgt $\lim_n \|\phi(v^{(n)})-\phi(a)\| = 0$ und somit $\lim_n \phi(v^{(n)}) = \phi(a)$. □

Das folgende Beispiel zeigt, dass nicht jede lineare Abbildung stetig ist.

(3.17) Proposition. Sei V der $\mathbb{R}$-Vektorraum der Polynomfunktionen auf dem Intervall $[0,1]$, ausgestattet mit der Supremumsnorm $\|f\|_\infty = \sup\{\,|f(x)| \mid x \in [0,1]\}$. Dann ist die Ableitungsabbildung $\phi_1 : V \to V$, $f \mapsto f'$ unstetig.

Beweis: Angenommen, es ist $\gamma \in \mathbb{R}^+$ eine Konstante mit der Eigenschaft $\|\phi_1(f)\|_\infty \le \gamma\|f\|_\infty$ für alle $f \in V$. Dann betrachten wir die Folge $(f_n)_{n\in\mathbb{N}}$ gegeben durch $f_n(x) = x^n$ für alle $n \in \mathbb{N}$. Der betragsmäßig größte Funktionswert,

der von f_n im Intervall $[0,1]$ angenommen wird, ist $|f_n(1)| = 1$, also gilt $\|f_n\|_\infty = 1$ für alle $n \in \mathbb{N}$. Die Bilder der Folgenglieder sind gegeben durch $\phi_1(f_n) = f_n'$ mit $f_n'(x) = nx^{n-1}$. Es gilt $\|\phi_1(f_n)\|_\infty = \|f_n'\|_\infty = |f_n'(1)| = n$ für alle $n \in \mathbb{N}$. Setzen wir nun die Funktionen f_n in die Abschätzung von oben ein, dann erhalten wir $\|\phi_1(f_n)\|_\infty \leq \gamma \|f_n\|_\infty$, also $n \leq \gamma$ für alle $n \in \mathbb{N}$. Dies aber widerspricht der Unbeschränktheit der Menge $\mathbb{N}$ der natürlichen Zahlen. Also existiert keine Konstante γ mit der angegebenen Eigenschaft. □

Seien V, W zwei normierte $\mathbb{R}$-Vektorräume. Dann bezeichnen wir mit $\mathscr{L}(V,W)$ den Untervektorraum vom $\mathbb{R}$-Vektorraum $\mathrm{Hom}_{\mathbb{R}}(V,W)$ bestehend aus den *stetigen* linearen Abbildungen $V \to W$.

(3.18) Satz. Für jedes $\phi \in \mathscr{L}(V,W)$ existiert das Supremum

$$\|\phi\| \quad = \quad \sup\{\, \|\phi(v)\| \mid v \in V \,,\, \|v\| \leq 1\}.$$

Durch die Zuordnung $\mathscr{L}(V,W) \to \mathbb{R}_+$, $\phi \mapsto \|\phi\|$ ist auf dem $\mathbb{R}$-Vektorraum $\mathscr{L}(V,W)$ eine Norm definiert, die sogenannte ***Operatornorm***.

Beweis: Das Supremum existiert, weil die angegebene Menge nichtleer und auf Grund des vorherigen Satzes nach oben beschränkt ist. Für die Nullabbildung gilt offenbar $\|0_{L(V,W)}\| = 0$. Sei umgekehrt $\phi \in \mathscr{L}(V,W)$ mit $\|\phi\| = 0$. Für jeden Vektor $0_V \neq v \in V$ gilt dann

$$\|\phi(v)\| \quad = \quad \|v\| \left\| \phi\left(\frac{v}{\|v\|}\right) \right\| \quad \leq \quad \|v\| \cdot 0 \quad = \quad 0$$

und somit $\phi(v) = 0$, d.h. ϕ ist die Nullabbildung. Die Gleichung $\|\lambda\phi\| = |\lambda| \|\phi\|$ für alle $\lambda \in \mathbb{R}$ und $\phi \in \mathscr{L}(V,W)$ ist offensichtlich, denn der Übergang von ϕ zu $\lambda\phi$ bewirkt, dass die Zahlen in der Menge $\{\|\phi(v)\| \mid v \in V \mid \|v\| \leq 1\}$ mit dem Wert $|\lambda|$ multipliziert werden. Seien nun $\phi, \psi \in \mathscr{L}(V,W)$ vorgegeben und $v \in V$ mit $\|v\| \leq 1$. Dann gilt

$$\|\phi(v) + \psi(v)\| \quad \leq \quad \|\phi(v)\| + \|\psi(v)\| \quad \leq \quad \|\phi\| + \|\psi\| \quad ,$$

also $\|\phi + \psi\| \leq \|\phi\| + \|\psi\|$ nach Definition der Operatornorm. □

Wir bemerken noch, dass die Operatornorm so definiert ist, dass $\|\phi(v)\| \leq \|\phi\| \|v\|$ für alle $\phi \in \mathscr{L}(V,W)$ und $v \in V$ erfüllt ist.

(3.19) Proposition. Seien $V = \mathbb{R}^n$, $W = \mathbb{R}^m$ jeweils mit der Supremums-Norm $\|\cdot\|_\infty$ ausgestattet und $A \in \mathscr{M}_{m\times n,\mathbb{R}}$ eine Matrix. Dann ist die Operatornorm von $\phi_A : V \to W$, $v \mapsto Av$ gegeben durch

$$\|\phi_A\| \quad = \quad \max\left\{ \sum_{j=1}^{n} |a_{ij}| \;\middle|\; 1 \leq i \leq m \right\} \tag{3}$$

Man bezeichnet diese Norm auch als ***Zeilensummennorm***.

Beweis: Wir zeigen, dass der Wert $\|\phi_A(v)\|$ für alle $v \in V$ mit $\|v\|_\infty \leq 1$ durch die Zahl γ_A auf der rechten Seite von (3) beschränkt ist, und dass es Vektoren $v \in V$ mit $\|v\|_\infty = 1$ gibt, für die $\|\phi_A(v)\|_\infty = \gamma_A$ erfüllt ist. Beide Aussagen zusammen beweisen die Gleichung $\|\phi_A\| = \gamma_A$.

Sei also $v \in V$ mit $\|v\|_\infty \leq 1$ vorgegeben. Dann sind die Komponenten von $w = \phi_A(v)$ durch $w_i = \sum_{j=1}^n a_{ij} v_j$ gegeben und können durch

$$\left| \sum_{j=1}^n a_{ij} v_j \right| \quad \leq \quad \sum_{j=1}^n |a_{ij}||v_j| \quad \leq \quad \sum_{j=1}^n |a_{ij}| \quad ,$$

also insbesondere durch das Maximum dieser Zahlen abgeschätzt werden. Andererseits wird das Maximum auch durch Vektoren $v \in V$ mit $\|v\|_\infty = 1$ angenommen: Ist i_0 der Index der Zeile mit der maximalen Betragssumme, also

$$\max \left\{ \sum_{j=1}^n |a_{ij}| \;\middle|\; 1 \leq i \leq m \right\} \quad = \quad \sum_{j=1}^n |a_{i_0 j}| \quad ,$$

dann definieren wir $v = (v_1, ..., v_n)$ durch $v_j = |a_{i_0 j}|/a_{i_0 j}$ falls $a_{i_0 j} \neq 0$ und $v_j = 1$ sonst, für $1 \leq j \leq n$. Die Gleichung $\|v\|_\infty = 1$ ist dann offensichtlich, denn die Einträge des Vektors sind alle gleich 1 oder -1. In der Summe $\sum_{j=1}^n a_{i_0 j} v_j$ sind alle Summanden nicht-negativ, und es gilt $\sum_{j=1}^n a_{i_0 j} v_j = \sum_{j=1}^n |a_{i_0 j}| = \|\phi_A\|$. □

(3.20) Folgerung. Jede lineare Abbildung von einem endlich-dimensionalen in einen normierten $\mathbb{R}$-Vektorraum beliebiger Dimension ist stetig. Jeder Isomorphismus zwischen $\mathbb{R}$-Vektorräumen derselben endlichen Dimension ist ein Homöomorphismus.

Beweis: Sei V endlich-dimensional, W beliebig und $(v_1, ..., v_n)$ eine Basis von V. Da sich an der Stetigkeit einer Abbildung bei Übergang zu einer äquivalenten Norm nichts ändert, können wir annehmen, dass die Norm auf V durch

$$\left\| \sum_{i=1}^n \lambda_i v_i \right\| \quad = \quad \max\{|\lambda_1|, ..., |\lambda_n|\}$$

gegeben ist. Sei nun $v \in V$ beliebig, $v = \sum_{i=1}^n \lambda_i v_i$. Setzen wir $\gamma = \max\{\|\phi(v_1)\|, ..., \|\phi(v_n)\|\}$, dann gilt

$$\|\phi(v)\| \quad = \quad \left\| \sum_{i=1}^n \lambda_i \phi(v_i) \right\| \quad \leq \quad \sum_{i=1}^n |\lambda_i| \|\phi(v_i)\| \quad \leq \quad n\gamma \|v\|.$$

Die zweite Aussage folgt unmittelbar aus der ersten. □

Das folgende Resultat werden wir benötigen, um später die Stetigkeit von Ableitungsfunktionen zu untersuchen. Denn wie wir im Kapitel über totale Differenzierbarkeit sehen werden, nimmt die (totale) Ableitung der Funktion im Mehrdimensionalen ihre Werte nicht in $\mathbb{R}$, sondern in einem Vektorraum an, der seinerseits aus linearen Abbildungen besteht.

(3.21) Satz. Sei X ein metrischer Raum, und seien V, W zwei endlich-dimensionale, normierte $\mathbb{R}$-Vektorräume. Eine Abbildung $\varphi : X \to \mathcal{L}(V, W)$ ist genau dann stetig, wenn die Zuordnung $\varphi_v : X \to W$, $x \mapsto \varphi(x)(v)$ für jeden Vektor $v \in V$ stetig ist.

Beweis: „⇒“ Sei $x \in X$ und $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge in X mit $\lim_n x^{(n)} = x$. Für jeden Vektor $v \in V$ ist $\lim_n \varphi(x^{(n)})(v) = \varphi(x)(v)$ nachzuweisen. Für $v = 0_V$ ist dies offensichtlich, weil dann $\varphi(x)(v) = 0_W$ und $\varphi(x^{(n)})(v) = 0_W$ für alle

$n \in \mathbb{N}$ erfüllt ist. Also können wir von nun an $v \neq 0_V$ voraussetzen. Auf Grund der Stetigkeit von φ finden wir für jedes ε ein $N \in \mathbb{N}$ mit $\|\varphi(x^{(n)}) - \varphi(x)\| < \varepsilon/\|v\|$ für alle $n \geq N$. Nach Definition der Operatornorm folgt daraus

$$\|\varphi(x^{(n)})(v) - \varphi(x)(v)\| \quad = \quad \|(\varphi(x^{(n)}) - \varphi(x))(v)\| \quad \leq \quad \|\varphi(x_n) - \varphi(x)\|\|v\| \quad < \quad \frac{\varepsilon}{\|v\|}\|v\| \quad = \quad \varepsilon$$

für alle $n \geq N$. Also ist die Gleichung $\lim_n \varphi(x^{(n)})(v) = \varphi(x)(v)$ erfüllt.

„$\Leftarrow$“ Sei $\mathscr{B} = (v_1, ..., v_d)$ eine Basis von V. Der Raum $\mathscr{L}(V, W)$ ist endlich-dimensional, deshalb sind je zwei Normen darauf äquivalent. Wir können also die Norm auf V und damit auch die Operatornorm auf $\mathscr{L}(V, W)$ abändern, ohne dass sich dadurch an der Stetigkeit von φ etwas ändert. Deshalb können wir davon ausgehen, dass die Norm auf V durch

$$\left\| \sum_{i=1}^{d} \lambda_i v_i \right\| \quad = \quad \max\{|\lambda_1|, ..., |\lambda_d|\}$$

definiert ist. Sei nun $x \in X$ und $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge in X mit $\lim_n x^{(n)} = x$. Zu zeigen ist, dass die Folge $\psi_n = \varphi(x^{(n)})$ bezüglich der Operatornorm gegen das Element $\psi = \varphi(x) \in \mathscr{L}(V, W)$ konvergiert. Weil nach Voraussetzung die Zuordnungen $x \mapsto \varphi(x)(v_i)$ für $1 \leq i \leq d$ stetig sind, existiert für jedes $\varepsilon \in \mathbb{R}^+$ ein $N \in \mathbb{N}$ mit der Eigenschaft, dass $\|\psi_n(v_i) - \psi(v_i)\| \leq \frac{\varepsilon}{d}$ für alle $n \geq N$ und $1 \leq i \leq d$ erfüllt ist. Sei nun $v \in V$ beliebig vorgegeben, $v = \sum_{i=1}^{d} \lambda_i v_i$. Dann gilt

$$\|\psi_n(v) - \psi(v)\| \quad \leq \quad \sum_{i=1}^{d} |\lambda_i| \|\psi_n(v_i) - \psi(v_i)\| \quad < \quad \sum_{i=1}^{d} |\lambda_i| \frac{\varepsilon}{d} \quad \leq \quad \sum_{i=1}^{d} \|v\| \frac{\varepsilon}{d} \quad = \quad \varepsilon\|v\|.$$

Insbesondere gilt $\|(\psi_n - \psi)(v)\| \leq \varepsilon$ für alle $v \in V$ mit $\|v\| \leq 1$ und alle $n \geq N$. Nach Definition der Operatornorm gilt also $\|\psi_n - \psi\| \leq \varepsilon$ für alle $n \geq N$. □

§ 4. *Offene und abgeschlossene Teilmengen*

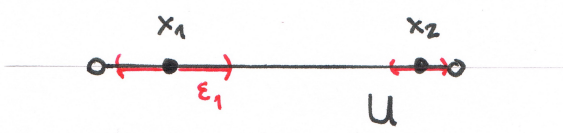
Zusammenfassung. Eine Teilmenge $U \subseteq X$ eines metrischen Raums (X, d_X) wird *offen* genannt, wenn um eine Punkt von U ein (eventuell sehr kleiner) ε-Ball gelegt werden kann, der immer noch vollständig vollständig in U liegt. Komplemente offener Teilmengen werden *abgeschlossen* genannt. Beispielsweise sind offene Intervalle offene Teilmengen von $\mathbb{R}$ (mit der Standard-Metrik), und abgeschlossene Intervalle sind abgeschlossene Teilmengen.

Mit dem Begriff der Offenheit können Abbildungen auf Stetigkeit getestet werden; umgekehrt kann mit Hilfe stetiger Abbildungen die Offenheit einer Menge nachgewiesen werden. Auch bei der Definition der Differenzierbarkeit werden wir den Begriff der offenen Teilmenge benötigen. Schränkt man die Metrik X eine metrischen Raums auf eine Teilmenge Y ein, dann kann auch Y als metrischer Raum betrachtet werden. Dieser besitzt dann seine „eigenen" offenen Teilmengen. Ist beispielsweise $X = \mathbb{R}^3$ und Y eine Ebene in X, dann kann man auch im metrischen Raum Y offene Teilmenge betrachten. Ist beispielsweise $p \in Y$ ein beliebiger Punkt, dann ist $Y \setminus \{p\}$ in Y offen. Aber weder $Y \setminus \{p\}$ noch Y selbst sind offen im metrischen Raum X. Statt dessen spricht man von *in Y relativ offenen* Mengen.

Wichtige Begriffe und Sätze in diesem Kapitel

- Definition der Offenheit, Umgebungsbegriff
- Eine Abbildung ist genau dann stetig, wenn das Urbild jeder offenen Teilmenge offen ist. Die Stetigkeit in einem Punkt kann entsprechend durch Umgebungen charakterisiert werden.
- Abgeschlossenheit
- Schachtelungsprinzip
- Grenzwerte von Folgen in abgeschlossenen Mengen
- relative Offenheit und Abgeschlossenheit

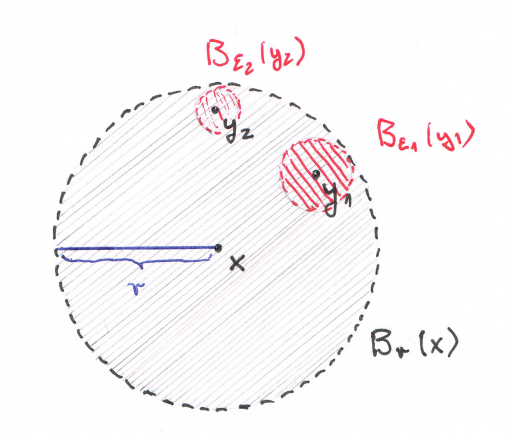
(4.1) Definition. Sei (X, d) ein metrischer Raum. Eine Teilmenge $U \subseteq X$ wird ***offen*** genannt, wenn für jedes $x \in U$ ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq U$ existiert.



Offene Intervalle sind offen. Gleichgültig wo die Punkte $x_1, x_2, \ldots$ innerhalb des Intervalls gewählt werden, es gibt immer eine ε-Umgebung, die ganz innerhalb des Intervalls liegt.

(4.2) Proposition. Jedes offene Intervall $]a, b[\subseteq \mathbb{R}$ ist eine offene Teilmenge bezüglich der Standard-Metrik $d(x, y) = |x - y|$.

Beweis: Ist $x \in]a, b[$ vorgegeben, dann setzen wir $\varepsilon = \min\{x - a, b - x\}$. Es gilt dann $B_\varepsilon(x) \subseteq]a, b[$, denn für jedes $y \in B_\varepsilon(x)$ gilt $|y - x| < \varepsilon \Leftrightarrow x - \varepsilon < y < x + \varepsilon$, also insbesondere $a = x - (x - a) < y < x + (b - x) = b$. □



Zur Offenheit der offenen Bälle in einem metrischen Raum. Egal wo die Punkte $y_1, y_2, \ldots$ innerhalb von $B_r(x)$ gewählt werden, eine hinreichend kleine offene Kreisscheibe um y_n ist vollständig in $B_r(x)$ enthalten.

(4.3) Proposition. Ist (X, d) ein metrischer Raum, dann ist jeder offene Ball $B_r(x)$ eine offene Teilmenge von X.

Beweis: Sei $y \in B_r(x)$ beliebig und $\varepsilon = r - d(y,x)$. Dann gilt $B_\varepsilon(y) \subseteq B_r(x)$, denn für alle $u \in X$ gilt

$$u \in B_\varepsilon(y) \quad \Rightarrow \quad d(u,y) < \varepsilon \quad \Rightarrow \quad d(u,x) \leq d(u,y) + d(y,x) < \varepsilon + d(y,x) = r.$$

und somit $u \in B_r(x)$. □

Ist d die *diskrete* Metrik auf einer Menge X, dann ist jede Teilmenge $U \subseteq X$ offen, denn für jedes $x \in X$ ist der offene Ball $B_1(x) = \{x\}$ in U enthalten.

(4.4) Proposition. Sei $(V, \|\cdot\|)$ ein normierter $\mathbb{R}$-Vektorraum und $\|\cdot\|'$ eine zu $\|\cdot\|$ äquivalente Norm. Eine Teilmenge $U \subseteq V$ ist genau dann offen bezüglich der Norm $\|\cdot\|$, wenn sie bezüglich der Norm $\|\cdot\|'$ offen ist.

Beweis: Sei U eine bezüglich $\|\cdot\|$ offene Menge und $x \in U$. Weil U offen ist, gibt es ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq U$. Auf Grund der Äquivalenz der Normen gibt es eine Konstante $\gamma \in \mathbb{R}^+$ mit $\|v\| \leq \gamma \|v\|'$ für alle $v \in V$. Bezeichnen wir nun mit $B' = B'_{\gamma^{-1}\varepsilon}(x)$ den offenen Ball um x vom Radius $\gamma^{-1}\varepsilon$ bezüglich $\|\cdot\|'$. Es gilt nun $B' \subseteq B_\varepsilon(x) \subseteq U$, denn für jedes $y \in B'$ gilt $\|y - x\|' < \gamma^{-1}\varepsilon$, damit $\|y - x\| \leq \gamma \|y - x\|' < \gamma\gamma^{-1}\varepsilon = \varepsilon$ und somit $y \in B_\varepsilon(x)$. □

Betrachtet man einen endlich-dimensionalen $\mathbb{R}$-Vektorraum V, ohne dass auf V explizit eine bestimmte Metrik angegeben wurde, dann ist mit der *Offenheit* einer Teilmenge $U \subseteq V$ immer die Offenheit bezüglich der von einer beliebigen Norm induzierten Metrik gemeint. Wegen (4.4) spielt die Wahl der Norm dabei keine Rolle.

Sei X eine Menge und $\mathfrak{P}(X)$ seine Potenzmenge, also die Menge aller Teilmengen von X. In den folgenden Abschnitt werden wir des öfteren mit *Familien* von Teilmengen arbeiten. Ein *Familie* von Teilmengen der Menge X über einer Indexmenge I ist einfach eine Abbildung $\phi : I \to \mathfrak{P}(X)$. Jedem Element aus I wird also eine Teilmenge von X zugeordnet.

In der Regel verwendet man an Stelle der Abbildung- die Indexschreibweise, wobei dann $(X_i)_{i\in I}$ für die Abbildung $I \to \mathfrak{P}(X)$, $i \mapsto X_i$ steht. Beispiele für Familien von Teilmengen der Menge $X = \mathbb{R}$ sind $(\,]n, n+1[\,)_{n\in\mathbb{N}}$ oder $(\{a, a+\frac{1}{3}\})_{a\in\mathbb{Z}}$. Elemente der zweiten Familie sind zum Beispiel $\{0, \frac{1}{3}\}$ oder $\{-2, -\frac{5}{3}\}$. Ein Element der ersten Familie ist das offene Intervall $]7, 8[$.

(4.5) Satz. Sei (X, d) ein metrischer Raum. Dann gilt

(i) Die Teilmengen $\varnothing$ und X sind offen.

(ii) Sind U und V offen, dann auch die Schnittmenge $U \cap V$.

(iii) Ist $(U_i)_{i\in I}$ eine Familie offener Teilmengen, dann ist auch die Vereinigung $\bigcup_{i\in I} U_i$ offen.

Beweis: zu (i) Die Menge $Y = \varnothing$ ist offen, da in diesem Fall kein Punkt $x \in Y$ existiert, für den die Existenz eines $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq Y$ gefordert wird. Die Menge X ist offen, weil in diesem Fall für jedes $x \in X$ und jedes $\varepsilon \in \mathbb{R}^+$ die Bedingung $B_\varepsilon(x) \subseteq X$ offensichtlich erfüllt ist; denn nach Definition ist $B_\varepsilon(x)$ eine Teilmenge von X.

zu (ii) Sei $x \in U \cap V$. Weil U offen ist, gibt es ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq U$. Weil V offen ist, existiert ein $\eta \in \mathbb{R}^+$ mit $B_\eta(x) \subseteq V$. Setzen wir $\xi = \min(\varepsilon, \eta)$, dann folgt $B_\xi(x) \subseteq B_\varepsilon(x) \subseteq U$ und $B_\xi(x) \subseteq B_\eta(x) \subseteq V$, insgesamt also $B_\xi(x) \subseteq U \cap V$.

zu (iii) Sei $U = \bigcup_{i\in I} U_i$ und $x \in U$ vorgegeben. Dann gibt es ein $i \in I$ mit $x \in U_i$. Weil U_i offen ist, existiert ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq U_i$. Wegen $U_i \subseteq U$ folgt $B_\varepsilon(x) \subseteq U$. □

Aus Teil (ii) von (4.5) kann mit vollständiger Induktion gefolgert werden, dass der Durchschnitt $U_1 \cap \ldots \cap U_r$ von endlich vielen offenen Mengen offen ist. Für unendliche Durchschnitte gilt dies aber nicht mehr. Betrachtet man im metrischen Raum $X = \mathbb{R}$ beispielsweise das System $(U_n)_{n\in\mathbb{N}}$ gegeben durch $U_n = \left]-\frac{1}{n}, \frac{1}{n}\right[$, so ist jedes einzelne U_n eine offene Teilmenge von $\mathbb{R}$, ihr Durchschnitt $\bigcap_{n\in\mathbb{N}} U_n = \{0\}$ aber nicht.

Sei X eine Menge. Ein Mengensystem $\mathscr{U} \subseteq \mathfrak{P}(X)$ mit den Eigenschaften $\varnothing, X \in \mathscr{U}$ und $U, V \in \mathscr{U} \Rightarrow U \cap V \in \mathscr{U}$ und $\bigcup_{i\in I} U_i \in \mathscr{U}$ für jede Familie $(U_i)_{i\in I}$ bestehend aus Menge $U_i \in \mathscr{U}$ wird ***Topologie*** auf X genannt. Der soeben bewiesene Satz zeigt also, dass die offenen Mengen in einem metrischen Raum X eine Topologie auf X bilden. Es gibt aber auch Topologien, die sich nicht durch eine Metrik definieren lassen.

(4.6) Definition. Sei (X, d) ein metrischer Raum und $x \in X$. Eine Teilmenge $U \subseteq X$ wird ***Umgebung*** von x genannt, wenn ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq U$ existiert.

Der Begriff der Umgebung lässt sich auf die Definition der Offenheit zurückführen.

(4.7) Proposition. Seien die Bezeichnungen wie in der Umgebungsdefinition gewählt. Eine Teilmenge U von X ist genau dann eine Umgebung von $x \in X$, wenn eine offene Menge V mit $V \subseteq U$ und $x \in V$ existiert.

Beweis: „⇒“ Nach Voraussetzung gilt $B_\varepsilon(x) \subseteq U$ für ein $\varepsilon \in \mathbb{R}^+$. Die Menge $B_\varepsilon(x)$ ist offen, wir können also $V = B_\varepsilon(x)$ setzen. „⇐“ Sei V eine offene Teilmenge mit $V \subseteq U$ und $x \in V$. Nach Definition der Offenheit gibt es ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq V$. Es folgt $B_\varepsilon(x) \subseteq U$. Also ist U eine Umgebung von x. □

Die Konvergenz von Folgen kann auch mit Hilfe von Umgebungen beschrieben werden: Eine Folge $(x_n)_{n\in\mathbb{N}}$ in einem metrischen Raum (X,d) konvergiert genau dann gegen einen Punkt $a \in X$, wenn für jede Umgebung U von a ein $N \in \mathbb{N}$ existiert, so dass $x_n \in U$ für alle $n \geq N$ erfüllt ist. (Beweis als Übung)

(4.8) Proposition. Sei X ein metrischer Raum, und seien $x, y \in X$ mit $x \neq y$. Dann gibt es Umgebungen U von x und V von y mit $U \cap V = \varnothing$. Man bezeichnet dieses Phänomen als *Hausdorffsche Trennungseigenschaft*.

Beweis: Sei $r = d(x,y)$; wegen $x \neq y$ gilt $r \in \mathbb{R}^+$. Dann sind $U = B_{r/2}(x)$ und $V = B_{r/2}(y)$ Umgebungen von x bzw. y mit der gewünschten Eigenschaft. Wäre nämlich $z \in U \cap V$, dann würde sich daraus der Widerspruch $r = d(x,y) \leq d(x,z) + d(z,y) < \frac{r}{2} + \frac{r}{2} = r$ ergeben. □

(4.9) Satz. Sei $f : X \to Y$ eine Abbildung zwischen metrischen Räumen (X, d_X) und (Y, d_Y), und sei $a \in X$. Genau dann ist f stetig in a, wenn für jede Umgebung V von $f(a)$ die Urbildmenge $f^{-1}(V)$ eine Umgebung von a ist.

Beweis: „⇐“ Wir zeigen, dass f unter der gegebenen Voraussetzung das ε-δ-Kriterium erfüllt. Sei also $\varepsilon \in \mathbb{R}^+$ vorgegeben. Setzen wir $V = B_\varepsilon(f(a))$, dann ist $f^{-1}(V)$ auf Grund des Voraussetzung eine Umgebung von a. Es gibt also ein $\delta \in \mathbb{R}^+$ mit $B_\delta(a) \subseteq f^{-1}(B_\varepsilon(f(a)))$. Sei nun $x \in X$ mit $d_X(a,x) < \delta$. Dann gilt $x \in B_\delta(a)$ und somit $f(x) \in B_\varepsilon(f(a))$, also $d_Y(f(a), f(x)) < \varepsilon$. Dies zeigt, dass das ε-δ-Kriterium erfüllt ist.

„⇒“ Sei V eine Umgebung von $f(a)$. Dann gibt es ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(f(a)) \subseteq V$. Weil f in a stetig ist, können wir das ε-δ-Kriterium anwenden. Demnach existiert ein $\delta \in \mathbb{R}^+$, so dass die Implikation $d_X(a,x) < \delta \Rightarrow d_Y(f(a), f(x)) < \varepsilon$ für alle $x \in X$ erfüllt ist. Aus $x \in B_\delta(a)$ folgt also $f(x) \in B_\varepsilon(f(a))$, d.h. es gilt $f(B_\delta(a)) \subseteq B_\varepsilon(f(a))$ und somit $B_\delta(a) \subseteq f^{-1}(B_\varepsilon(f(a))) \subseteq f^{-1}(V)$. Dies zeigt, dass $f^{-1}(V)$ eine Umgebung von a ist. □

(4.10) Folgerung. Eine Abbildung $f : X \to Y$ zwischen metrischen Räumen ist genau dann stetig, wenn das Urbild $f^{-1}(V)$ jeder offenen Menge $V \subseteq Y$ offen ist.

Beweis: „⇒“ Sei f stetig und $V \subseteq Y$ eine offene Menge. Wir müssen zeigen, dass $U = f^{-1}(V)$ offen ist. Sei dazu $a \in U$ beliebig vorgegeben. Als offene Teilmenge ist V eine Umgebung von $f(a)$. Nach (4.9) ist U damit eine Umgebung von a. Es gibt also ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(a) \subseteq U$. Weil a beliebig gewählt war, folgt daraus die Offenheit der Teilmenge U.

„⇐“ Nach Voraussetzung ist das Urbild jeder offenen Teilmenge von Y eine offene Teilmenge von X. Sei $a \in X$ vorgegeben. Zu zeigen ist, dass f in a stetig ist. Dafür wiederum genügt es nach (4.9) zu zeigen, dass für jede Umgebung V von $f(a)$ die Menge $U = f^{-1}(V)$ eine Umgebung von a ist. Seien also V und U wie angegeben. Da V

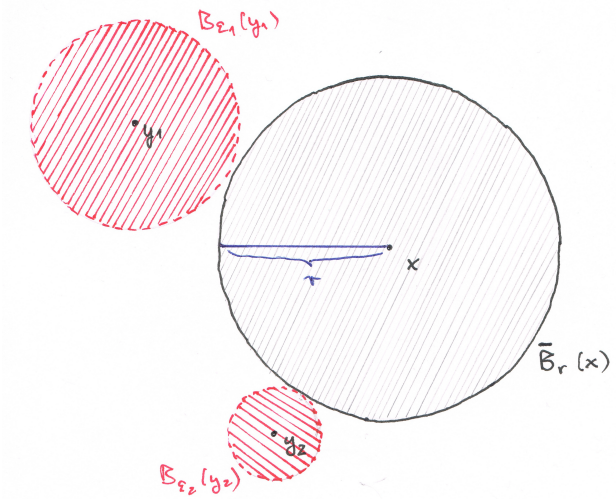
eine Umgebung von $f(a)$ ist, gibt es eine offene Menge V' mit $V' \subseteq V$ und $f(a) \in V'$. Auf Grund der Voraussetzung ist $U' = f^{-1}(V')$ offen, und aus $V' \subseteq V$ folgt $U' \subseteq U$. Wegen $f(a) \in V'$ gilt $a \in U'$. Dies zeigt, dass U eine Umgebung von a ist. □

(4.11) Definition. Sei (X, d) ein metrischer Raum. Eine Teilmenge $V \subseteq X$ wird (genau dann) als ***abgeschlossen*** bezeichnet, wenn ihr Komplement $U = X \setminus V$ offen ist.

Auch hier gilt: Ist X ein endlich-dimensionaler $\mathbb{R}$-Vektorraum, und wurde auf X keine Metrik angegeben, dann bezieht sich *abgeschlossen* immer auf die durch eine beliebige Norm induzierte Metrik. Weil die offenen Mengen nach (4.4) nicht von der Wahl der Norm abhängen, gilt dasselbe auch für die abgeschlossenen Mengen.

(4.12) Proposition. Die abgeschlossenen Intervalle in $\mathbb{R}$ der Form $[a, b] \subseteq \mathbb{R}$ mit $a, b \in \mathbb{R}$, $a < b$ sind abgeschlossen.

Beweis: Zu zeigen ist, dass die Menge $U = \mathbb{R} \setminus [a, b]$ offen ist. Sei $x \in U$ vorgegeben. Dann gilt entweder $x < a$ oder $x > b$. Setzen wir im ersten Fall $\varepsilon = a - x$, dann ist $B_\varepsilon(x) \subseteq U$ erfüllt, denn aus $y \in B_\varepsilon(x)$ folgt $|y - x| < \varepsilon$, damit insbesondere $y < x + \varepsilon = x + (a - x) = a$. Also ist y nicht im Intervall $[a, b]$ enthalten und liegt somit in U. Im zweiten Fall setzt man $\varepsilon = x - b$ und überprüft genauso, dass $B_\varepsilon(x) \subseteq U$ gilt. □



Zur Abgeschlossenheit der abgeschlossenen Bälle in einem metrischen Raum. Egal wie nahe die Punkte $y_1, y_2, \ldots$ an $\bar{B}_r(x)$ gewählt werden, eine hinreichend kleine offene Kreisscheibe um y_n ist zu $\bar{B}_r(x)$ disjunkt.

(4.13) Proposition. Sei (X, d) ein metrischer Raum, $a \in X$ und $r \in \mathbb{R}^+$. Dann ist der abgeschlossene Ball $\bar{B}_r(a)$ um a vom Radius r abgeschlossen.

Beweis: Zu zeigen ist, dass es sich bei $U = X \setminus \bar{B}_r(a)$ um eine offene Menge handelt. Sei $x \in U$ vorgegeben. Dann gilt $d(a, x) > r$. Setzen wir $\varepsilon = d(a, x) - r$, dann ist $B_\varepsilon(x) \subseteq U$ erfüllt. Wäre dies nicht der Fall, dann gäbe es einen Punkt $y \in B_\varepsilon(x) \cap \bar{B}_r(a)$. Daraus würde sich aber der Widerspruch $d(a, x) \leq d(a, y) + d(y, x) < r + \varepsilon = r + d(x, a) - r = d(x, a)$ ergeben. □

Die Aussage, dass jede Teilmenge eines metrischen Raums X entweder offen oder abgeschlossen sein muss, ist in mehrfacher Hinsicht ***falsch***, auch wenn dies durch die Wahl der Begriffe nahegelegt wird. Es gibt im allgemeinen sehr viele Teilmengen, die weder offen noch abgeschlossen sind, in $\mathbb{R}$ zum Beispiel die halboffenen Intervalle der Form $[a, b[$. Ebenso gibt es in metrischen Räumen X immer Teilmengen, die sowohl offen als auch abgeschlossen sind, zumindest die Teilmengen $\varnothing$ oder X, wie wir gleich sehen werden.

(4.14) Satz. Sei (X, d) ein metrischer Raum. Dann gilt

(i) Die Teilmengen $\varnothing$ und X sind abgeschlossen.

(ii) Sind U und V abgeschlossen, dann auch $U \cup V$.

(iii) Ist $(V_i)_{i\in I}$ eine Familie abgeschlossener Teilmengen, dann ist auch $\bigcap_{i\in I} V_i$ abgeschlossen.

Beweis: zu (i) Nach (4.5) sind die Mengen $\varnothing$ und X offen, also sind $X = X \setminus \varnothing$ und $\varnothing = X \setminus X$ abgeschlossen.

zu (ii) Die Mengen $U' = X \setminus U$ und $V' = X \setminus V$ sind nach Voraussetzung offen. Wegen (4.5) ist damit auch $U' \cap V' = X \setminus (U \cup V)$ eine offene Menge. Also ist $U \cap V$ abgeschlossen.

zu (iii) Sei $U_i = X \setminus V_i$ für alle $i \in I$. Dann ist jedes U_i offen, nach (4.5) also auch die Vereinigungsmenge $U = \bigcup_{i\in I} U_i$. Wegen $\bigcap_{i\in I} V_i = X \setminus U$ ist $\bigcap_{i\in I} V_i$ damit abgeschlossen. □

(4.15) Satz. Eine Abbildung $f : X \to Y$ zwischen metrischen Räumen ist genau dann stetig, wenn das Urbild $f^{-1}(V)$ jeder abgeschlossenen Teilmenge $V \subseteq Y$ selbst abgeschlossen ist.

Beweis: „⇒“ Ist f stetig, dann ist nach (4.10) das Urbild jeder offenen Menge offen. Sei nun $V \subseteq Y$ abgeschlossen. Setzen wir $U = Y \setminus V$, dann ist $f^{-1}(U)$ also offen. Wegen $f^{-1}(V) = X \setminus f^{-1}(U)$ ist $f^{-1}(V)$ damit abgeschlossen.

„⇐“ Sei $U \subseteq Y$ offen. Dann ist $V = Y \setminus U$ abgeschlossen, und auf Grund der Voraussetzung ist $f^{-1}(V)$ ebenfalls abgeschlossen. Daraus folgt, dass $f^{-1}(U) = X \setminus f^{-1}(V)$ offen ist. Wir haben somit gezeigt, dass das Urbild jeder offenen Menge offen ist. Nach (4.10) folgt daraus die Stetigkeit von f. □

(4.16) Satz. Eine Teilmenge $A \subseteq X$ in einem metrischen Raum (X, d_X) ist genau dann abgeschlossen, wenn mit jeder Folge $(x^{(n)})_{n\in\mathbb{N}}$ in A, die in (X, d_X) konvergiert, auch der Grenzwert $a = \lim\limits_{n\to\infty} x^{(n)}$ in A enthalten ist.

Beweis: „⇒“ Setzen wir voraus, dass $A \subseteq X$ eine abgeschlossene Teilmenge von X ist. Sei $(x^{(n)})_{n\in\mathbb{N}}$ eine in A liegende, konvergente Folge und $a = \lim_n x^{(n)}$ ihr Grenzwert. Angenommen, es gilt $a \notin A$. Weil A abgeschlossen ist, handelt es sich bei $U = X \setminus A$ um eine Umgebung von a. Es gibt also ein $N \in \mathbb{N}$ mit $x^{(n)} \in U$ für alle $n \geq N$, im Widerspruch zur Voraussetzung $x^{(n)} \in A$ für alle $n \in \mathbb{N}$.

„⇐“ Nehmen wir an, dass für jede in A liegende, in (X, d_X) konvergente Folge auch der Grenzwert in A liegt, dass aber A nicht abgeschlossen ist. Dann gibt es ein $a \in X \setminus A$, so dass für jedes $\varepsilon \in \mathbb{R}^+$ der Durchschnitt $A \cap B_\varepsilon(a)$ nichtleer ist. Insbesondere finden wir für jedes $n \in \mathbb{N}$ ein $x^{(n)} \in A \cap B_{\frac{1}{n}}(a)$. Wir erhalten so eine Folge $(x^{(n)})_{n\in\mathbb{N}}$ mit $\lim x^{(n)} = a$, deren Folgenglieder alle in A liegen, mit einem Grenzwert außerhalb von A. Dies widerspricht unserer Annahme. □

Beispielsweise ist das Intervall $I = [0, 1[$ nicht abgeschlossen in $\mathbb{R}$. Die Zahlen $x^{(n)} = 1 - \frac{1}{n}$ bilden eine Folge $(x^{(n)})_{n\in\mathbb{N}}$ in I, aber deren Grenzwert $\lim\limits_{n\to\infty} x^{(n)} = 1$ ist nicht in I enthalten.

(4.17) Definition. Eine Teilmenge $A \subseteq X$ eines metrischen Raums (X, d_X) wird ***beschränkt*** genannt, wenn die Menge $D(A) = \{d_X(a, b) \mid a, b \in A\} \subseteq \mathbb{R}_+$ beschränkt ist. Ist dies der Fall, dann bezeichnen wir $d(A) = \sup D(A)$ als den ***Durchmesser*** der Teilmenge A. Der leeren Menge $\varnothing$ wird der Durchmesser $d(\varnothing) = 0$ zugeordnet.

Eine Teilmenge A eines metrischen Raums (X, d_X) ist genau dann beschränkt, wenn ein Punkt $a \in A$ und ein $\gamma \in \mathbb{R}^+$ mit $d_X(a, x) < \gamma$ für alle $x \in A$ existiert. Ist nämlich a ein solcher Punkt, dann folgt $d(x, y) \le d(a, x) + d(a, y) < 2\gamma$ für alle $x, y \in A$. Die Umkehrung ist offensichtlich.

(4.18) Satz. *(Schachtelungsprinzip)*

Sei (X, d_X) ein vollständiger metrischer Raum und $(A_n)_{n\in\mathbb{N}}$ eine Folge nichtleerer, abgeschlossener, beschränkter Teilmengen von X mit $A_n \supseteq A_{n+1}$ für alle $n \in \mathbb{N}$. Gilt $\lim_n d(A_n) = 0$, dann gibt es einen eindeutig bestimmten Punkt $a \in X$ mit $\bigcap_{n\in\mathbb{N}} A_n = \{a\}$.

Beweis: Zunächst beweisen wir die Eindeutigkeit. Sind $a, a' \in A$ zwei verschiedene Punkte mit $a, a' \in A_n$ für alle $n \in \mathbb{N}$, dann gilt $d(a, a') \le d(A_n)$ für alle $n \in \mathbb{N}$. Aus $\lim_n d(A_n) = 0$ folgt $d(a, a') = 0$ und $a = a'$, im Widerspruch zur Voraussetzung.

Zum Nachweis der Existenz wählen wir in jeder Menge A_n einen Punkt $x^{(n)}$. Sei $\varepsilon \in \mathbb{R}^+$ vorgegeben und $N \in \mathbb{N}$ so groß gewählt, dass $d(A_n) < \varepsilon$ für alle $n \ge N$ erfüllt ist. Sind nun $m, n \in \mathbb{N}$ mit $n \ge m \ge N$, dann folgt $d(x^{(n)}, x^{(m)}) \le d(A_m) < \varepsilon$, also ist $(x^{(n)})_{n\in\mathbb{N}}$ eine Cauchyfolge. Auf Grund der Vollständigkeit von (X, d_X) konvergiert diese gegen einen Punkt $a \in X$. Ist $m \in \mathbb{N}$ beliebig, dann gilt $x^{(n)} \in A_n \subseteq A_m$ für alle $n \ge m$. Weil A_m abgeschlossen ist, liegt nach (4.16) auch der Grenzwert a der Folge in A_m. Weil m beliebig vorgegeben war, haben wir somit $a \in A_m$ für alle $m \in \mathbb{N}$ und somit $a \in \bigcap_{n\in\mathbb{N}} A_n$ nachgewiesen. □

Wie bereits oben angesprochen, gibt es in der Regel viele Teilmengen eines metrischen Raums, die weder offen noch abgeschlossen sind. Man kann aber jeder Teilmenge A eine offene Menge A° und eine abgeschlossene Menge $\bar{A}$ mit $A^\circ \subseteq A \subseteq \bar{A}$ zuordnen.

(4.19) Definition. Sei (X, d_X) ein metrischer Raum und $A \subseteq X$ eine Teilmenge.

(i) Man nennt x einen ***inneren Punkt*** von A, wenn ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq A$ existiert. Die Menge der inneren Punkte von A wird mit A° bezeichnet und das ***Innere*** von A genannt.

(ii) Ein Punkt $x \in X$ liegt im ***Abschluss*** von A, wenn für jedes $\varepsilon \in \mathbb{R}^+$ der Durchschnitt $A \cap B_\varepsilon(x)$ nicht leer ist.

(iii) Die Menge $\partial A = \bar{A} \setminus A^\circ$ wird der ***Rand*** von A genannt.

Man beachte, dass A° leer sein kann; dies gilt zum Beispiel für jede einelementige Teilmenge von $\mathbb{R}$, oder für jede Teilmenge von $\mathbb{R}^3$, die in einer Ebene enthalten ist. Ebenso ist $\bar{A} = X$ möglich. Wenn das der Fall ist, bezeichnet man A als ***dichte*** Teilmenge von $\mathbb{R}$. Beispielsweise ist $\mathbb{Q}$ eine dichte Teilmenge von $\mathbb{R}$ bezüglich der Standard-Metrik.

(4.20) Proposition. Sei (X, d_X) ein metrischer Raum und $A \subseteq X$ eine Teilmenge.

(i) Die Menge A° ist die größte offene Teilmenge von X, die in A enthalten ist. Ist also $U \subseteq A$ offen in X, dann gilt $U \subseteq A^\circ$.

(ii) Die Menge $\bar{A}$ ist die kleinste abgeschlossene Teilmenge von X, die A enthält. Ist also $Z \subseteq X$ eine beliebige abgeschlossene Teilmenge mit $Z \subseteq A$, dann folgt $\bar{A} \subseteq Z$.

(iii) Ein Punkt $x \in X$ gehört genau dann zum Rand ∂A, wenn $B_\varepsilon(x)$ für jedes $\varepsilon \in \mathbb{R}^+$ jeweils mindestens einen Punkt von A und einen Punkt des Komplements $X \setminus A$ enthält.

Beweis: zu (i) Zunächst einmal gilt $A^\circ \subseteq A$. Ist nämlich $x \in A^\circ$ und $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq A$, dann gilt $x \in B_\varepsilon(x)$ und somit $x \in A$. Die Menge A° ist auch offen. Ist nämlich $x \in A^\circ$, dann gibt es nach Definition ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq A$. Weil $B_\varepsilon(x)$ nach (4.3) offen ist, gibt es auch für jedes $y \in B_\varepsilon(x)$ ein $\varepsilon' \in \mathbb{R}^+$ mit $B_{\varepsilon'}(y) \subseteq B_\varepsilon(x) \subseteq A$. Dies zeigt, dass jedes $y \in B_\varepsilon(x)$ ein innerer Punkt von A, insgesamt also $B_\varepsilon(x) \subseteq A^\circ$ gilt. Sei nun U eine beliebige offene Teilmenge von X mit $U \subseteq A$, und sei $x \in U$. Auf Grund der Offenheit von U gibt es ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq U$. Es gilt dann auch $B_\varepsilon(x) \subseteq A$, und daraus wiederum folgt $x \in A^\circ$.

zu (ii) Zum Nachweis von $\bar{A} \supseteq A$ sei $x \in A$ vorgegeben. Für jedes $\varepsilon \in \mathbb{R}^+$ gilt $x \in B_\varepsilon(x)$. Wegen $x \in A$ folgt daraus $x \in \bar{A}$. Die Menge $\bar{A}$ ist abgeschlossen. Ist nämlich $x \in X \setminus \bar{A}$, dann existiert nach Definition von $\bar{A}$ ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \cap A = \emptyset$. Weil $B_\varepsilon(x)$ offen ist, gibt es für jedes $y \in B_\varepsilon(x)$ ein $\varepsilon' \in \mathbb{R}^+$ mit $B_{\varepsilon'}(y) \subseteq B_\varepsilon(x)$. Wegen $B_\varepsilon(x) \cap A = \emptyset$ gilt auch $B_{\varepsilon'}(y) \cap A = \emptyset$ und damit $y \in X \setminus \bar{A}$. Damit ist $B_\varepsilon(x) \subseteq X \setminus \bar{A}$ nachgewiesen; dies zeigt, dass $\bar{A}$ abgeschlossen ist. Sei nun Z eine abgeschlossene Teilmenge von X mit $Z \supseteq A$, und sei $x \in \bar{A}$ vorgegeben. Zu zeigen ist $x \in Z$, was zu $x \notin X \setminus Z$ äquivalent ist. Wäre x in $X \setminus Z$ enthalten, dann gäbe es auf Grund der Offenheit dieser Menge ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq X \setminus Z$. Dies ist mit $B_\varepsilon(x) \cap Z = \emptyset$ gleichbedeutend, und wegen $Z \supseteq A$ folgt daraus $B_\varepsilon \cap A = \emptyset$. Aber wegen $x \in \bar{A}$ muss jeder offene ε-Ball mit A einen nichtleeren Schnitt haben. Also war die Annahme $x \in X \setminus Z$ falsch, und es muss $x \in Z$ gelten.

zu (iii) „⇒“ Sei $x \in \partial A$. Dann gilt insbesondere $x \in \bar{A}$, somit enthält $B_\varepsilon(x)$ für jedes $\varepsilon \in \mathbb{R}^+$ einen Punkt aus A. Andererseits gilt auch $x \notin A^\circ$. Gäbe es ein $\varepsilon \in \mathbb{R}^+$ mit der Eigenschaft, dass $B_\varepsilon(x)$ keine Punkte aus $X \setminus A$ enthält, dann würde $B_\varepsilon(x) \subseteq A$ folgen. Dies würde $x \in A^\circ$ bedeuten, im Widerspruch zu unserer Annahme. „⇐“ Weil $B_\varepsilon(x)$

für jedes $\varepsilon \in \mathbb{R}^+$ einen Punkt aus A enthält, gilt $x \in \bar{A}$. Nehmen wir nun an, es würde $x \in A^\circ$ gelten. Dann gäbe es ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq A$; der Schnitt von $B_\varepsilon(x)$ mit $X \setminus A$ wäre also leer, im Widerspruch zur Voraussetzung. So aber gilt $x \notin A^\circ$, insgesamt ist x also in $\partial A = \bar{A} \setminus A^\circ$ enthalten. □

(4.21) Definition. Sei (X, d_X) ein metrischer Raum und $Y \subseteq X$ eine Teilmenge. Dann erhalten wir durch Einschränkung von d_X auf $Y \times Y$ eine Metrik d_Y auf Y mit $d_Y(a,b) = d_X(a,b)$ für alle $a, b \in Y$. Wir bezeichnen eine Teilmenge $U \subseteq Y$ als ***relativ offen*** bzw. ***abgeschlossen in*** Y, wenn U bezüglich d_Y offen bzw. abgeschlossen ist.

(4.22) Satz. Sei (X, d_X) ein metrischer Raum, $Y \subseteq X$ eine Teilmenge und $U \subseteq Y$.

(i) Genau dann ist U relativ offen in Y, wenn eine offene Teilmenge $\tilde{U} \subseteq X$ mit der Eigenschaft $U = \tilde{U} \cap Y$ existiert.

(ii) Genau dann ist U relativ abgeschlossen in Y, wenn eine abgeschlossene Teilmenge $\tilde{U} \subseteq X$ mit $U = \tilde{U} \cap Y$ existiert.

Beweis: Wir beschränken uns auf den Beweis von (i), weil der Beweis des anderen Teils weitgehend analog verläuft. Es sei d_Y die Metrik auf Y, die man durch Einschränkung von d_X auf $Y \times Y$ erhält.

„⇐“ Sei $\tilde{U} \subseteq X$ eine Menge mit den angegebenen Eigenschaften und $a \in U$. Weil $\tilde{U}$ offen ist, gibt es ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(a) \subseteq \tilde{U}$, wobei $B_\varepsilon(a)$ den offenen Ball vom Radius ε um U bezüglich der Metrik d_X bezeichnet. Für alle $x \in Y$ gilt die Äquivalenz $x \in B_\varepsilon(a) \Leftrightarrow d_X(a,x) < \varepsilon \Leftrightarrow d_Y(a,x) < \varepsilon$, also ist $B_\varepsilon(a) \cap Y$ der offene Ball vom Radius ε um a bezüglich d_Y. Die Inklusion $B_\varepsilon(a) \cap Y \subseteq \tilde{U} \cap Y = U$ zeigt, dass U in Y relativ offen ist.

„⇒“ Sei $U \subseteq Y$ eine in Y relativ offene Teilmenge. Dann gibt es für jedes $a \in U$ ein $\varepsilon(a) \in \mathbb{R}^+$, so dass der offene Ball

$$B_{\varepsilon(a),Y}(a) \quad = \quad \{y \in Y \mid d_Y(a,y) < \varepsilon(a)\} \quad = \quad \{y \in Y \mid d_X(a,y) < \varepsilon(a)\} \quad = \quad B_{\varepsilon(a)}(a) \cap Y$$

um a vom Radius $\varepsilon(a)$ bezüglich d_Y in U enthalten ist. Es gilt also $B_{\varepsilon(a)}(a) \cap Y = B_{\varepsilon(a),Y}(a) \subseteq U$ für alle $a \in U$. Die Menge $\tilde{U} = \bigcup_{a \in U} B_{\varepsilon(a)}$ ist offen in (X, d_X), und es gilt

$$\tilde{U} \cap Y \quad = \quad \left(\bigcup_{a \in U} B_{\varepsilon(a)}(a) \right) \cap Y \quad = \quad \bigcup_{a \in U} (B_{\varepsilon(a)}(a) \cap Y) \quad = \quad U. \qquad \square$$

Die Teilmenge $U =]-1,1[\times \{0\} \subseteq \mathbb{R}^2$ ist zwar keine offene Menge in $X = \mathbb{R}^2$, da zum Beispiel keine ε-Umgebung $B_\varepsilon((0,0))$ von $(0,0)$ in X enthalten ist. Sie ist aber relativ offen in $Y = \mathbb{R} \times \{0\}$, denn sie kann als Durchschnitt von Y und der in $\mathbb{R}^2$ offenen Menge $\tilde{U} =]-1,1[\times \mathbb{R}$ dargestellt werden.

§ 5. Kompaktheit

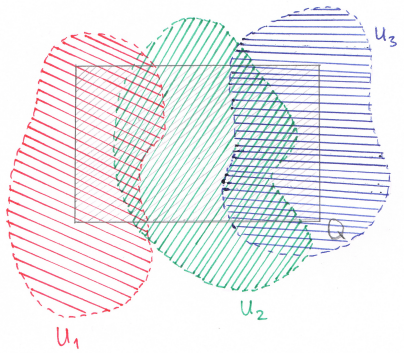
Zusammenfassung. Die kompakten Teilmengen eines metrischen Raums sind das Analogon der endlichen abgeschlossenen Intervalle in der eindimensionalen Analysis. Für sie gilt das Maximumsprinzip und der Satz von Bolzano-Weierstraß (der besagt, dass jede Folge in einem endlichen abgeschlossenen Intervall bzw. einem metrischen Raum eine konvergente Teilfolge besitzt).

Für die Definition der Kompaktheit in allgemeinen metrischen Räumen benötigt man den Begriff der *offenen Überdeckung*. Im $\mathbb{R}^n$ sind die kompakten Teilmengen einfach die beschränkten und abgeschlossenen. Diese Charakterisierung ist aber in allgemeinen metrischen Räumen nicht mehr gültig.

Wichtige Begriffe und Sätze in diesem Kapitel

- offene Überdeckungen, Definition der Kompaktheit
- Endliche abgeschlossene Quader sind kompakt.
- Die kompakten Teilmengen im $\mathbb{R}^n$ sind genau die beschränkten und abgeschlossenen Mengen.
- Satz von Bolzano-Weierstrass
- Maximumsprinzip
- gleichmäßige Stetigkeit einer Abbildung zwischen metrischen Räumen
- Stetige Abbildungen auf kompakten metrischen Räumen sind gleichmäßig stetig.

(5.1) Definition. Sei I eine Menge, (X,d) ein metrischer Raum und $A \subseteq X$ eine Teilmenge. Eine ***offene Überdeckung*** von A ist eine Familie $(U_i)_{i\in I}$ offener Teilmengen $U_i \subseteq X$ mit der Eigenschaft $\bigcup_{i\in I} U_i \supseteq A$.



Überdeckung einer Menge Q durch drei offene Mengen U_1, U_2, U_3

Für $n \in \mathbb{Z}$ sei beispielsweise $U_n =]n, n+2[$ und $V_n =]n, n+1[$. Dann ist $(U_n)_{n\in\mathbb{Z}}$ eine offene Überdeckung von $\mathbb{R}$, die Familie $(V_n)_{n\in\mathbb{Z}}$ aber nicht, denn die Teilmenge $\mathbb{Z} \subseteq \mathbb{R}$ ist in $\bigcup_{n\in\mathbb{Z}} V_n$ nicht enthalten.

(5.2) Definition. Sei (X, d_X) ein metrischer Raum. Eine Teilmenge $A \subseteq X$ wird ***kompakt*** genannt, wenn zu jeder offenen Überdeckung $(U_i)_{i\in I}$ von A eine endliche Teilmenge $J \subseteq I$ existiert, so dass bereits $\bigcup_{i\in J} U_i \supseteq A$ erfüllt ist. Den metrischen Raum selbst bezeichnen wir als kompakt, wenn die Teilmenge $A = X$ in (X, d_X) kompakt ist.

Häufig verwendet man für die Kompaktheit von A die Formulierung „In jeder offenen Überdeckung von A gibt es eine endliche Teilüberdeckung." Dies ist aber ***nicht*** einfach damit gleichbedeutend, dass A eine endliche offene Überdeckung besitzt. Letzteres trifft für jede Teilmenge $A \subseteq X$ zu: Die einelementige Familie $(U_i)_{i\in\{1\}}$ gegeben durch $U_1 = X$ ist offensichtlich eine Überdeckung von A. Bei der Definition der Kompaktheit liegt die Betonung darauf, dass in ***jeder*** offenen Überdeckung von A eine endliche Teilüberdeckung gewählt werden kann.

Schauen wir uns nun einige Beispiele und Gegenbeispiele für kompakte Mengen an.

(5.3) Proposition. Sei (X, d_X) ein metrischer Raum und $A \subseteq X$ eine endliche Teilmenge. Dann ist A kompakt.

Beweis: Sei $r \in \mathbb{N}_0$, und seien $a_1, ..., a_r$ die verschiedenen Elemente von A. Sei außerdem $(U_i)_{i\in I}$ eine offene Überdeckung von A. Wegen $a_k \in A$ und $A \subseteq \bigcup_{i\in I} U_i$ gibt es für jedes $k \in \{1, ..., r\}$ ein $i(k) \in I$ mit $a_k \in U_{i(k)}$. Setzen wir $J = \{i(1), ..., i(r)\}$, dann ist $\bigcup_{i\in J} U_i \supseteq A$ offenbar erfüllt. □

(5.4) Proposition. Die Menge $\mathbb{R}$ im metrischen Raum $(\mathbb{R}, d)$ mit der Metrik $d(x, y) = |x - y|$ für $x, y \in \mathbb{R}$ ist nicht kompakt.

Beweis: Die Familie $(U_n)_{n\in\mathbb{Z}}$ gegeben durch $U_n =]n, n + 2[$ ist eine offene Überdeckung von $\mathbb{R}$. In $(U_n)_{n\in\mathbb{Z}}$ kann aber keine endliche Teilüberdeckung gewählt werden, denn jede Vereinigung der Form $U_{n_1} \cup ... \cup U_{n_r}$ mit $r \in \mathbb{N}_0$ und $n_1, ..., n_r \in \mathbb{Z}$ hat nur endlichen Durchmesser und enthält somit nicht ganz $\mathbb{R}$. □

(5.5) Proposition. Sei $(x^{(n)})_{n\in\mathbb{N}}$ eine konvergente Folge in einem metrischen Raum (X, d) mit Grenzwert $a = \lim_n x^{(n)}$. Dann ist die Menge $A = \{x^{(n)} \mid n \in \mathbb{N}\} \cup \{a\}$ kompakt.

Beweis: Sei $(U_i)_{i\in I}$ eine offene Überdeckung von A. Dann gibt es insbesondere ein $i_0 \in I$ mit $a \in U_{i_0}$, und U_{i_0} ist eine Umgebung von a. Auf Grund der Konvergenz finden wir somit ein $N \in \mathbb{N}$ mit $x^{(n)} \in U_{i_0}$ für alle $n \geq N$. Für jedes $n \in \mathbb{N}$ mit $n < N$ gibt es außerdem auf Grund der Überdeckungseigenschaft ein $i_n \in I$ mit $x^{(n)} \in U_{i_n}$. Ein Folgenglied $x^{(n)}$ liegt für $n < N$ also in U_{i_n} und für $n \geq N$ in U_{i_0}. Dies zeigt, dass die endliche Indexmenge $J = \{i_0, i_1, ..., i_{N-1}\}$ eine endliche Teilüberdeckung von $(U_i)_{i\in I}$ definiert. □

Nimmt man aus der Menge A den Grenzwert a heraus, so erhält man im allgemeinen keine kompakte Menge mehr. Ist die Folge beispielsweise durch $x^{(n)} = \frac{1}{n}$ für alle $n \in \mathbb{N}$ gegeben, dann in der Überdeckung $(U_n)_{n\in\mathbb{N}}$ mit $U_n = \left]\frac{1}{2n}, \frac{3}{2n}\right[$ für $n \in \mathbb{N}$ keine endliche Teilüberdeckung.

Wie schon bei der Offenheit und der Abgeschlossenheit bezieht sich der Begriff *kompakt* bei Teilmengen eines endlichdimensionalen $\mathbb{R}$-Vektorraums auf die induzierte Metrik bezüglich einer beliebigen Norm, solange nicht explizit eine andere Metrik vorgegeben wurde. Weil die Kompaktheit direkt auf dem Begriff der offenen Teilmenge basiert, folgt aus (4.4) unmittelbar, dass auch diese Eigenschaft nicht von der Wahl der Norm abhängig ist.

(5.6) Definition. Ein ***abgeschlossener Quader*** in $\mathbb{R}^n$ ist eine Teilmenge $Q \subseteq \mathbb{R}^n$ der Form $Q = I_1 \times \ldots \times I_n$, wobei $I_k \subseteq \mathbb{R}$ für $1 \leq k \leq n$ jeweils ein abgeschlossenes Intervall $[a_k, b_k]$ mit $a_k < b_k$ bezeichnet.

Man überprüft leicht, dass jeder abgeschlossene Quader dieser Form eine abgeschlossene Teilmenge von $\mathbb{R}^n$ ist. Der Durchmesser $d(Q)$ von Q bezüglich der Maximums-Norm $\|\cdot\|_\infty$ ist gegeben durch $d(Q) = \max\{b_k - a_k \mid 1 \leq k \leq n\}$.

(5.7) Satz. Jeder abgeschlossene Quader $Q \subseteq \mathbb{R}^n$ ist kompakt.

Beweis: Nehmen wir an, dass Q nicht kompakt ist. Dann gibt es in $\mathbb{R}^n$ eine offene Überdeckung $(U_i)_{i\in I}$ von Q, in der aber keine endliche Teilüberdeckung existiert. Wir definieren nun eine absteigende Folge

$$Q \quad = \quad Q_0 \quad \supseteq \quad Q_1 \quad \supseteq \quad Q_2 \quad \supseteq \quad Q_3 \quad \supseteq \quad \ldots$$

von abgeschlossenen Quadern in $\mathbb{R}^n$ mit $d(Q_m) = 2^{-m} d(Q)$, so dass keiner dieser Quader in $(U_i)_{i\in I}$ eine endliche Teilüberdeckung besitzt. Sei $m \in \mathbb{N}_0$ und Q_m bereits definiert, wobei $Q_m = 2^{-m} d(Q)$ gilt und Q_m keine endliche Teilüberdeckung in $(U_i)_{i\in I}$ besitzt. Dann erhalten wir Q_{m+1} durch das folgende Verfahren: Ist $Q_m = I_1 \times \ldots \times I_n$ mit abgeschlossenen Intervallen $I_k = [a_k, b_k]$, dann zerlegen wir das Intervall I_k jeweils in die beiden Teilstücke

$$I_k^{(1)} \quad = \quad [a_k, m_k] \qquad \text{und} \qquad I_k^{(2)} \quad = \quad [m_k, b_k] \quad ,$$

wobei $m_k = \frac{1}{2}(a_k + b_k)$ den Mittelpunkt von I_k bezeichnet. Für jedes Tupel $s = (s_1, \ldots, s_n) \in \{1,2\}^n$ sei $Q^{(s)} = I_1^{(s_1)} \times \ldots \times I_n^{(s_n)}$. Auf diese Weise haben wir Q_m in insgesamt 2^n Teilquader zerlegt. Weil die Intervalle, aus denen $Q^{(s)}$ gebildet wird, jeweils halb so lang wie die Intervalle von Q_m sind, gilt $d(Q^{(s)}) = \frac{1}{2} d(Q_m) = 2^{-(m+1)} d(Q)$ für alle $s \in \{1,2\}^n$. Es gibt mindestens ein $t \in \{1,2\}^n$, so dass $Q^{(t)}$ in $(U_i)_{i\in I}$ keine endliche Teilüberdeckung besitzt. Ansonsten könnten wir nämlich die 2^s endlichen Überdeckungen der Quader $Q^{(s)}$ zu einer endlichen Überdeckung von Q_m zusammenfügen. Definieren wir nun $Q_{m+1} = Q^{(t)}$, dann erfüllt Q_{m+1} alle angegebenen Bedingungen.

Nun zeigen wir, dass diese Konstruktion zu einem Widerspruch führt. Nach dem Schachtelungsprinzip (4.18) gibt es ein $a \in Q$ mit $\bigcap_{m\in\mathbb{N}} Q_m = \{a\}$. Weil $(U_i)_{i\in I}$ eine Überdeckung von Q ist, gibt es ein $i_0 \in I$ mit $a \in U_{i_0}$. Sei $\varepsilon \in \mathbb{R}^+$ so klein gewählt, dass $B_\varepsilon(a)$ in U_{i_0} enthalten ist, wobei $B_\varepsilon(a)$ den offenen Ball bezüglich $\|\cdot\|_\infty$ bezeichnet. Sei $m \in \mathbb{N}$ mit $2^{-m} < \varepsilon$. Wir zeigen, dass Q_m in U_{i_0} liegt. Sei $x \in Q_m$ vorgegeben. Aus $a, x \in Q_m$ und $d(Q_m) < 2^{-m} < \varepsilon$ folgt $d(a,x) < \varepsilon$. Dies wiederum bedeutet $x \in B_\varepsilon(a) \subseteq U_{i_0}$. Aber die Inklusion $Q_m \subseteq U_{i_0}$ widerspricht der Annahme, dass Q_m nicht durch endlich viele Mengen aus der Familie $(U_i)_{i\in I}$ überdeckt werden kann. □

(5.8) Satz. Sei A eine abgeschlossene Teilmenge eines kompakten metrischen Raums (X, d_X). Dann ist A kompakt.

Beweis: Sei $(U_i)_{i\in I}$ eine offene Überdeckung von A. Weil $U = X \setminus A$ offen ist, bildet $(U_i)_{i\in I}$ zusammen mit U eine offene Überdeckung von X. Weil X kompakt ist, können wir eine endliche Teilmenge $J \subseteq I$ wählen, so dass $(U_i)_{i\in J}$ zusammen mit U eine endliche Überdeckung von X bildet. Es gilt also

$$A \subseteq X = U \cup \bigcup_{j\in J} U_j = (X \setminus A) \cup \bigcup_{j\in J} U_j.$$

Die Gleichung zeigt $A \subseteq \bigcup_{j\in J} U_j$, wir haben also eine endliche Teilüberdeckung von A in $(U_i)_{i\in I}$ gefunden. □

(5.9) Folgerung. Jede beschränkte und abgeschlossene Teilmenge $A \subseteq \mathbb{R}^n$ ist kompakt.

Beweis: Weil A beschränkt ist, gibt es ein $r \in \mathbb{R}^+$, so dass A im offenen Ball $B_r(0_{\mathbb{R}^n})$ bezüglich der Maximums-Norm $\|\cdot\|_\infty$ enthalten ist. Nach Definition ist der *abgeschlossene* Ball $\bar{B}_r(0_{\mathbb{R}^n})$ gerade der abgeschlossene Quader $[-r, r]^n$, und dieser ist nach (5.7). eine kompakte Teilmenge von $\mathbb{R}^n$. Als abgeschlossene Teilmenge von $\bar{B}_r(0_{\mathbb{R}^n})$ ist A nach (5.8) ebenfalls kompakt. □

Kompakte Teilmengen metrischer Räume lassen sich auch durch das Konvergenzverhalten von Folgen beschreiben.

(5.10) Definition. Ein Punkt x in einem metrischen Raum (X, d_X) wird ***Häufungspunkt*** einer Teilmenge $A \subseteq X$ genannt, wenn in jeder Umgebung von x jeweils unendlich viele Elemente aus A liegen.

In Analogie zu den endlichen, abgeschlossenen Intervallen von I gilt nun

(5.11) Satz. *(Satz von Bolzano-Weierstraß)*
Jede Folge in einem kompakten metrischen Raum (X, d_X) besitzt eine konvergente Teilfolge.

Beweis: Sei $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge in X. Ist die Menge $A = \{x^{(n)} \mid n \in \mathbb{N}\}$ der Folgenglieder endlich, dann gibt es ein $a \in A$ mit $x^{(n)} = a$ für unendlich viele $n \in \mathbb{N}$. Wir finden dann eine Folge $(n_k)_{k\in\mathbb{N}}$ natürlicher Zahlen mit $x^{(n_k)} = a$ für alle $k \in \mathbb{N}$. Damit ist $(x^{(n_k)})_{k\in\mathbb{N}}$ eine konstante, insbesondere eine konvergente Teilfolge von $(x^{(n)})_{n\in\mathbb{N}}$.

Setzen wir von nun an voraus, dass A eine unendliche Menge ist. Wir beweisen, dass A unter dieser Voraussetzung einen Häufungspunkt besitzt. Wäre dies nicht der Fall, dann gäbe es für jedes $x \in X$ eine Umgebung $U_x \subseteq X$, so dass die Menge $U_x \cap A$ jeweils endlich ist. Die Familie $(U_x)_{x\in X}$ bildet eine Überdeckung von A. Auf Grund der Kompaktheit existiert eine endliche Teilmenge $J \subseteq X$, so dass A bereits von $(U_x)_{x\in J}$ überdeckt wird. Aber dies würde bedeuten, dass

$$A = A \cap X = A \cap \left(\bigcup_{x\in J} U_x \right) = \bigcup_{x\in J} (A \cap U_x)$$

eine endliche Menge ist, im Widerspruch zur Annahme. Sei $a \in X$ also ein Häufungspunkt von A. Dann enthält $B_{1/k}(a)$ für jedes $k \in \mathbb{N}$ jeweils unendlich viele Folgenglieder, wir finden also ein $n_k \in \mathbb{N}$ mit $x^{(n_k)} \in B_{1/k}(a)$. Nach Konstruktion konvergiert die Teilfolge $(x^{(n_k)})_{k\in\mathbb{N}}$ dann gegen den Punkt a. □

(5.12) Folgerung. Jede beschränkte Folge in $\mathbb{R}^n$ besitzt eine konvergente Teilfolge.

Beweis: Dies folgt aus dem Satz von Bolzano-Weierstraß, weil jede beschränkte Folge in einem hinreichend groß gewählten kompakten Quader enthalten ist. □

(5.13) Folgerung. Jede kompakte Teilmenge $A \subseteq X$ eines metrischen Raums (X, d_X) ist beschränkt und abgeschlossen.

Beweis: Nehmen wir an, A wäre nicht beschränkt. Dann wählen wir einen beliebigen Punkt $a \in A$ und definieren eine offene Überdeckung von A durch $(U_n)_{n\in\mathbb{N}}$ durch $U_n = B_n(a)$ für alle $n \in \mathbb{N}$. Da A unbeschränkt ist, gibt es in dieser offenen Überdeckung keine endliche Teilüberdeckung. Also ist A nicht kompakt, im Widerspruch zur Annahme.

Gehen wir nun davon aus, dass A nicht abgeschlossen ist. Dann gibt es nach (4.16) eine Folge $(x^{(n)})_{n\in\mathbb{N}}$ mit einem Grenzwert $x \in X$, der außerhalb von A liegt. Jede Teilfolge von $(x^{(n)})_{n\in\mathbb{N}}$ konvergiert dann ebenfalls gegen x. Aber nach dem Satz von Bolzano-Weierstraß gibt es eine Teilfolge, die gegen einen Punkt in A konvergiert. Weil eine Folge nicht gegen zwei verschiedene Punkt konvergieren kann (siehe (2.2)), erhalten wir auch hier einen Widerspruch. □

Zusammen mit (5.9) erhalten wir

(5.14) Satz. *(Satz von Heine-Borel)*

Eine Teilmenge $A \subseteq \mathbb{R}^n$ ist genau dann kompakt, wenn sie beschränkt und abgeschlossen ist.

Abeschlossene und beschränkte Teilmengen allgemeiner metrischer Räume sind nicht notwendigerweise auch kompakt. Beispielsweise gilt

(5.15) Proposition. Sei X eine unendliche Menge und δ_X die diskrete Topologie auf X. Dann ist X zwar beschränkt und abgeschlossen, aber nicht kompakt.

Beweis: Weil δ_X nur die Werte 0 und 1 annimmt, gilt $d(X) = 1$ für den Durchmesser von X. Also ist X beschränkt. Wie in jedem metrischen Raum ist die Gesamtmenge X abgeschlossen. Darüber hinaus ist, wie wir in 2.4 gesehen haben, jede Teilmenge eines diskreten metrischen Raums offen und damit auch abgeschlossen. Andererseits ist X nicht kompakt. Weil nämlich jede Teilmenge von X offen ist, erhalten wir durch $(\{x\})_{x\in X}$ eine offene Überdeckung von X. Weil aber X unendlich ist, können wir in $(\{x\})_{x\in X}$ keine endliche Teilüberdeckung wählen. □

(5.16) Satz. Sei $f : X \to Y$ eine stetige Abbildung zwischen metrischen Räumen, wobei X kompakt sei. Dann ist auch die Bildmenge $f(X)$ kompakt.

Beweis: Sei $(V_i)_{i\in I}$ eine offene Überdeckung von $f(X)$. Dann sind die Urbildmengen $U_i = f^{-1}(V_i)$ auf Grund der Stetigkeit von f offen (siehe (4.10)), und sie bilden eine Überdeckung des Definitionsbereichs X der Abbildung. Weil X kompakt ist, gibt es eine endliche Teilmenge $J \subseteq I$, so dass X bereits von $(U_i)_{i\in J}$ überdeckt wird. Dann ist $(V_i)_{i\in J}$ eine Überdeckung von $f(X)$. Ist nämlich $y \in f(X)$ vorgegeben, dann existiert ein $x \in X$ mit $f(x) = y$ und ein $i \in J$ mit $x \in U_i = f^{-1}(V_i)$. Es folgt dann $y = f(x) \in V_i$. □

(5.17) Satz. *(Maximumsprinzip)*

Jede stetige, reellwertige Funktion $f : X \to \mathbb{R}$ auf einem kompakten metrischen Raum X ist beschränkt und nimmt auf X ihr Maximum und Minimum an.

Beweis: Nach (5.16) ist $f(X) \subseteq \mathbb{R}$ beschränkt und abgeschlossen. Weil $f(X)$ beschränkt ist, besitzt diese Menge ein Infimum m_- und ein Supremum m_+. Nehmen wir an, dass m_+ nicht in $f(X)$ liegt. Dann gibt es nach Definition des Supremums eine Folge $(y^{(n)})_{n\in\mathbb{N}}$ in $f(X)$ mit $\lim_n y^{(n)} = m_+$. Aber weil $f(X)$ abgeschlossen ist, muss nach (4.16) auch der Grenzwert m_+ in $f(X)$ liegen. Genauso beweist man $m_- \in f(X)$. □

(5.18) Definition. Eine Abbildung $f : X \to Y$ zwischen metrischen Räumen (X, d_X) und (Y, d_Y) wird ***gleichmäßig stetig*** auf X genannt, wenn für jedes $\varepsilon \in \mathbb{R}^+$ ein $\delta \in \mathbb{R}^+$ existiert, so dass die Implikation $d_X(x, y) < \delta \Rightarrow d_Y(f(x), f(y)) < \varepsilon$ für alle $x, y \in X$ erfüllt ist.

(5.19) Satz. Sei $f : X \to Y$ eine stetige Abbildung zwischen metrischen Räumen, wobei X kompakt sei. Dann ist f sogar gleichmäßig stetig auf X.

Beweis: Sei $\varepsilon \in \mathbb{R}^+$ vorgegeben. Weil f auf X stetig ist, können wir für jeden Punkt $a \in X$ das ε-δ-Kriterium ((3.5)) anwenden und erhalten ein $\delta_a \in \mathbb{R}^+$, so dass die Implikation $d_X(a, x) < \delta_a \Rightarrow d_Y(f(a), f(x)) < \frac{1}{2}\varepsilon$ für alle $x \in X$ erfüllt ist. Lassen wir a die gesamte Menge X durchlaufen, dann bilden die offenen Bälle $U_a = B_{\frac{1}{2}\delta_a}(a)$ bilden einen offene Überdeckung von X. Weil X kompakt ist, können wir eine endliche Teilmenge $J \subseteq X$ wählen, so dass X bereits durch die Mengen U_a mit $a \in J$ überdeckt wird.

Sei nun $\delta = \min\{\frac{1}{2}\delta_a \,\big|\, a \in J\}$, und seien $x, y \in X$ mit $d_X(x, y) < \delta$ vorgegeben. Auf Grund der Überdeckungseigenschaft finden wir ein $a \in J$ mit $x \in U_a$. Es folgt $d_X(a, x) < \frac{1}{2}\delta_a$, und zusammen mit $d_X(x, y) < \delta \leq \frac{1}{2}\delta_a$ erhalten wir $d_X(a, y) < \delta_a$. Aus $d_X(a, x) < \frac{1}{2}\delta_a < \delta_a$ folgt $d_Y(f(a), f(x)) < \frac{1}{2}\varepsilon$, und aus $d_X(a, y) < \delta_a$ folgt $d_Y(f(a), f(y)) < \frac{1}{2}\varepsilon$. Mit der Dreiecksungleichung erhalten wir insgesamt $d_Y(f(x), f(y)) \leq d_Y(f(a), f(x)) + d_Y(f(a), f(y)) < \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon$. □

§ 6. Zusammenhang

Zusammenfassung. Besteht eine Teilmenge A eines metrischen Raums X aus mehreren voneinander getrennten Komponenten (wie beispielsweise die Teilmenge $[0,1] \cup [2,3]$ von $\mathbb{R}$), dann sind diese Komponenten (hier $[0,1]$ und $[2,3]$) in A offene Teilmengen. Man kann A also als disjunkte Vereinigung mehrerer echter relativ offener Teilmengen darstellen. Die Menge A soll zusammenhängend sein, wenn eine solche Zerlegung nicht möglich ist. Das Problem bei dieser Herangehensweise ist natürlich, dass wir zu Beginn nicht gesagt haben, was „getrennte Komponenten" sein sollen, auch wenn anschaulich vielleicht klar ist, was gemeint war.

Um diesen Schwierigkeiten aus dem Weg zu gehen, machen wir statt dessen die disjunkte Vereinigung zur Grundlage unserer Definition: Ein metrischer Raum A ist zusammenhängend, wenn er nicht als Vereinigung echter relativ offener Teilmengen dargestellt werden kann. Zusammenhängende Teilmengen sind das Analogon der Intervalle in der eindimensionalen Analysis. Unter anderem es dies genau die Teilmenge $A \subseteq X$ mit der Eigenschaft, dass für stetige Funktionen $f : A \to X$ der Zwischenwertsatz gültig ist.

Anschaulich besser zugänglich ist der Begriff des Wegzusammenhangs. Unter einem *Weg* in A versteht man eine stetige Abbildung $\gamma : [a,b] \to A$. Man verbindet damit die Vorstellung, sich in A zu „bewegen", indem man bei $\gamma(a)$ startet und über die Punkte $\gamma(t)$ mit $a < t < b$ schließlich zum Ziel $\gamma(b)$ genannt. Wegzusammenhang bedeutet nun, dass zwei beliebig gewählte Punkte jeweils durch einen Weg in A verbunden werden können.

Wichtige Begriffe und Sätze in diesem Kapitel

- Die Intervalle sind die zusammenhängenden Teilmengen von $\mathbb{R}$.
- Zwischenwertsatz für stetige Funktionen auf zusammenhängenden Teilmengen
- Definition wegzusammenhängender Mengen
  (Es gilt „wegzusammenhängend ⇒ „zusammenhängend", aber nicht die Umkehrung.)
- konvexe Teilmengen normierter $\mathbb{R}$-Vektorräume als Beispiel für Wegzusammenhang

(6.1) Definition. Sei (X, d_X) ein metrischer Raum. Eine Teilmenge $A \subseteq X$ wird ***zusammenhängend*** genannt, wenn es keine disjunkten, nichtleeren, und in A relativ offenen Mengen $U, V \subseteq A$ mit $A = U \cup V$ gibt.

Wir bezeichnen den metrischen Raum (X, d_X) selbst als *zusammenhängend*, wenn die Teilmenge $A = X$ zusammenhängend ist.

(6.2) Proposition. Eine Teilmenge A eines metrischen Raums (X, d_X) ist genau dann zusammenhängend, wenn $\varnothing$ und A die einzigen Teilmengen von A sind, die sowohl relativ offen als auch relativ abgeschlossen in A sind.

Beweis: Wir beweisen beide Richtungen durch Kontraposition. „⇒" Angenommen, es gibt eine Teilmenge $U \subseteq A$ mit $U \neq \varnothing, A$, die sowohl relativ offen als auch relativ abgeschlossen in A ist. Dann ist auch $V = A \setminus U$ in A relativ offen, und es gilt $V \neq \varnothing$. Somit ist durch $A = U \cup V$ eine Zerlegung von A in disjunkte, nichtleere, in A relativ offene Mengen gegeben.

„⇐“ Sei $A = U \cup V$ eine Zerlegung von A in nichtleere, disjunkte, in A relativ offene Mengen. Neben $U \neq \emptyset$ gilt auch $U \neq A$, dennansonsten wäre $V = \emptyset$. Also ist U eine in A relativ offene Menge ungleich $\emptyset$ und A. □

Die Teilmenge $A = \{(x, \frac{1}{x}) \mid 0 \neq x \in \mathbb{R}\} \subseteq \mathbb{R}^2\}$ ist nicht zusammenhängend. Ist nämlich $\mathbb{R}^-$ die Menge der negativen reellen Zahlen, $U = \mathbb{R}^- \times \mathbb{R}$ und $V = \mathbb{R}^+ \times \mathbb{R}$, dann ist $A = U' \cup V'$ mit $U' = U \cap A$ und $V' = V \cap A$ eine Zerlegung von A in disjunkte, nichtleere und in A relativ offene Teilmengen.

Nach unserer Definition aus der Analysis einer Variablen wird eine Teilmenge $I \subseteq \mathbb{R}$ als ***Intervall*** bezeichnet, wenn mit $a, b \in I$ auch $[a, b]$ in I enthalten ist.

(6.3) Satz. Sei $M \subseteq \mathbb{R}$ eine Menge, die mindestens zwei verschiedene Elemente enthält. Genau dann ist M eine zusammenhängende Teilmenge von $\mathbb{R}$, wenn M ein Intervall ist.

Beweis: „⇐“ Angenommen, M ist ein Intervall. Sei $M = U \cup V$ eine Zerlegung von M in disjunkte, nichtleere und in M relativ offene Teilmengen. Sei $u \in U$ und $v \in V$; aus Symmetriegründen können wir $u < v$ voraussetzen. Weil M ein Intervall ist, gilt $[u, v] \subseteq M$, und $U' = [u, v] \cap U$, $V' = [u, v] \cap V$ ist eine Zerlegung von $[u, v]$ in disjunkte, nichtleere Teilmengen, die beide in $[u, v]$ relativ offen sind. Sei nun $s = \sup U'$. Dann gibt es in U' eine Folge $(s^{(n)})_{n\in\mathbb{N}}$, die gegen s konvergiert. Weil U' als Komplement von V' in $[u, v]$ relativ abgeschlossen ist, liegt s als Grenzwert der Folge in U'. Nach Definition des Supremums gilt $x \notin U'$ und somit $x \in V'$ für alle $x \in [u, v]$ mit $x > s$. Andererseits gibt es auf Grund der relativen Offenheit von U' in $[u, v]$ ein $\varepsilon \in \mathbb{R}^+$, so dass $[s, s + \varepsilon[$ noch in U' enthalten ist. Dies würde bedeuten, dass zum Beispiel $s + \frac{1}{2}\varepsilon$ sowohl in U' als auch in V' liegt, im Widerspruch dazu, dass U' und V' disjunkt sind.

„⇒“ Nehmen wir an, M ist zusammenhängend, aber kein Intervall. Dann gibt es zwei verschiedene Punkte $u, v \in M$ mit $u < v$ und einen Punkt $s \in [u, v]$, der nicht in M liegt. Wir definieren nun $U =]-\infty, s[$ und $V =]s, +\infty[$. Dann ist $M = (U \cap M) \cup (V \cap M)$ eine Zerlegung von M in disjunkte, nichtleere Teilmengen, die in M relativ offen sind. Dies widerspricht der Annahme, dass M zusammenhängend ist. □

(6.4) Proposition. Seien (X, d_X) und (Y, d_Y) metrische Räume und $A \subseteq X$ zusammenhängend. Sei $f : X \to Y$ eine stetige Abbildung. Dann ist $f(A)$ eine zusammenhängende Teilmenge von Y.

Beweis: Angenommen, es gibt nichtleere, disjunkte Teilmengen $U, V \subseteq f(A)$, die in $f(A)$ relativ offen sind. Weil f und damit auch $f|_A$ stetig ist, sind die Urbildmengen $U' = (f|_A)^{-1}(U)$ und $V' = (f|_A)^{-1}(V)$ relativ offen in A, nach (4.10). Weil U und V beide nichtleer sind (und $f|_A$ die Menge A surjektiv auf $f(A) = U \cup V$ abbildet), sind auch U' und V' nichtleer. Wären U' und V' nicht disjunkt, $x \in U' \cap V'$, dann hätten U und V mit $f(x)$ ebenfalls einen gemeinsamen Punkt, im Widerspruch zur Voraussetzung. Außerdem gilt $U \cup V = A$, denn für jedes $x \in A$ gilt $f(x) \in f(A)$, also $f(x) \in U$ oder $f(x) \in V$ und damit $x \in U'$ oder $x \in V'$. Insgesamt haben wir also A in nichtleere disjunkte, in A relativ offene Teilmengen zerlegt. Aber dies widerspricht der Voraussetzung, dass A zusammenhängend ist. □

(6.5) Satz. *(Zwischenwertsatz)*

Sei (X, d_X) ein metrischer Raum, $A \subseteq X$ eine zusammenhängende Teilmenge und $f : X \to \mathbb{R}$ eine stetige Funktion. Seien $a, b \in A$ vorgegeben. Dann nimmt f auf A jeden Wert $c \in \mathbb{R}$ mit $f(a) \leq c \leq f(b)$ an.

Beweis: Nach (6.4) ist $f(A) \subseteq \mathbb{R}$ zusammenhängend. Ist $f(A)$ die leere Menge oder besteht $f(A)$ nur aus einem Element, dann ist die Aussage offenbar erfüllt. Andernfalls ist $f(A)$ nach (6.3) ein Intervall. Dies bedeutet, dass mit $f(a)$ und $f(b)$ auch jeder Wert dazwischen in $f(A)$ enthalten ist. □

(6.6) Definition. Sei (X, d_X) ein metrischer Raum. Eine Teilmenge $A \subseteq X$ heißt ***wegzusammenhängend***, wenn für beliebige Punkte $a, b \in A$ jeweils eine stetige Abbildung $\gamma : [0,1] \to A$ mit $\gamma(0) = a$ und $\gamma(1) = b$ existiert. Man bezeichnet γ als ***Weg***, der die Punkte a und b verbindet.

Ein wichtiger Spezialfall für wegzusammenhängende Mengen sind die konvexen Teilmengen von normierten $\mathbb{R}$-Vektorräumen.

(6.7) Definition. Sei V ein normierter $\mathbb{R}$-Vektorraum, und seien $p, q \in V$. Dann ist die ***Verbindungsstrecke*** zwischen p und q die Menge $[p,q] = \{(1-t)p + tq \mid t \in [0,1]\}$. Eine Teilmenge $A \subseteq V$ wird ***konvex*** genannt, wenn für alle $p, q \in A$ jeweils $[p,q] \subseteq A$ gilt.

Einfache Beispiele konvexer Teilmengen sind die offenen und abgeschlossenen Bälle in einem metrischen Raum.

(6.8) Proposition. Sei V ein normierter $\mathbb{R}$-Vektorraum. Dann ist jeder offene und jeder abgeschlossene Ball in V konvex.

Beweis: Wir beschränken uns auf den offenen Fall. Sei $a \in V$, $r \in \mathbb{R}^+$ und $A = B_r(a) = \{x \in V \mid \|x - a\| < r\}$. Seien außerdem $p, q \in B_r(a)$ vorgegeben; zu zeigen ist $[p,q] \subseteq A$. Wegen $p, q \in B_r(a)$ gilt $\|p - a\| < r$ und $\|q - a\| < r$. Sei nun $z \in [p,q]$. Dann gibt es nach Definition der Verbindungsstrecke ein $t \in [0,1]$ mit $z = (1-t)p + tq$. Aus der Dreiecksungleichung folgt

$$\begin{aligned} \|z - a\| &= \|(1-t)p + tq - a\| = \|(1-t)(p-a) + t(q-a)\| \leq (1-t)\|p-a\| + t\|q-a\| \\ &< (1-t)r + tr = r. \end{aligned}$$

Also ist z in $B_r(a)$ enthalten. □

Jede konvexe Teilmenge A in einem $\mathbb{R}$-Vektorraum V ist wegzusammenhängend: Seien $p, q \in A$ beliebig vorgegebene Punkte. Weil die Verbindungsstrecke $[p,q]$ vollständig in A liegt, können wir durch $\gamma : [0,1] \to A$, $t \mapsto (1-t)p + tq$ einen in A verlaufenden Weg definieren, der die Punkte p und q verbindet.

(6.9) Proposition. Jeder wegzusammenhängende Teilmenge $A \subseteq X$ eines metrischen Raums (X, d_X) ist zusammenhängend.

Beweis: Angenommen, $A = U \cup V$ ist eine Zerlegung von A in nichtleere, disjunkte, in A relativ offene Mengen U, V. Dann wählen wir Punkte $p \in U$, $q \in V$ und verbinden diese durch einen Weg $\gamma : [0,1] \to A$. Nun gilt $\gamma^{-1}(U) \cup \gamma^{-1}(V) = [0,1]$, weil $\gamma(t)$ für jedes $t \in [0,1]$ in U oder V liegt, und dies ist eine Zerlegung des Intervalls $[0,1]$ in nichtleere und disjunkte Mengen: disjunkt, weil die Mengen U und V disjunkt sind, und nichtleer wegen $0 \in \gamma^{-1}(U)$ und $1 \in \gamma^{-1}(V)$. Weil γ stetig ist, sind die Teilmengen $\gamma^{-1}(U)$ und $\gamma^{-1}(V)$ nach (4.10) offen in $[0,1]$. Weil aber $[0,1]$ als metrischer Raum nach (6.3) zusammenhängend ist, kann es eine solche Zerlegung nicht geben. □

Ergänzung:
Beispiel für eine zusammenhängende, nicht wegzusammenhängende Menge

Bereits in der Analysis einer Variablen war uns die Funktion $f : \mathbb{R} \to \mathbb{R}$ gegeben durch

$$f(x) \quad = \quad \begin{cases} \sin(\frac{1}{x}) & \text{für } x \neq 0 \\ 0 & \text{für } x = 0 \end{cases}$$

begegnet, als Beispiel für eine (im Punkt 0) unstetige Funktion, bei der sich die Unstetigkeit aber nicht am Funktionsgraphen ablesen lässt. Hier zeigen wir, dass der Funktionsgraph von f, also die Menge $A = \Gamma_f = \{(x, f(x)) \mid x \in \mathbb{R}\}$ eine zusammenängende, aber nicht wegzusammenhängende Teilmenge von $\mathbb{R}^2$ ist.

(1) $A = \Gamma_f$ ist zusammenhängend
Nehmen wir an, die Menge A ist nicht zusammenhängend. Dann gibt es eine disjunkte Zerlegung $A = U \cup V$ in nichtleere Teilmengen U, V, die beide in A relativ offen und damit auch relativ abgeschlossen sind. Wir bemerken nun zunächst, dass die Mengen $A^+ = \{(x, f(x)) \mid x \in \mathbb{R}^+\}$ und $A^- = \{(x, f(x)) \mid x \in \mathbb{R}, x < 0\}$ wegzusammenhängend sind. Sind nämlich $p = (y, f(y))$ und $q = (z, f(z))$ in A^+ vorgegeben mit $0 < y < z$, dann ist auf Grund der Stetigkeit von $f|_{\mathbb{R}^+}$ durch $\gamma(t) = ((1-t)y + tz, f((1-t)y + tz))$ eine stetige Abbildung $\gamma : [0,1] \to A^+$ mit $\gamma(0) = (y, f(y)) = p$ und $\gamma(1) = (z, f(z)) = q$ definiert. Genauso zeigt man, dass A^- wegzusammenhängend ist.

Daraus folgt nun, dass sowohl A^+ als auch A^- entweder vollständig in U oder vollständig in V liegt. Würde nämlich zum Beispiel weder $A^+ \subseteq U$ noch $A^+ \subseteq V$ gelten, dann wäre $A^+ = (A^+ \cap U) \cup (A^+ \cap V)$ eine Zerlegung von A^+ in relativ offene, nichtleere, disjunkte Teilmengen, im Widerspruch dazu, dass A^+ wegzusammenhängend und nach (6.9) damit auch zusammenhängend ist. Nach eventueller Vertauschung von U und V können wir also $A^- \subseteq U$ und $A^+ \subseteq V$ annehmen. Weil nun U in A auch relativ abgeschlossen ist und die Folge gegeben durch $x^{(n)} = (\frac{1}{2\pi n}, \sin(2\pi n)) = (\frac{1}{2\pi n}, 0)$ vollständig in V liegt, muss nach (4.16) auch der Grenzwert $(0,0) \in A$ dieser Folge in U liegen. Mit Hilfe der Folge $y^{(n)} = (-\frac{1}{2\pi n}, \sin(-2\pi n)) = (-\frac{1}{2\pi n}, 0)$ kann genauso begründet werden, dass $(0,0)$ in V liegt. Aber $(0,0) \in U \cap V$ widerspricht der Voraussetzung, dass U und V disjunkt sind.

(2) $A = \Gamma_f$ ist nicht wegzusammenhängend
Nehmen wir an, dass A wegzusammenhängend ist, und setzen wir $p = (0, f(0)) = (0,0)$ und $q = (1, f(1))$. Dann gibt es eine stetige Abbildung $\gamma : [0,1] \to A$ mit $\gamma(0) = p$ und $\gamma(1) = q$. Nach (3.6) ist dann auch die erste Komponente γ_1 von γ stetig, und wegen $\gamma_1(0) = 0$ und $\gamma_1(1) = 1$ muss es nach dem Zwischenwertsatz für jedes $n \in \mathbb{N}$ ein $t^{(n)} \in]0,1[$ mit $\gamma_1(t^{(n)}) = \frac{1}{2\pi(n+\frac{1}{4})}$ geben. Die Folge $(t^{(n)})_{n\in\mathbb{N}}$ ist beschränkt und besitzt nach dem Satz von Bolzano-Weierstrass eine in $[0,1]$ konvergente Teifolge $(t^{(n_k)})_{k\in\mathbb{N}}$, deren Grenzwert wir mit t_0 bezeichnen. Weil γ stetig ist und $\gamma(t)$ für jedes $t \in [0,1]$ in A liegt und somit $\gamma(t) = (\gamma_1(t), f(\gamma_1(t)))$ gilt, folgt nun

$$\begin{aligned} (\gamma_1(t_0), f(\gamma_1(t_0))) \quad &= \quad \gamma(t_0) \quad = \quad \lim_{k\to\infty} \gamma(t^{(n_k)}) \quad = \quad \lim_{k\to\infty} (\gamma_1(t^{(n_k)}), f(\gamma_1(t^{(n_k)}))) \\ = \quad \lim_{k\to\infty} (\tfrac{1}{2\pi(n_k+\frac{1}{4})}, f(\tfrac{1}{2\pi(n_k+\frac{1}{4})})) \quad &= \quad \lim_{k\to\infty} (\tfrac{1}{2\pi(n_k+\frac{1}{4})}, \sin(2\pi(n_k+\tfrac{1}{4}))) \quad = \quad (0,1). \end{aligned}$$

Durch Vergleich der beiden Komponenten erhalten wir $\gamma_1(t_0) = 0$ und $f(\gamma_1(t_0)) = 1$, was aber zu $f(\gamma_1(t_0)) = f(0) = 0$ im Widerspruch steht. Dies zeigt, dass eine solche Abbildung γ nicht existiert. Also ist A nicht wegzusammenhängend.

§ 7. Partielle Ableitungen und Richtungsableitungen

Zusammenfassung. In der eindimensionalen Analysis ist die Ableitung einer Funktion in einem Punkt lediglich eine reelle Zahl, die die Steigung der Funktion in diesem Punkt angibt. Bei einer mehrdimensionalen Funktion ist eine einzelne Zahl zur Beschreibung des Steigungsverhaltens nicht mehr ausreichend, weil die Steigung in der Regel davon abhängt, in welche Richtung man sich innerhalb des Definitionsbereichs bewegt. Zum Beispiel wächst der Funktionswert von $f(x,y) = x + 2y$ in y-Richtung doppelt so schnell wie in x-Richtung.

Man kann aber der Funktion in einem Punkt p für jeden Richtungsvektor v einen Steigungswert zuordnen. Dies geschieht durch die *Richtungsableitung* $\partial_v f(p)$. Ist v einer der Einheitsvektoren im $\mathbb{R}^n$, dann spricht man auch von einer *partiellen Ableitung*. Neben der Definition dieser Ableitungsarten behandeln wir in diesem Kapitel auch höhere (mehrfache) partielle Ableitungen und verallgemeinern den Mittelwertsatz aus der Analysis einer Variablen auf diesen neuen Ableitungstyp.

Wichtige Begriffe und Sätze in diesem Kapitel

- Definition der partiellen Ableitungen und Richtungsableitungen
- Mittelwertsatz für Richtungsableitungen
- höhere partielle Ableitungen
- Vertauschbarkeit der partiellen Ableitungen (Satz von Schwarz)

Im gesamten Kapitel sei V ein endlich-dimensionaler normierter $\mathbb{R}$-Vektorraum. Sei $U \subseteq V$ eine offene Teilmenge und $f : U \to \mathbb{R}$ eine Funktion auf U. Sei außerdem $a \in U$ ein beliebiger Punkt des Definitionsbereichs und $v \in V$. Unser Ziel besteht darin, das Steigungsverhalten von f im Punkt a in Richtung des Vektors v zu untersuchen.

Sei dazu $\phi : \mathbb{R} \to V$ definiert durch $\phi(t) = a + tv$. Weil U eine offene Teilmenge ist, die den Punkt $\phi(0) = a$ enthält, und auf Grund der Stetigkeit von ϕ, ist $\phi^{-1}(U) \subseteq \mathbb{R}$ nach Satz (4.9) eine Umgebung von 0. Es gibt also ein $\varepsilon \in \mathbb{R}^+$ mit $]-\varepsilon, \varepsilon[\subseteq \phi^{-1}(U)$, und folglich ist die $\mathbb{R}$-wertige Funktion $f \circ \phi$ zumindest auf dem Intervall $]-\varepsilon, \varepsilon[$ definiert. Es kann somit von der Differenzierbarkeit der Funktion $f \circ \phi$ im Punkt 0 gesprochen werden.

(7.1) Definition. Ist die Funktion $f \circ \phi$ im Punkt 0 differenzierbar, dann nennt man die Ableitung $\partial_v f(a) = (f \circ \phi)'(0)$ die ***Richtungsableitung*** von f im Punkt a in Richtung v.

Die Abbildung ϕ dient bei dieser Definition also lediglich dazu, aus der mehrdimensionalen Funktion f eine eindimensionale Funktion zu machen, was uns ermöglicht, auf den Differenzierbarkeitsbegriff der Analysis einer Variablen zurückzugreifen. Man kann mit dieser Hilfsfunktion aber nicht nur die Richtungsableitung $\partial_v f$ an der Stelle a, sondern an jedem Punkt der Form $\phi(t) = a + tv$ berechnen, der im Definitionsbereich U von f enthalten ist.

(7.2) Lemma. Für alle $t \in \mathbb{R}$ mit $\phi(t) \in U$ existiert $\partial_v f(a + tv)$ genau dann, wenn $f \circ \phi$ an der Stelle t differenzierbar ist, und in diesem Fall gilt $\partial_v f(a + tv) = (f \circ \phi)'(t)$.

Beweis: Sei $t_0 \in \mathbb{R}$ ein Punkt mit $\phi(t_0) \in U$, und sei $\psi : \mathbb{R} \to V$ definiert durch $\psi(t) = a + t_0 v + tv = a + (t_0 + t)v$. Nach Definition existiert die Richtungsableitung $\partial_v f(a + t_0 v)$ genau dann, wenn die Funktion $f \circ \psi$ an der Stelle 0 differenzierbar ist, und in diesem Fall gilt $\partial_v f(a + t_0 v) = (f \circ \psi)'(0)$. Sei nun $\tau : \mathbb{R} \to \mathbb{R}$ gegeben durch $\tau(t) =$

$t_0 + t$. Dann hängen die beiden Abbildungen ϕ und ψ durch $\psi = \phi \circ \tau$ zusammen. Auf Grund der Kettenregel der eindimensionalen Analysis folgt aus der Differenzierbarkeit von τ an der Stelle 0 und der Differenzierbarkeit von $f \circ \phi$ an der Stelle $t_0 = \tau(0)$ die Differenzierbarkeit von $f \circ \phi \circ \tau = f \circ \psi$ an der Stelle 0, und es gilt dann

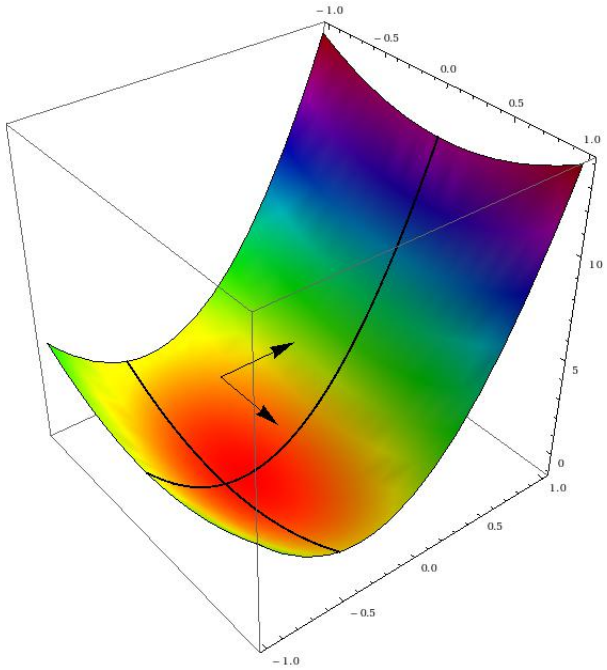
$$\begin{aligned} \partial_v f(a + t_0 v) &= (f \circ \psi)'(0) = (f \circ \phi \circ \tau)'(0) = (f \circ \phi)'(\tau(0)) \cdot \tau'(0) \\ &= (f \circ \phi)'(t_0) \cdot 1 = (f \circ \phi)'(t_0). \end{aligned}$$

Umgekehrt folgt aus der Existenz von $\partial_v f(a + t_0 v)$ nach Definition die Differenzierbarkeit von $f \circ \psi$ an der Stelle 0. Wegen $\phi = \psi \circ \tau^{-1}$ folgt daraus wiederum die Differenzierbarkeit von $f \circ \psi \circ \tau^{-1} = f \circ \phi$ an der Stelle t_0. □

Für konkrete Anwendungen ist vor allem der Fall $V = \mathbb{R}^n$ interessant. Hier arbeitet man vor allem mit den Richtungsableitungen bezüglich der Einheitsvektoren $e_1, ..., e_n$.

(7.3) Definition. Sei $U \subseteq \mathbb{R}^n$, $a \in U$ und $k \in \{1, ..., n\}$. Die Richtungsableitung von f im Punkt a bezüglich des k-ten Einheitsvektors e_k wird (sofern sie existiert) die ***k-te partielle Ableitung*** $\partial_k f(a)$ von f im Punkt a genannt.

Wird f in Abhängigkeit von Variablen x, y, ... dargestellt, zum Beispiel in der Form $f(x, y) = x^2 + 2xy + y^2$, dann verwendet man an Stelle von $\partial_1 f$ und $\partial_2 f$ auch die Bezeichnungen $\partial f / \partial x$ und $\partial f / \partial y$ für die partiellen Ableitungen.



Die partiellen Ableitungen geben die Steigung einer Funktion in x- und y-Richtung an.

Ist $a = (a_1, ..., a_n) \in U$ ein Punkt des Definitionsbereichs, und ist die Abbildung $\phi : \mathbb{R} \to \mathbb{R}^n$ definiert durch $\phi(t) = (a_1, ..., a_{k-1}, t, a_{k+1}, ..., a_n)$, dann ist $\partial_k f(a)$ nach (7.2) die Ableitung von $(f \circ \phi) = f(a_1, ..., a_{k-1}, t, a_{k+1}, ..., a_n)$ im Punkt a_k, sofern die Ableitung dort existiert. Man $\partial_k f(a)$ also berechnen, indem man f nur als Funktion der k-ten

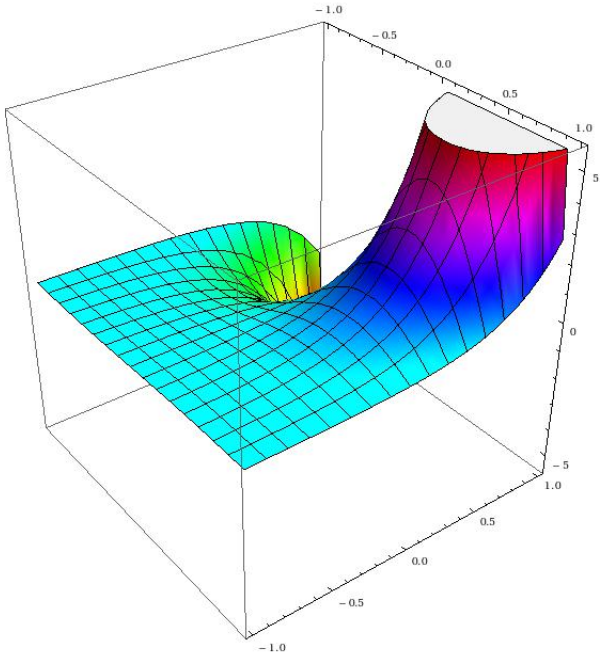
Koordinate betrachtet, während die übrigen Koordinaten auf die konstanten Werte $a_1, ..., a_{k-1}, a_{k+1}, ..., a_n$ gesetzt werden, und diese Funktion dann an der Stelle a_k ableitet. Ist f beispielsweise eine Funktion auf dem $\mathbb{R}^2$, dann ist $\partial_1 f(a_1, a_2)$ die Ableitung von $t \mapsto f(t, a_2)$ an der Stelle a_1, und $\partial_2 f(a_1, a_2)$ ist die Ableitung von $t \mapsto f(a_1, t)$ an der Stelle a_2. Wir betrachten nun einige konkrete Beispiele.

Beispiel 1: Sei $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch $f(x, y) = 2x^2 + 7y^2 + 5y$ und $p = (x, y) \in \mathbb{R}^2$ ein beliebig gewählter Punkt. Um $\partial_1 f(p)$ zu berechnen, müssen wir die Ableitung der Funktion $t \mapsto f(t, y)$ im Punkt $t = x$ bestimmen. Es gilt

$$f(t, y) \quad = \quad 2t^2 + 7y^2 + 5y$$

für alle $t \in \mathbb{R}$, und die Ableitung dieser Funktion ist $t \mapsto 4t$. Also ist $\partial_1 f(x, y) = 4x$. Zur Berechnung von $\partial_2 f(x, y)$ betrachten wir die Funktion $t \mapsto f(x, t)$. Es gilt $f(x, t) = 2x^2 + 7t^2 + 5t$, und die Ableitung ist $t \mapsto 14t + 5$. Wir erhalten somit $\partial_2 f(x, y) = 14y + 5$.

Beispiel 2: Sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch $f(x, y) = \sin(2x)e^{3y}$. Dann sind die partiellen Ableitungen von f gegeben durch $\partial_1 f(x, y) = 2\cos(2x)e^{3y}$ und $\partial_2 f(x, y) = 3\sin(2x)e^{3y}$.



Graph der Funktion aus Beispiel 2

Auch die Berechnung von Richtungsableitungen sehen wir uns an einem Beispiel an.

Beispiel 3: Sei $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch $f(x, y) = 2x^2 + 7y^2 + 5y$ für alle $(x, y) \in \mathbb{R}^2$. Wir bestimmen die Ableitung von f im Punkt (x, y) in Richtung des Vektors $v = (1, 1)$. Diese ist gegeben durch die Ableitung der Funktion $f \circ \phi$ im Punkt 0, mit der Hilfsfunktion $\phi : \mathbb{R} \to \mathbb{R}^2$, $t \mapsto (x, y) + tv$. Dabei ist $(x, y) + tv = (x + t, y + t)$.

Den Wert der Ableitung erhält man nun durch die Rechnungen

$$\begin{aligned}(f \circ \phi)(t) \quad &= \quad f(x + t, y + t) \quad = \quad 2(x + t)^2 + 7(y + t)^2 + 5(y + t) \\ &= \quad 2x^2 + 4xt + 2t^2 + 7y^2 + 14yt + 7t^2 + 5y + 5t\end{aligned}$$

und $(f \circ \phi)'(t) = 4x + 4t + 14y + 14t + 5$ für alle $t \in \mathbb{R}$. Wir erhalten $\partial_{(1,1)}(x,y) = (f \circ \phi)'(0) = 4x + 14y + 5$. Ein Vergleich mit dem Ergebnis von oben zeigt, dass es sich um die Summe $\partial_1 f(x,y) + \partial_2 f(x,y)$ der beiden partiellen Ableitungen handelt.

Wir wiederholen die Rechnung noch einmal mit einem beliebigen Richtungsvektor $v = (a,b) \in \mathbb{R}^2$. In diesem Fall müssen wir die Ableitung der Funktion $(f \circ \phi)(t) = f(x + ta, y + tb) = 2(x + ta)^2 + 7(y + tb)^2 + 5(y + tb)$ an der Stelle 0 bestimmen. Es gilt $(f \circ \phi)'(t) = 4a(x + ta) + 14b(y + tb) + 5b$ und somit

$$\partial_v f(x,y) \quad = \quad (f \circ \phi)'(0) \quad = \quad 4ax + 14by + 5b \quad = \quad a \cdot \partial_1 f(x,y) + b \cdot \partial_2 f(x,y).$$

Wir leiten nun einige allgemeine Regeln für das Rechnen mit Richtungsableitungen her.

(7.4) Lemma. Sei $U \subseteq V$ offen. Sei $v \in V$, und seien $f, g : U \to \mathbb{R}$ reellwertige Funktionen.

(i) Ist f konstant, dann gilt $\partial_v f(x) = 0$ für alle $x \in U$.

(ii) Ist $x \in U$ ein Punkt mit der Eigenschaft, dass die Richtungsableitungen $\partial_v f(x)$ und $\partial_v g(x)$ existieren, dann existiert auch $\partial_v(f+g)(x)$ und $\partial_v(fg)(x)$, und es gilt

$$\partial_v(f+g)(x) = \partial_v f(x) + \partial_v g(x) \qquad \text{und} \qquad \partial_v(fg)(x) = f(x)\partial_v g(x) + g(x)\partial_v f(x).$$

(iii) Existiert $\partial_v f(x)$ im Punkt $x \in U$, dann existiert auch $\partial_{cv} f(x)$ für alle $c \in \mathbb{R}$, und es gilt $\partial_{cv} f = c\, \partial_v f$.

Beweis: Sei $x \in U$, und sei $\varepsilon \in \mathbb{R}^+$ hinreichend klein gewählt, so dass $x + tv \in U$ für alle $t \in]{-}\varepsilon, \varepsilon[$ gilt. Sei außerdem $\phi : \mathbb{R} \to V$ definiert durch $\phi(t) = x + tv$ für alle $t \in \mathbb{R}$. Nach Definition der Richtungsableitung gilt $\partial_v f(x) = (f \circ \phi)'(0)$ und $\partial_v g(x) = (g \circ \phi)'(0)$.

zu (i) Ist f konstant, dann ist auch $f \circ \phi$ konstant, und es folgt $\partial_v f(x) = \phi'(0) = 0$.

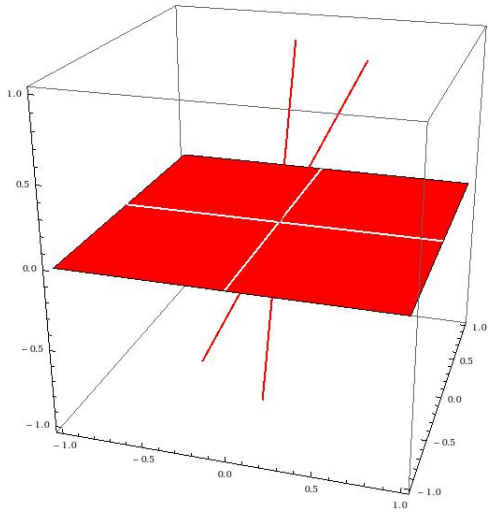
zu (ii) Nach Definition der punktweisen Summe zweier Funktionen gilt $(f+g) \circ \phi = (f \circ \phi) + (g \circ \phi)$. Mit der Summenregel erhalten wir $\partial_v(f+g)(x) = ((f \circ \phi) + (g \circ \phi))'(0) = (f \circ \phi)'(0) + (g \circ \phi)'(0) = \partial_v f(x) + \partial_v g(x)$. Ebenso gilt $(fg) \circ \phi = (f \circ \phi)(g \circ \phi)$. Die Produktregel liefert somit

$$\begin{aligned} \partial_v(fg) \quad &= \quad ((f \circ \phi)(g \circ \phi))'(0) \quad = \quad (f \circ \phi)'(0)(g \circ \phi)(0) + (f \circ \phi)(0)(g \circ \phi)'(0) \\ &= \quad \partial_v f(x) \cdot g(x) + f(x) \cdot \partial_v g(x). \end{aligned}$$

zu (iii) Sei $\phi_c : \mathbb{R} \to V$ definiert durch $\phi_c(t) = x + ctv$ und $\varepsilon_c \in \mathbb{R}^+$ so gewählt, dass $\phi_c(]{-}\varepsilon_c, \varepsilon_c[) \subseteq U$ gilt. Wegen $\phi_c(t) = \phi(ct)$ gilt $\phi_c = \phi \circ \pi_c$ mit $\pi_c : \mathbb{R} \to \mathbb{R}$ gegeben durch $\pi_c(t) = ct$ für alle $t \in \mathbb{R}$. Nach Definition der Richtungsableitung gilt $\partial_{cv} f(x) = (f \circ \phi_c)(t)$. Die Kettenregel aus der Analysis einer Variablen liefert nun

$$\partial_{cv} f(x) \quad = \quad (f \circ \phi_c)'(0) \quad = \quad ((f \circ \phi) \circ \pi_c)'(0) \quad = \quad (f \circ \phi)'(\pi_c(0))\pi_c'(0) \quad = \quad \partial_v f(x) \cdot c. \qquad \square$$

Das letzte Beispiel wirft die Frage auf, ob sich allgemein jede Richtungsableitung als Linearkombination der partiellen Ableitungen darstellen lässt. Wir werden im nächsten Abschnitt untersuchen, welche Bedingung die Funktion f erfüllen muss, damit dies der Fall ist. Bei einer beliebigen Funktion ist es jedenfalls möglich, dass die partiellen Ableitungen sehr wenig mit den sonstigen Richtungsableitungen zu tun haben, wie das folgende Beispiel zeigt.



Graph der Funktion aus Beispiel 4

Beispiel 4: Sei die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch

$$f(x,y) \quad = \quad \begin{cases} 2x & \text{für } y = 0 \\ 3y & \text{für } x = 0 \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt $\partial_1 f(0,0) = 2$, $\partial_2 f(0,0) = 3$, aber für $v \notin \text{lin}(e_1) \cup \text{lin}(e_2)$ ist $\partial_v f(0,0) = 0$. Man beachte, dass f im Nullpunkt stetig, aber in allen übrigen Punkten der x- und der y-Achse unstetig ist.

Die Existenz der Richtungsableitungen in einem Punkt a sagt im allgemeinen auch wenig über das Verhalten der Funktion in diesem Punkt aus. Es kann beispielsweise vorkommen, dass sämtliche Richtungsableitungen in einem Punkt a existieren, ohne das die Funktion im Punkt a stetig ist. Wir betrachten dazu das folgende Beispiel.

Beispiel 5: Sei die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch

$$f(x,y) \quad = \quad \begin{cases} \dfrac{xy^2}{x^2+y^6} & \text{für } (x,y) \neq (0,0) \\ 0 & \text{für } (x,y) = (0,0). \end{cases}$$

Dann existiert $\partial_v f(0,0)$ für jedes $v \in \mathbb{R}^2$, aber f ist in $(0,0)$ unstetig.

Beweis: Zum Nachweis der ersten Aussage sei $v = (a,b) \in \mathbb{R}^2$ beliebig vorgegeben. Zur Berechnung von $\partial_v f(0,0)$ verwenden wir die Hilfsfunktion $\phi(t) = (0,0) + tv = (ta, tb)$. Im Fall $a \neq 0$ gilt für $t \neq 0$ jeweils

$$(f \circ \phi)(t) \quad = \quad f(ta,tb) \quad = \quad \frac{(ta)(tb)^2}{(ta)^2+(tb)^6} \quad = \quad \frac{t^3ab^2}{t^2a^2+t^6b^6} \quad = \quad \frac{tab^2}{a^2+t^4b^6} \quad ,$$

und diese Gleichung ist auch für $t = 0$ gültig, da $(f \circ \phi)(0) = f(0,0) = 0$ gilt. Wir bilden nun die Ableitung

$$(f \circ \phi)'(t) \quad = \quad \frac{(ab^2) \cdot (a^2+t^4b^6) - (tab^2) \cdot (4t^3b^6)}{(a^2+t^4b^6)^2} \quad = \quad \frac{a^3b^2+t^4ab^8-4t^4ab^8}{(a^2+t^4b^6)^2} \quad = \quad \frac{a^3b^2-3t^4ab^8}{(a^2+t^4b^6)^2}$$

und erhalten $\partial_v f(0,0) = (f \circ \phi)'(0) = a^{-1}b^2$. Betrachten wir nun den Fall $a = 0$. Im Fall $b \neq 0$ gilt für $t \neq 0$ die Gleichung

$$(f \circ \phi)(t) \quad = \quad f(0, tb) \quad = \quad \frac{0 \cdot (tb)^2}{0^2 + (tb)^6} \quad = \quad 0$$

und ebenso $(f \circ \phi)(0) = f(0,0) = 0$. Also gilt in diesem Fall $\partial_v f(0,0) = 0$. Für $v = (0,0)$ erhalten wir schließlich $\phi(t) = (0,0)$ für alle $t \in \mathbb{R}$, also $(f \circ \phi)(t) = 0$ und damit ebenso $\partial_v f(0,0) = 0$. Damit ist gezeigt, dass $\partial_v f(0,0)$ für alle $v \in \mathbb{R}^2$ existiert. Zum Nachweis der Unstetigkeit in $(0,0)$ betrachten wir die Folge $((x_n, y_n))_{n\in\mathbb{N}}$ gegeben durch $(x_n, y_n) = (\frac{1}{n^3}, \frac{1}{n})$, die gegen $(0,0)$ konvergiert. Es gilt

$$f(x_n, y_n) \quad = \quad \frac{\frac{1}{n^5}}{\frac{1}{n^6} + \frac{1}{n^6}} \quad = \quad \tfrac{1}{2}n.$$

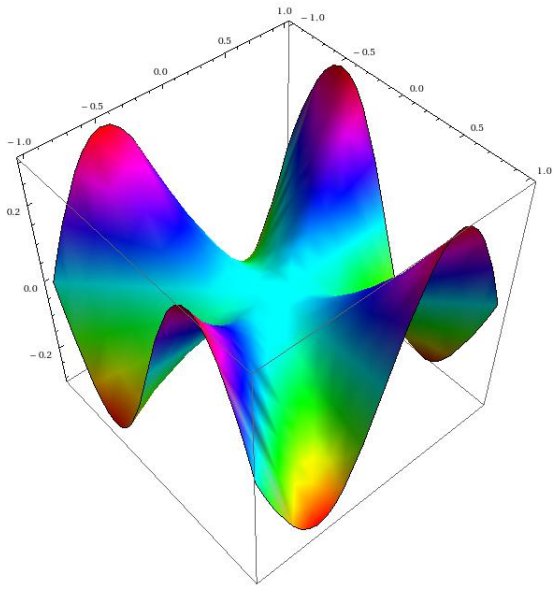
Die Folge der Funktionswerte $f(x_n, y_n)$ konvergiert nicht gegen $f(0,0) = 0$, also ist f in $(0,0)$ unstetig.

In vielen Anwendungen werden Funktionen auch mehrfach partiell abgeleitet. Man bezeichnet eine Funktion $f : U \to \mathbb{R}$ auf einer offenen Teilmenge $U \subseteq \mathbb{R}^n$ als ***stetig partiell differenzierbar***, wenn die partiellen Ableitungen $\partial_i f$ für $1 \leq i \leq n$ auf U existieren und stetig sind.

Falls die Funktionen $\partial_i f$ ihrerseits alle partiell differenzierbar sind, spricht man von einer ***zweifach partiell differenzierbaren*** Funktion. Sind auch die Funktionen $\partial_i \partial_j f$ für $1 \leq i, j \leq n$ wieder stetig, nennt man f zweifach stetig partiell differenzierbar. Auf naheliegende Weise definiert man m-fach (stetig) partiell differenzierbar für beliebige $m \geq 3$. An Stelle von $\partial_{i_1} \ldots \partial_{i_m} f$ verwendet man zur Abkürzung auch die Schreibweise $\partial_{i_1 \ldots i_m} f$ für die höheren partiellen Ableitungen.

Beispiel 6: Wir betrachten wieder die Funktion $f(x,y) = \sin(2x)e^{3y}$ mit den beiden partiellen Ableitungen $\partial_1 f(x,y) = 2\cos(2x)e^{3y}$ und $\partial_2 f(x,y) = 3\sin(2x)e^{3y}$. Beide sind erneut partiell differenzierbar. Es gilt $\partial_{11} f(x,y) = -4\sin(2x)e^{3y}$, $\partial_{12} f(x,y) = \partial_{21} f(x,y) = 6\cos(2x)e^{3y}$ und $\partial_{22} = 9\sin(2x)e^{3y}$.

Das letzte Beispiel scheint darauf hinzudeuten, dass partielle Ableitungen miteinander vertauschbar sind, dass also $\partial_{ij} f = \partial_{ji} f$ für $1 \leq i, j \leq n$ gilt. Ohne stärkere Voraussetzungen an die Funktion f als die zweifache partielle Differenzierbarkeit braucht dies aber nicht zu gelten, wie das folgende Beispiel zeigt.



Graph der Funktion aus Beispiel 7

Beispiel 7: Sei die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch

$$f(x,y) \quad = \quad \begin{cases} \dfrac{x^3y - xy^3}{x^2+y^2} & \text{für } (x,y) \neq (0,0) \\ 0 & \text{für } (x,y) = (0,0) \end{cases}$$

Dann gilt $\partial_{12} f(0,0) \neq \partial_{21} f(0,0)$.

Beweis: Zunächst berechnen wir die partiellen Ableitungen $\partial_1 f$ und $\partial_2 f$ in den Punkten $(x,y) \neq (0,0)$. Weil die Funktion in der Umgebung eines solchen Punktes als Quotient zweier Polynome in x und y gegeben ist, können wir hier die gewöhnlichen Ableitungsregeln verwenden, um $\partial_1 f$ und $\partial_2 f$ zu bestimmen, wobei wir jeweils eine der Unbekannten als variabel und die andere als Konstante betrachten. Wir erhalten so

$$\partial_1 f(x,y) \quad = \quad \frac{(3x^2y - y^3)\cdot(x^2+y^2) - (x^3y - xy^3)\cdot(2x)}{(x^2+y^2)^2} \quad =$$

$$\frac{3x^4y - x^2y^3 + 3x^2y^3 - y^5 - 2x^4y + 2x^2y^3}{(x^2+y^2)^2} \quad = \quad \frac{x^4y + 4x^2y^3 - y^5}{(x^2+y^2)^2}.$$

Mit einer analogen Rechnung bestimmen wir $\partial_2 f(x,y)$ für $(x,y) \neq (0,0)$.

$$\partial_2 f(x,y) \quad = \quad \frac{(x^3 - 3xy^2)(x^2+y^2) - (x^3y - xy^3)\cdot(2y)}{(x^2+y^2)^2} \quad =$$

$$\frac{x^5 - 3x^3y^2 + x^3y^2 - 3xy^4 - 2x^3y^2 + 2xy^4}{(x^2+y^2)^2} \quad = \quad \frac{x^5 - 4x^3y^2 - xy^4}{(x^2+y^2)^2}.$$

Seien die Funktionen $\phi_1, \phi_2 : \mathbb{R} \to \mathbb{R}^2$ gegeben durch $\phi_1(t) = (t,0)$ und $\phi_2(t) = (0,t)$ für alle $t \in \mathbb{R}$. Weil die Funktion $f \circ \phi_1$ konstant Null ist, gilt $\partial_1 f(0,0) = (f \circ \phi_1)'(0) = 0$. Ebenso ist $f \circ \phi_2$ konstant Null, und wir erhalten für die partielle Ableitung nach y entsprechend $\partial_2 f(0,0) = (f \circ \phi_2)'(0) = 0$. Insgesamt gilt also

$$\partial_1 f(x,y) \quad = \quad \begin{cases} \dfrac{x^4y + 4x^2y^3 - y^5}{(x^2+y^2)^2} & \text{für } (x,y) \neq (0,0) \\ 0 & \text{für } (x,y) = (0,0) \end{cases}$$

und

$$\partial_2 f(x,y) \quad = \quad \begin{cases} \dfrac{x^5 - 4x^3y^2 - xy^4}{(x^2+y^2)^2} & \text{für } (x,y) \neq (0,0) \\ 0 & \text{für } (x,y) = (0,0) \end{cases}$$

Um nun $\partial_{21} f(0,0)$ zu bestimmen, betrachten wir die Funktion

$$(\partial_1 f \circ \phi_2)(t) \quad = \quad \partial_1 f(0,t) \quad = \quad \frac{(-t^5)}{t^4} \quad = \quad -t.$$

Wie im Beispiel oben ist darauf zu achten, dass diese Gleichung auch für $t = 0$ gültig ist. Wir erhalten $\partial_{21} f(0,0) = (\partial_1 f \circ \phi_2)'(0) = -1$. Mit Hilfe der Funktion

$$(\partial_2 f \circ \phi_1)(t) \quad = \quad \partial_2 f(t,0) \quad = \quad \frac{t^5}{t^4} \quad = \quad t$$

ergibt sich $\partial_{12} f(0,0) = (\partial_2 f \circ \phi_1)'(0) = 1$, insgesamt also $\partial_{12} f(0,0) \neq \partial_{21} f(0,0)$.

Wir untersuchen nun, welche hinreichende Bedingung es für die Vertauschbarkeit der beiden Ableitungen gibt. Dazu führen wir die folgende Notation ein: Ist V ein endlich-dimensionaler $\mathbb{R}$-Vektorraum, und sind $a, b \in V$ beliebig vorgegeben, dann bezeichnen wir die Menge

$$[a,b] \quad = \quad \{(1-t)a + tb \mid 0 \leq t \leq 1$$

als ***Verbindungsstrecke*** zwischen a und b. Außerdem setzen wir $]a,b[= [a,b] \setminus \{a,b\}$.

(7.5) Satz. *(Mittelwertsatz für Richtungsableitungen)*

Sei $U \subseteq V$ eine offene Teilmenge und $f : U \to \mathbb{R}$ eine reellwertige Funktion. Seien $a, b \in U$ zwei verschiedene Punkte mit $[a,b] \subseteq U$ und der Eigenschaft, dass die Richtungsableitung $\partial_v f$ für $v = b - a$ auf ganz U existiert. Dann gibt es ein $p \in]a,b[$ mit

$$f(b) - f(a) \quad = \quad \partial_v f(p).$$

Beweis: Sei $\phi : \mathbb{R} \to V$ gegeben durch $\phi(t) = (1-t)a + tb = a + t(b-a) = a + tv$ für alle $t \in \mathbb{R}$. Nach (7.2) gilt $\partial_v f(a+tv) = (f \circ \phi)'(t)$ für alle $t \in \mathbb{R}$ mit $\phi(t) \in U$; insbesondere ist $f \circ \phi$ in diesen Punkten t differenzierbar. Nach Voraussetzung gilt $\phi(t) \in U$ für $0 \leq t \leq 1$. Weil f und ϕ stetig sind, ist $f \circ \phi$ insgesamt also eine auf $[0,1]$ definierte und stetige und auf $]0,1[$ differenzierbare Funktion. Nach dem Mittelwertsatz der Differenzialrechnung in einer Variablen gibt es also einen Punkt $t_0 \in]0,1[$ mit $(f \circ \phi)(1) - (f \circ \phi)(0) = (f \circ \phi)'(t_0) \cdot (1-0) = (f \circ \phi)'(t_0)$. Definieren wir $p = \phi(t_0)$, dann gilt $p \in]a,b[$ und

$$\partial_v f(p) \quad = \quad \partial_v f(a + t_0 v) \quad = \quad (f \circ \phi)'(t_0) \quad = \quad (f \circ \phi)(1) - (f \circ \phi)(0) \quad = \quad f(b) - f(a). \qquad \square$$

Für unseren Satz über die Vertauschbarkeit von Richtungsableitungen benötigen wir noch die folgende Hilfsaussage.

(7.6) Lemma. Sei $U \subseteq V$ offen. Seien $v, w \in V$, und sei $f : U \to \mathbb{R}$ eine reellwertige Funktion mit der Eigenschaft, dass die doppelte Richtungsableitung $\partial_v \partial_w f$ auf ganz U existiert. Sei außerdem $a \in U$ ein Punkt, so dass die Menge

$$R \quad = \quad \{a + tv + t'w \mid t, t' \in [0,1]\}$$

vollständig in U enthalten ist. Dann gibt es ein $p \in R$ mit

$$f(a+v+w) - f(a+v) - f(a+w) + f(a) \quad = \quad \partial_v \partial_w f(p).$$

Beweis: Im Fall $v = 0$ oder $w = 0$ ist die Aussage offenbar erfüllt, da in diesem Fall beide Seiten der Gleichung Null sind. Deshalb können wir von nun an $v, w \neq 0$ voraussetzen. Sei $\tau_v : V \to V$ die Translationsabbildung $x \mapsto x + v$. Damit für einen Punkt u die Differenz $f(x+v) - f(x)$ definiert ist, müssen sowohl x als auch $\tau_v(x)$ in U liegen. Setzen wir also $U_v = U \cap \tau_v^{-1}(U)$, dann ist durch $f_v(x) = f(x+v) - f(x)$ eine Abbildung $U_v \to \mathbb{R}$ definiert. Weil τ_v stetig und U offen ist, handelt es sich auch bei U_v um eine offene Teilmenge von U. Für jedes $x \in U_v$ gilt außerdem

$$\partial_w f_v(x) \quad = \quad \partial_w f(x+v) - \partial_w f(x).$$

Definieren wir nämlich Hilfsfunktionen $\phi :]-\varepsilon,\varepsilon[\to V$, $t \mapsto x + tw$ und $\phi_v :]-\varepsilon,\varepsilon[\to V$, $t \mapsto x + v + tw$ (wobei $\varepsilon \in \mathbb{R}^+$ so klein gewählt ist, dass das Bild des Intervalls in U_v liegt), dann gilt

$$\partial_w f_v(x) \quad = \quad (f_v \circ \phi)'(0) \qquad \text{und} \qquad \partial_w f(x+v) - \partial_w f(x) \quad = \quad (f \circ \phi_v)'(0) - (f \circ \phi)'(0)$$

nach Definition der Richtungsableitungen. Diese stimmen überein, denn für alle $t \in]-\varepsilon,\varepsilon[$ gilt

$$(f_v \circ \phi)(t) \quad = \quad f_v(x+tw) \quad = \quad f(x+v+tw) - f(x+tw) \quad = \quad (f \circ \phi_v)(t) - (f \circ \phi)(t).$$

Zu zeigen ist nun $f_v(a+w) - f_v(a) = \partial_v \partial_w f(p)$ für ein $p \in R$. Dazu bemerken wir zunächst, dass die Strecke $[a, a+w]$ ist in U_v enthalten. Denn für jedes $t \in [0,1]$ liegen die beiden Punkte $a+tw$ und $\tau_v(a+tw) = a+v+tw$ in $R \subseteq U$, also liegt $a+tw$ in $U \cap \tau_v^{-1}(U) = U_v$. Die Anwendung von (7.5) auf die Funktion f_v liefert nun einen Punkt $p' \in [a, a+w]$ mit

$$f_v(a+w) - f_v(a) \quad = \quad \partial_w f_v(p') \quad = \quad \partial_w f(p'+v) - \partial_w f(p').$$

Nochmalige Anwendung von (7.5), diesmal auf die Funktion $\partial_w f$, liefert einen Punkt $p \in [p', p'+v]$ mit $\partial_v \partial_w f(p) = \partial_w f(p'+v) - \partial_w f(p')$. Ingesamt gilt also $\partial_v \partial_w f(p) = f_v(a+w) - f_v(a)$ wie gewünscht. □

(7.7) Satz. *(Satz von Schwarz)*

Sei $U \subseteq V$ offen und $f : U \to \mathbb{R}$ eine reellwertige Funktion, und seien $0 \neq v, w \in V$ derart, dass die zweifachen Richtungsableitungen $\partial_v \partial_w f$ und $\partial_w \partial_v f$ auf U existieren und stetig sind. Dann gilt $\partial_v \partial_w f = \partial_w \partial_v f$ auf ganz U.

Beweis: Sei $a \in U$ vorgegeben, und seien $(\alpha_n)_{n\in\mathbb{N}}$ und $(\beta_n)_{n\in\mathbb{N}}$ Folgen positiver reeller Zahlen mit $\lim_n \alpha_n = \lim_n \beta_n = 0$ und der Eigenschaft, dass der Bereich $Q_n = \{a + tv + t'w \mid t \in [0, \alpha_n], t' \in [0, \beta_n]\}$ jeweils vollständig in U enthalten ist. Nach Lemma (7.6) gibt es für jedes $n \in \mathbb{N}$ jeweils Punkte $p_n, q_n \in Q_n$ mit

$$\begin{aligned} \alpha_n \beta_n \partial_v \partial_w f(p_n) \quad &= \quad \partial_{(\alpha_n v)} \partial_{(\beta_n w)} f(p_n) \quad = \quad f(a + \alpha_n v + \beta_n w) - f(a + \alpha_n v) - f(a + \beta_n w) + f(a) \\ &= \quad \partial_{(\beta_n w)} \partial_{(\alpha_n v)} f(q_n) \quad = \quad \alpha_n \beta_n \partial_w \partial_v f(q_n). \end{aligned}$$

Division durch $\alpha_n \beta_n$ liefert die Gleichung $\partial_v \partial_w f(p_n) = \partial_w \partial_v f(q_n)$ für alle $n \in \mathbb{N}$. Weil die Folgen $(\alpha_n)_{n\in\mathbb{N}}$ und $(\beta_n)_{n\in\mathbb{N}}$ gegen Null konvergieren, gilt $\lim_n p_n = \lim_n q_n = a$. Weil die $\partial_v \partial_w f$ und $\partial_w \partial_v f$ stetig sind, folgt schließlich $\partial_w \partial_v f(a) = \lim_n \partial_w \partial_v f(q_n) = \lim_n \partial_v \partial_w f(p_n) = \partial_v \partial_w f(a)$. □

§ 8. Totale Differenzierbarkeit

Zusammenfassung. Im letzten Abschnitt haben wir gesehen, dass der Begriff der Richtungsableitung nur teilweise das leistet, was man von der eindimensionalen Differenzierbarkeit her erwartet. Zum Beispiel haben wir gesehen, dass selbst die Existenz aller Richtungsableitungen in einem Punkt nicht die Stetigkeit der Funktion in diesem Punkt sicherstellt. Die in diesem Abschnitt eingeführte *totalen Ableitung* weist diese Mängel nicht mehr auf und kann als passendes Analogon zur eindimensionalen Ableitung angesehen werden. Allerdings handelt es sich bei der totalen Ableitung $f'(a)$ einer Funktion $f : \mathbb{R}^n \to \mathbb{R}^m$ in einem Punkt $a \in \mathbb{R}^n$ nicht um eine reelle Zahl, sondern um eine lineare Abbildung $\mathbb{R}^n \to \mathbb{R}^m$, die auch in Form einer $m \times n$-Matrix angegeben werden kann.

Neben der Definition klären wir in diesem Abschnitt auch, in welchem Zusammenhang die die totale Differenzierbarkeit mit der Existenz der Richtungsableitungen steht. Wie wir sehen werden, können die Richtungsableitungen von f auf einfache Weise aus der totalen Ableitung berechnet werden, denn es gilt $\partial_v f(a) = f'(a)(v)$ für alle $v \in \mathbb{R}^n$. Wie in der eindimensionalen Analysis existieren auch für die mehrdimensionale totale Ableitung Summen-, Produkt-, Ketten- und Umkehrregel.

Wichtige Begriffe und Sätze in diesem Kapitel

- Definition der totalen Ableitung einer Funktion
- Zusammenhang zwischen totaler Ableitung und den Richtungsableitungen
- hinreichendes Kriterium für totale Differenzierbarkeit
- Ableitungsregeln: Summen-, Produkt-, Ketten und Umkehrregel

In der Analysis einer Variablen haben wir gesehen, dass sich die Differenzierbarkeit einer Funktion $f : I \to \mathbb{R}$ in einem Punkt $a \in I$ dadurch charakterisieren lässt, dass sich $f(a+h)$ für kleines h durch eine affin-lineare Funktion der Form $h \mapsto f(a) + f'(a)h$ approximieren lässt. Dies bedeutet, dass eine Darstellung von f der Form

$$f(a+h) \quad = \quad f(a) + f'(a)h + \psi(h)$$

existiert, mit einer Funktion ψ, die für $h \to 0$ sehr schnell gegen Null geht. Diese Vorstellung von der Differenzierbarkeit einer Funktion soll nun auf höherdimensionale Funktionen verallgemeinert werden.

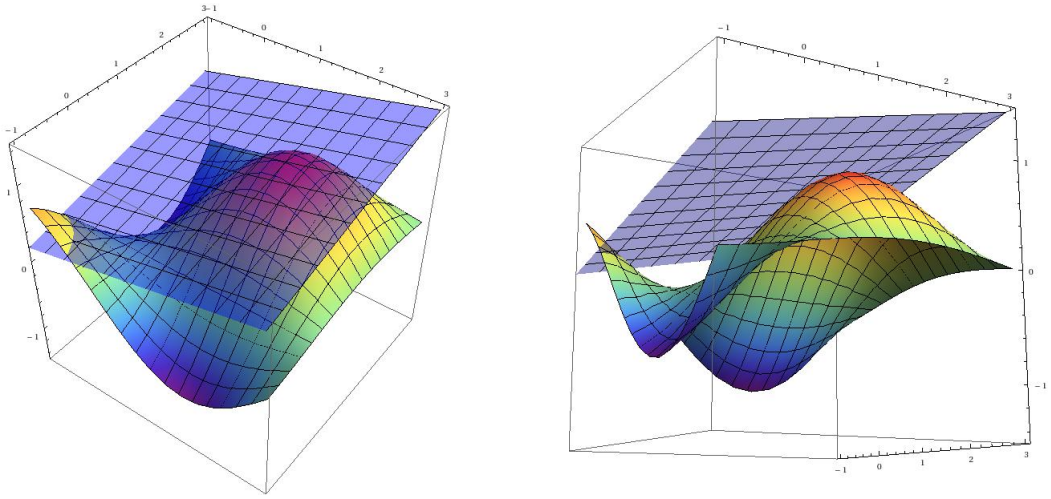
Im gesamten Abschnitt seien V, W jeweils endlich-dimensionale, normierte $\mathbb{R}$-Vektorräume.

(8.1) Definition. Sei $U \subseteq V$ offen, $a \in U$ und $U_a = \{x \in V \mid a + x \in U\}$. Außerdem sei $f : U \to W$ eine Abbildung. Man sagt, die Funktion f ist im Punkt a ***total differenzierbar***, wenn eine lineare Abbildung $\phi : V \to W$ und eine Funktion $\psi : U_a \to W$ existieren, so dass

$$f(a+h) = f(a) + \phi(h) + \psi(h) \text{ für alle } h \in U_a \quad \text{und außerdem} \quad \lim_{h \to 0_V} \frac{\psi(h)}{\|h\|} = 0_W$$

erfüllt ist. Man nennt ϕ dann die ***Ableitung*** von f an der Stelle a und bezeichnet sie mit $f'(a)$.

Im Gegensatz zum eindimensionalen Fall ist die Ableitung also keine reelle Zahl mehr, sondern eine lineare Abbildung, genauer ein Element des $\mathbb{R}$-Vektorraums $\mathscr{L}(V, W)$ der linearen Abbildungen von V nach W. Häufig wird der Zusatz „total“ auch weggelassen; wenn im Mehrdimensionalen von einer differenzierbaren Funktion gesprochen wird, dann ist immer totale Differenzierbarkeit gemeint.



Veranschaulichung der totalen Ableitung einer Funktion f auf $\mathbb{R}^2$. Die blaue Fläche stellt die affin-lineare Näherung $\tilde{f}(a+h) = f(p) + f'(a)(h)$ dar, die der Ableitung der Funktion in einem Punkt $a \in \mathbb{R}^2$ entspricht.

(8.2) Proposition. Sei $f : U \to W$ eine Abbildung auf einer offenen Teilmenge $U \subseteq V$, und $a \in U$ ein beliebiger Punkt. Ist f in a differenzierbar, dann ist f auch in a stetig.

Beweis: Sei $f(a+h) = f(a) + f'(a)(h) + \psi(h)$ eine Darstellung von f wie in der Definition der Differenzierbarkeit angegeben. Als lineare Abbildung auf einem endlich-dimensionalen $\mathbb{R}$-Vektorraum ist $f'(a)$ stetig, und wegen $\lim_{h \to 0_V} \psi(h)/\|h\| = 0_W$ gilt auch $\lim_{h \to 0_V} \psi(h) = 0_W$. Es folgt

$$\lim_{h \to 0} f(a+h) \quad = \quad f(a) + \lim_{h \to 0_V} f'(a)(h) + \lim_{h \to 0_V} \psi(h) \quad = \quad f(a) + f'(a)(0_V) \quad = \quad f(a). \qquad \square$$

(8.3) Proposition. Wir betrachten den Spezialfall, dass $W = \mathbb{R}^m$ für ein $m \in \mathbb{N}$ gilt. Sei $U \subseteq V$ offen und $a \in U$. Eine Abbildung $f : U \to W$ ist genau dann in a differenzierbar, wenn die Komponentenfunktionen $f_i : U \to \mathbb{R}$ in a differenzierbar sind, für $1 \leq i \leq m$. Die Komponentenfunktionen der Ableitung $f'(a)$ sind dann die Ableitungen $f_1'(a), ..., f_m'(a)$.

Beweis: Sei $\phi \in \mathscr{L}(V, \mathbb{R}^m)$ eine lineare Abbildung, und seien $\phi_i \in \mathscr{L}(V, \mathbb{R})$ ihre Komponentenfunktionen, für $1 \leq i \leq m$. Definieren wir die Abbildung $\psi : U_a \to \mathbb{R}^m$ auf $U_a = \{x \in V \mid a + x \in U\}$ durch $\psi(h) = f(a+h) - f(a) - \phi(h)$, dann sind die Komponentenfunktionen von ψ durch $\psi_i(h) = f_i(a+h) - f_i(a) - \phi_i(h)$ gegeben, für $1 \leq i \leq m$. Ist $(h^{(n)})_{n \in \mathbb{N}}$ eine Folge in U_a, die gegen 0_V konvergiert, so gilt nach (2.5) die Äquivalenz

$$\lim_{n \to \infty} \frac{\psi(h^{(n)})}{\|h^{(n)}\|} = 0_{\mathbb{R}^m} \qquad \Leftrightarrow \qquad \lim_{n \to \infty} \frac{\psi_i(h^{(n)})}{\|h^{(n)}\|} = 0 \quad \text{für} \quad 1 \leq i \leq m.$$

Also ist $\lim_{h \to 0_V} \psi(h)/\|h\| = 0_{\mathbb{R}^m}$ äquivalent zu $\lim_{h \to 0_V} \psi_i(h)/\|h\| = 0$ für $1 \leq i \leq m$.

„⇒“ Sei f in a differenzierbar. Dann setzen wir $\phi = f'(a)$ und definieren $\psi : U_a \to \mathbb{R}^m$ in Abhängigkeit von ϕ wie oben angegeben. Auf Grund der Differenzierbarkeit gilt $\lim_{h \to 0_V} \psi(h)/\|h\| = 0_{\mathbb{R}^m}$, und es folgt $\lim_{h \to 0_V} \psi_i(h)/\|h\| = 0$ für $1 \leq i \leq m$. Dies zusammen mit der Gleichung $f_i(a+h) = f_i(a) + \phi_i(h) + \psi_i(h)$ für $h \in U_a$ zeigt, dass f_i im Punkt a differenzierbar ist, und dass die Komponentenfunktion ϕ_i jeweils die Ableitung von f_i im Punkt a ist.

„⇐“ Sei f_i in a differenzierbar, $\phi_i = f_i'(a)$ für $1 \le i \le m$, und sei $\phi \in \mathscr{L}(V, \mathbb{R}^m)$ die eindeutig bestimmte lineare Abbildung mit den Komponentenfunktionen $\phi_i \in \mathscr{L}(V, \mathbb{R})$. Wiederum sei die Abbildung ψ in Abhängigkeit von ϕ wie oben definiert. Aus der Differenzierbarkeit der Funktionen f_i folgt jeweils $\lim_{h \to 0_V} \psi_i(h)/\|h\| = 0$, für $1 \le i \le m$. Nach unserer Vorüberlegung bedeutet dies $\lim_{h \to 0_V} \psi(h)/\|h\| = 0_{\mathbb{R}^m}$. Also ist f im Punkt a differenzierbar, und es gilt $\phi = f'(a)$. □

(8.4) Proposition. Sei $U \subseteq V$ offen, $f : U \to \mathbb{R}$ eine reellwertige Funktion, und $a \in U$. Ist f im Punkt a differenzierbar, dann existiert für jedes $v \in V$ die Richtungsableitung $\partial_v f(a)$, und es gilt $\partial_v f(a) = f'(a)(v)$.

Beweis: Nach Definition gibt es auf $U_a = \{x \in V \mid a + x \in U\}$ eine Abbildung $\psi : U_a \to \mathbb{R}$ mit

$$f(a+h) \quad = \quad f(a) + f'(a)(h) + \psi(h) \quad \text{für alle} \quad h \in U_a \qquad \text{und} \qquad \lim_{h \to 0} \psi(h)/\|h\| = 0.$$

Weil die Gleichung auch für $h = 0$ erfüllt ist, muss $\psi(0) = 0$ gelten. Sei $\phi : \mathbb{R} \to V$ gegeben durch $\phi(t) = a + tv$ für alle $t \in \mathbb{R}$ und $\varepsilon \in \mathbb{R}^+$ hinreichend klein gewählt, so dass $\phi(]-\varepsilon, \varepsilon[) \subseteq U$ erfüllt ist. Nach Definition gilt $\partial_v f(a) = (f \circ \phi)'(0)$. Zu zeigen ist also die Gleichung $(f \circ \phi)'(0) = f'(a)(v)$. Für alle $t \in \mathbb{R}$ mit $|t| < \varepsilon$ gilt

$$(f \circ \phi)(t) \quad = \quad f(a+tv) \quad = \quad f(a) + f'(a)(tv) + \psi(tv) \quad = \quad f(a) + tf'(a)(v) + \psi(tv).$$

Nach Definition der Ableitung in einer Variablen gilt $(f \circ \phi)'(0) = \lim_{t \to 0} h(t)$, wobei $h(t)$ für $0 < |t| < \varepsilon$ durch

$$h(t) \quad = \quad \frac{(f \circ \phi)(t) - (f \circ \phi)(0)}{t} \quad = \quad \frac{f(a+tv) - f(a)}{t} \quad = \quad f'(a)(v) + \frac{\psi(tv)}{t}$$

definiert ist. Betrachten wir nun zunächst den Fall $v = 0$. Wegen $\psi(0) = 0$ gilt hier $h(t) = 0$ für alle t mit $0 < |t| < \varepsilon$ und somit $(f \circ \phi)'(0) = 0 = f'(a)(0)$. Im Fall $v \neq 0$ betrachten wir eine Folge $(t_n)_{n \in \mathbb{N}}$ mit $\lim_n t_n = 0$ und $0 < |t_n| < \varepsilon$ für alle $n \in \mathbb{N}$. Dann konvergiert die Folge $(t_n v)_{n \in \mathbb{N}}$ in V gegen Null. Auf Grund der Voraussetzung $\lim_{h \to 0} \psi(h)/\|h\| = 0$ gilt

$$\|v\|^{-1} \lim_{n \to \infty} \frac{\psi(t_n v)}{|t_n|} \quad = \quad \lim_{n \to \infty} \frac{\psi(t_n v)}{|t_n|\|v\|} \quad = \quad \lim_{n \to \infty} \frac{\psi(t_n v)}{\|t_n v\|} \quad = \quad 0$$

und folglich $(f \circ \phi)'(0) = \lim_{n \to \infty} h(t_n) = f'(a)(v) + \lim_{n \to \infty} \psi(t_n v)/t_n = f'(a)(v)$. □

(8.5) Folgerung. Ist $U \subseteq V$ offen, $a \in U$ und $f : U \to \mathbb{R}$ in a differenzierbar, dann gilt $\partial_{v+w} f(a) = \partial_v f(a) + \partial_w f(a)$ für alle $v, w \in V$.

Beweis: Seien $v, w \in V$. Durch (8.4) ist sichergestellt, dass die drei angegebenen Richtungsableitungen existieren. Weil $f'(a)$ eine lineare Abbildung ist, gilt außerdem

$$\partial_{v+w} f(a) \quad = \quad f'(a)(v+w) \quad = \quad f'(a)(v) + f'(a)(w) \quad = \quad \partial_v f(a) + \partial_w f(a).$$ □

(8.6) Definition. Seien $m, n \in \mathbb{N}$, $U \subseteq \mathbb{R}^n$ offen, $a \in U$ und $f : U \to \mathbb{R}^m$ eine in a differenzierbare Abbildung, mit Komponentenfunktionen $f_1, ..., f_m$. Dann nennt man

$$\mathrm{Jac}(f)(a) \quad = \quad \begin{pmatrix} \partial_1 f_1(a) & \dots & \partial_n f_1(a) \\ \vdots & & \vdots \\ \partial_1 f_m(a) & \dots & \partial_n f_m(a) \end{pmatrix} \quad \in \mathscr{M}_{m\times n,\mathbb{R}}$$

die ***Jacobi-*** oder ***Funktionalmatrix*** von f an der Stelle a.

Wir werden in Kürze ein Kriterium kennenlernen, mit dem die Differenzierbarkeit einer Funktion in vielen Fällen leicht zu erkennen ist. Dieses Kriterium wird unter anderem zeigen, dass alle Abbildungen, die durch Polynomausdrücke in den Variablen dargestellt werden können, differenzierbar sind, beispielsweise die Abbildungen

$$f : \mathbb{R}^2 \to \mathbb{R}\ ,\ (x,y) \mapsto 2x^2 + 7y^2 + 5y \quad \text{und} \qquad g : \mathbb{R}^3 \to \mathbb{R}^2\ ,\ (x,y,z) \mapsto (3x^2 - 5y + z, z^4 + xy).$$

Für jedes $(x,y) \in \mathbb{R}^2$ ist $\mathrm{Jac}(f)(x,y) \in \mathscr{M}_{1\times 2,\mathbb{R}}$ gegeben durch $(4x \;\; 14y + 5)$. Die Jacobi-Matrizen der Funktion g sind gegeben durch

$$\mathrm{Jac}(g)(x,y,z) \quad = \quad \begin{pmatrix} 6x & -5 & 1 \\ y & x & 4z^3 \end{pmatrix} \in \mathscr{M}_{2\times 3,\mathbb{R}} \qquad \text{für} \quad (x,y,z) \in \mathbb{R}^3.$$

(8.7) Proposition. Seien die Bezeichnungen wie in (8.6) gewählt, und sei $J = \mathrm{Jac}(f)(a)$. Dann gilt $f'(a)(v) = J \cdot v$ für alle $v \in \mathbb{R}^n$, wobei der Ausdruck $J \cdot v$ das Matrix-Vektor-Produkt zwischen der Matrix $J \in \mathscr{M}_{m\times n,\mathbb{R}}$ und dem Vektor $v \in \mathbb{R}^n$ bezeichnet. Die Matrix J ist also die Darstellungsmatrix der linearen Abbildung $f'(a) : \mathbb{R}^n \to \mathbb{R}^m$ bezüglich der Einheitsbasen.

Beweis: Für $1 \le i \le m$ seien $f'(a)_i$ und $(\phi_J)_i$ jeweils die i-te Komponentenfunktion linearen Abbildungen $f'(a)$ und $\phi_J : \mathbb{R}^n \to \mathbb{R}^m$, $v \mapsto J \cdot v$. Nach (8.3) gilt jeweils $f'(a)_i = f_i'(a)$. Es genügt also, $(\phi_J)_i = f_i'(a)$ für $1 \le i \le m$ zu beweisen. Auf Grund der Linearität der beiden Abbildungen genügt es wiederum, die Gleichungen $f_i'(a)(e_j) = (\phi_J)_i(e_j)$ für $1 \le i \le m$ und $1 \le j \le n$ zu überprüfen, wobei $e_1, ..., e_n$ die Einheitsvektoren von $\mathbb{R}^n$ bezeichnen. Für $i \in \{1, ..., m\}$ ist die Abbildung $(\phi_J)_i$ gegeben durch das Matrix-Vektor-Produkt mit der einzeiligen Matrix

$$(\partial_1 f_i(a) \;\cdots\; \partial_n f_i(a)) \quad \in \mathscr{M}_{1\times n,\mathbb{R}}.$$

Durch Einsetzen des j-ten Einheitsvektors wird der j-te Eintrag ausgewählt, es gilt also $(\phi_J)_i(e_j) = \partial_j f_i(a)$ für $1 \le j \le n$. Aus (8.4) folgt andererseits auch $f_i'(a)(e_j) = \partial_{e_j} f_i(a) = \partial_j f_i(a)$. □

Um die Notation möglichst einfach zu halten, schreiben wir in Zukunft für die Jacobi-Matrix ebenfalls $f'(a)$ statt $\mathrm{Jac}(f)(a)$. Für jede Funktion $f : U \to \mathbb{R}^m$ auf einer offenen Teilmenge $U \subseteq \mathbb{R}^n$, jeden Punkt $a \in U$ und jedes $v \in \mathbb{R}^n$ gilt also $f'(a)(v) = f'(a) \cdot v$ wobei auf der linken Seite der Gleichung $f'(a)$ als lineare Abbildung $\mathbb{R}^n \to \mathbb{R}^m$ interpretiert wird, während auf der rechten Seite das Matrix-Vektor-Produkt von $f'(a) \in \mathscr{M}_{m\times n,\mathbb{R}}$ und $v \in \mathbb{R}^n$ gemeint ist.

Nachdem wir nun einen Weg gefunden haben, die totale Ableitung einer Abbildung explizit anzugeben, soll hier noch einmal die Definition der totalen Differenzierbarkeit an einem konkreten Beispiel illustriert werden. Sei die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch $f(x,y) = x^2 + y^2$. Dann gilt $\partial_1 f(x,y) = 2x$ und $\partial_2 f(x,y) = 2y$. Im Falle der totalen Differenzierbarkeit muss also $f'(x,y) = (2x\ 2y)$ gelten. Ist f insbesondere an der Stelle $(3,4)$ total differenzierbar, dann gilt $f'(3,4) = (6\ \ 8)$, und der Fehlerterm $\psi(h_1,h_2)$ ist gegeben durch

$$\begin{aligned}\psi(h_1,h_2) &= f(3+h_1,4+h_2) - f(3,4) - \begin{pmatrix}6 & 8\end{pmatrix}\begin{pmatrix}h_1\\h_2\end{pmatrix} = \\ &(3+h_1)^2 + (4+h_2)^2 - 3^2 - 4^2 - 6h_1 - 8h_2 = \\ &9 + 6h_1 + h_1^2 + 16 + 8h_2 + h_2^2 - 9 - 16 - 6h_1 - 8h_2 = h_1^2 + h_2^2.\end{aligned}$$

Für jedes $h = (h_1,h_2) \in \mathbb{R}^2$ gilt

$$\frac{\psi(h)}{\|h\|_\infty} = \frac{h_1^2+h_2^2}{\max\{|h_1|,|h_2|\}} = \frac{h_1^2}{\max\{|h_1|,|h_2|\}} + \frac{h_2^2}{\max\{|h_1|,|h_2|\}} \leq |h_1| + |h_2| = \|h\|_1$$

und somit $\lim_{h\to 0} \psi(h)/\|h\|_\infty = \lim_{h\to 0} \|h\|_1 = 0$. Wir haben also direkt anhand der Definition überprüft, dass unsere Funktion f an der Stelle $(3,4)$ total differenzierbar ist. Andererseits wird anhand des nun folgenden hinreichendes Kriteriums sofort klar werden, dass f in *jedem* Punkt seines Definitionsbereichs $\mathbb{R}^2$ total differenzierbar ist.

(8.8) Satz. Sei $U \subseteq \mathbb{R}^n$ eine offene Teilmenge und $f : U \to \mathbb{R}$ eine partiell differenzierbare Funktion, und sei $a \in U$ ein Punkt, in dem die partiellen Ableitungen $\partial_i f$ stetig sind, für $1 \leq i \leq n$. Dann ist f in a differenzierbar.

Beweis: Sei $\varepsilon \in \mathbb{R}^+$ hinreichend klein gewählt, so dass der offene Ball $B_\varepsilon(a)$ bezüglich der Maximums-Norm $\|\cdot\|_\infty$ in U enthalten ist. Jedem Vektor $v \in B_\varepsilon(0_{\mathbb{R}^n})$, $v = (v_1,\ldots,v_n)$ ordnen wir durch

$$z^{(k)}(v) = a + \sum_{i=1}^{k} v_i e_i \qquad \text{für} \quad 0 \leq k \leq n$$

insgesamt $n+1$ Punkte $z^{(k)}(v) \in B_\varepsilon(a)$ zu, wobei $e_1,\ldots,e_n$ wie immer die Einheitsvektoren bezeichnen und $z^{(0)}(v) = a$, $z^{(n)}(v) = a + v$ gilt. Weil $B_\varepsilon(a)$ als offener Ball konvex ist, liegen alle Verbindungsstrecken $[z^{(k-1)}(v), z^{(k)}(v)]$ mit $1 \leq k \leq n$ jeweils in $B_\varepsilon(a)$. Nach dem Mittelwertsatz für Richtungsableitungen (7.5) gibt es für $1 \leq k \leq n$ im Fall $v_k \neq 0$ jeweils ein $y^{(k)}(v) \in [z^{(k-1)}(v), z^{(k)}(v)]$ mit

$$f(z^{(k)}(v)) - f(z^{(k-1)}(v)) = \partial_{v_k e_k} f(y^{(k)}(v)) = v_k \partial_k f(y^{(k)}(v)).$$

Im Fall $v_k = 0$ ist die Gleichung mit $y^{(k)}(v) = z^{(k-1)}(v) = z^{(k)}(v)$ ebenfalls erfüllt. Definieren wir nun eine lineare Abbildung $\phi : \mathbb{R}^n \to \mathbb{R}$ und eine weitere Abbildung $\psi : B_\varepsilon(0_{\mathbb{R}^n}) \to \mathbb{R}$ durch

$$\phi(v) = \sum_{k=1}^{n} v_k \partial_k f(a) \qquad \text{und} \qquad \psi(v) = \sum_{k=1}^{n} v_k (\partial_k f(y^{(k)}(v)) - \partial_k f(a)) \quad ,$$

dann erhalten wir

$$\begin{aligned}f(a+v) &= f(z^{(n)}(v)) = f(z^{(0)}(v)) + \sum_{k=1}^{n} \big(f(z^{(k)}(v)) - f(z^{(k-1)}(v))\big) = \\ f(a) + \sum_{k=1}^{n} v_k \partial_k f(y^{(k)}(v)) &= f(a) + \sum_{k=1}^{n} v_k \partial_k f(a) + \psi(v) = f(a) + \phi(v) + \psi(v).\end{aligned}$$

Betrachten wir nun den Grenzübergang $\lim_{v\to 0} \psi(v)/\|v\|_\infty$. Wenn v bezüglich $\|\cdot\|_\infty$ gegen Null konvergiert, dann laufen sowohl die Punkte $z^{(k)}(v)$ für $0 \le k \le n$ als auch die Punkte $y^{(k)}(v)$ für $0 < k \le n$ gegen a. Auf Grund der Stetigkeit der partiellen Ableitungen $\partial_k f$ im Punkt a konvergieren damit die Differenzen $\partial_k f(y^{(k)}(v)) - \partial_k f(a)$ gegen Null. Darüber hinaus ist $v_k/\|v\|_\infty$ durch 1 nach oben beschränkt. Insgesamt erhalten wir also $\lim_{v\to 0} \psi(v)/\|v\|_\infty = 0$. Folglich ist f im Punkt a differenzierbar, und es gilt $f'(a) = \phi$. □

Wir untersuchen die bisherigen Beispiele auf totale Differenzierbarkeit.

(i) Die Funktionen $f : \mathbb{R}^2 \to \mathbb{R}$, $(x,y) \mapsto 2x^2 + 7y^2 + 5y$ und $g : \mathbb{R}^2 \to \mathbb{R}$, $(x,y) \mapsto \sin(2x)e^{3y}$ sind in allen Punkten ihres Definitionsbereichs differenzierbar, weil ihre partiellen Ableitungen alle auf ganz $\mathbb{R}^2$ stetig sind.

(ii) Die Funktion aus Beispiel 4 in § 10 ist im Punkt $(0,0)$ nicht differenzierbar, weil die Gleichung

$$\partial_v f(0,0) + \partial_w f(0,0) \quad = \quad \partial_{v+w} f(0,0)$$

für $v = e_1$ und $w = e_2$ nicht gilt. Bei einer differenzierbaren Funktion wäre die Gleichung nach (8.5) erfüllt.

(iii) Die Funktion f aus Beispiel 5 in § 10 ist im Punkt $(0,0)$ nicht differenzierbar, weil sie dort nicht einmal stetig ist, vgl. (8.2).

Ist eine Funktion auf ihrem Definitionsbereich also stetig partiell differenzierbar, dann ist sie auch (total) differenzierbar. Aus der Differenzierbarkeit wiederum folgen sowohl partielle Differenzierbarkeit als auch Stetigkeit.

Für die totale Ableitung gelten wie im eindimensionalen Fall eine Reihe von Ableitungsregeln.

(8.9) Satz. *(Additionsregel)*

Sei $U \subseteq V$ offen, $a \in U$, und seien $f, g : U \to W$ zwei Abbildungen, die beide in $a \in U$ differenzierbar sind. Dann sind auch $f+g$ und λf für alle $\lambda \in \mathbb{R}$ im Punkt a differenzierbar, und es gilt

$$(f+g)'(a) = f'(a) + g'(a) \quad \text{und} \quad (\lambda f)'(a) = \lambda f'(a).$$

Beweis: Nach Definition der Differenzierbarkeit im Punkt a können wir beide Funktionen in der Form

$$\begin{aligned} f(a+h) &= f(a) + f'(a)(h) + \varphi(h) \\ g(a+h) &= g(a) + g'(a)(h) + \psi(h) \end{aligned}$$

darstellen, wobei $\lim\limits_{h\to 0} \varphi(h)/\|h\| = \lim\limits_{h\to 0} \psi(h)/\|h\| = 0$ ist. Es folgt

$$(f+g)(a+h) \quad = \quad (f+g)(a) + (f'(a) + g'(a))(h) + (\varphi + \psi)(h)$$

mit

$$\lim_{h\to 0} \frac{(\varphi+\psi)(h)}{\|h\|} \quad = \quad \lim_{h\to 0} \frac{\varphi(h)}{\|h\|} + \lim_{h\to 0} \frac{\psi(h)}{\|h\|} \quad = \quad 0.$$

Also ist $f+g$ in a differenzierbar, und es gilt $(f+g)'(a) = f'(a) + g'(a)$. Ebenso erhalten wir $(\lambda f)(a+h) = (\lambda f)(a) + (\lambda f'(a))(h) + \lambda\varphi(h)$, und für den Restterm gilt $\lim\limits_{h\to 0} \lambda\varphi(h)/\|h\| = 0$. Somit ist auch λf in a differenzierbar, und die Ableitung ist durch $(\lambda f)'(a) = \lambda f'(a)$ gegeben. □

(8.10) Satz. *(Produktregel)*

Sei $U \subseteq V$ offen, $a \in U$, und seien $f, g : U \to \mathbb{R}$ beide in a differenzierbar. Dann ist auch die Funktion $f\,g$ in a differenzierbar, und es gilt

$$(f\,g)'(a) \quad = \quad f(a)g'(a) + g(a)f'(a).$$

Beweis: Durch Ausmultiplizieren der beiden Gleichungen

$$\begin{aligned} f(a+h) &= f(a) + f'(a)(h) + \varphi(h) \\ g(a+h) &= g(a) + g'(a)(h) + \psi(h) \end{aligned}$$

(wobei die Funktionen φ und ψ dieselbe Grenzwert-Bedingung wie im vorherigen Beweis erfüllen) erhalten wir

$$(f\,g)(a+h) = (f\,g)(a) + f(a)g'(a)(h) + g(a)f'(a)(h) + \rho(h) \tag{4}$$

mit dem Restterm

$$\rho(h) \quad = \quad f'(a)(h)g'(a)(h) + g(a)\varphi(h) + f(a)\psi(h) + g'(a)(h)\varphi(h) + f'(a)(h)\psi(h) + \varphi(h)\psi(h).$$

Jeder der sechs Summanden des Restterms läuft für $h \to 0$ auch nach Division durch $\|h\|$ noch gegen Null. Für die hinteren fünf Summanden ist dies klar auf Grund der Voraussetzung an die Funktionen $\varphi(h)$ und $\psi(h)$ und wegen der Stetigkeit von linearen Abbildungen auf endlich-dimensionalen $\mathbb{R}$-Vektorräumen. Für den ersten Summanden ist zu beachten, dass die linearen Abbildungen $f'(a)$ und $g'(a)$ durch Konstanten $\gamma_1, \gamma_2 \in \mathbb{R}^+$ mit $\|f'(a)(h)\| \leq \gamma_1\|h\|$ und $\|g'(a)(h)\| \leq \gamma_2\|h\|$ für alle $h \in V$ abgeschätzt werden können. Für das Produkt gilt dann

$$\|f'(a)(h)g'(a)(h)\| \quad \leq \quad \gamma_1\gamma_2\|h\|^2 \quad ,$$

und der Ausdruck rechts konvergiert auch nach Division durch $\|h\|$ noch gegen Null. Die Gleichung (4) zeigt also, dass $f\,g$ tatsächlich in a differenzierbar ist, und dass die Ableitung die angegebene Form besitzt. □

(8.11) Folgerung. Sei $U \subseteq V$ offen, $f : U \to \mathbb{R}$, $n \in \mathbb{N}$ und $g(x) = f(x)^n$ für alle $x \in U$. Ist f in einem Punkt $a \in U$ differenzierbar, dann auch g, und es gilt

$$g'(a) \quad = \quad nf(a)^{n-1}f'(a).$$

Beweis: Wir führen den Beweis durch vollständige Induktion über n, wobei für $n = 1$ nichts zu zeigen ist. Setzen wir die Aussage nun für n voraus. Es sei $g(x) = f(x)^{n+1}$ und $h(x) = f(x)^n$ für alle $x \in U$. Nach Induktionsvoraussetzung ist h in a differenzierbar, mit $h'(a) = nf(a)^{n-1}f'(a)$. Durch Anwendung der Produktregel erhalten wir

$$\begin{aligned} g'(a) \quad = \quad (fh)'(a) \quad = \quad f(a)h'(a) + h(a)f'(a) \quad &= \quad nf(a)^n f'(a) + f(a)^n f'(a) \\ &= \quad (n+1)f(a)^n f'(a). \end{aligned}$$

□

Als Beispiel betrachten wir die Funktion $f : \mathbb{R}^3 \to \mathbb{R}$ gegeben durch $(x,y,z) \mapsto x+y+z$. Dann ist $f'(x,y,z) = (1\ 1\ 1)$ für alle $(x,y,z) \in \mathbb{R}^3$, und wir erhalten für beliebiges $v \in \mathbb{R}^3$, $v = (v_1, v_2, v_3)$

$$f'(x,y,z)(v) \quad = \quad \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad = \quad v_1 + v_2 + v_3$$

für alle $(x,y,z) \in \mathbb{R}^3$. Sei nun $g : \mathbb{R}^3 \to \mathbb{R}$ gegeben durch $g(x,y,z) = f(x,y,z)^3 = (x+y+z)^3$. Dann gilt $g'(x,y,z) = 3f(x,y,z)^2 f'(x,y,z)$ und somit

$$g'(x,y,z)(v) \quad = \quad 3(x+y+z)^2 \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad = \quad 3(x+y+z)^2(v_1+v_2+v_3).$$

Um den Beweis der folgenden Differentiationsregel zu vereinfachen, führen wir die folgende Vorüberlegung durch: Sei $U \subseteq V$ eine offene Teilmenge, $f : U \to W$ eine Abbildung, und seien $v \in V$, $w \in W$ beliebige Elemente. Dann ist auf der offenen Teilmenge $\tilde{U} = \{(-v) + u \mid u \in U\}$ von V durch $g : \tilde{U} \to W$, $x \mapsto f(v+x) + w$ ebenfalls eine Abbildung definiert.

(8.12) Lemma. Ist $a \in U$ und f in a differenzierbar, dann ist g im Punkt $(-v)+a$ differenzierbar, und es gilt $f'(a) = g'((-v)+a)$.

Beweis: Auf Grund der Voraussetzung gibt es eine Funktion ψ auf einer geeigneten offenen Teilmenge $U_a \subseteq V$, so dass $f(a+h) = f(a) + f'(a)(h) + \psi(h)$ für alle $h \in U_a$ und $\lim\limits_{h \to 0} \psi(h)/\|h\| = 0$ erfüllt ist. Es folgt $g((-v)+a+h) - w = g((-v)+a) - w + f'(a)(h) + \psi(h)$ und somit

$$g((-v)+a+h) \quad = \quad g((-v)+a) + f'(a)(h) + \psi(h)$$

für alle $h \in U_a$. Dies zeigt, dass g im Punkt $(-v)+a$ differenzierbar ist und $f'(a) = g'((-v)+a)$ gilt. □

(8.13) Satz. *(Kettenregel)*

Seien V, V', V'' endlich-dimensionale normierte $\mathbb{R}$-Vektorräume und $U \subseteq V$, $U' \subseteq V'$ offene Teilmengen. Seien $f : U \to V'$ und $g : U' \to V''$ Abbildungen mit $f(U) \subseteq U'$. Ferner sei $a \in U$ ein Punkt mit der Eigenschaft, dass f in a und g in $f(a)$ differenzierbar ist. Dann ist die Abbildung $g \circ f : U \to V''$ in a differenzierbar, und es gilt

$$(g \circ f)'(a) \quad = \quad g'(f(a)) \circ f'(a).$$

Beweis: Um die nachfolgenden Rechnungen zu vereinfachen, führen wir den Beweis zunächst auf die Situation zurück, dass $a = 0$, $f(0) = 0$ und $(g \circ f)(0) = 0$ gilt. Dazu setzen wir $\tilde{U} = (-a) + U = \{(-a) + x \mid x \in U\}$, $\tilde{U}' = (-f(a)) + U'$ und definieren $\tilde{f} : \tilde{U} \to V'$ und $\tilde{g} : \tilde{U}' \to V''$ durch $\tilde{f}(x) = f(a+x) - f(a)$ und $\tilde{g}(x) = g(f(a)+x) - (g \circ f)(a)$. Offenbar gilt dann $\tilde{f}(0) = 0$, $\tilde{g}(0) = 0$ und somit auch $(\tilde{g} \circ \tilde{f})(0) = 0$. Nach Lemma (8.12)

ist $\tilde{f}$ an der Stelle 0 differenzierbar, und es gilt $\tilde{f}'(0) = f'(a)$. Ebenso ist $\tilde{g}$ an der Stelle 0 differenzierbar, und es gilt $\tilde{g}'(0) = g'(f(a))$. Nehmen wir nun an, es wurde gezeigt, dass die Funktion $\tilde{g} \circ \tilde{f}$ im Nullpunkt differenzierbar ist und $(\tilde{g} \circ \tilde{f})'(0) = \tilde{g}'(0) \circ \tilde{f}'(0)$ gilt. Wegen $(\tilde{g} \circ \tilde{f})(0) = \tilde{g}(\tilde{f}(0)) = \tilde{g}(f(x+a) - f(a)) = (g \circ f)(x+a) - (g \circ f)(a)$ liefert eine erneute Anwendung von Lemma (8.12) die Gleichung $(\tilde{g} \circ \tilde{f})'(0) = (g \circ f)'(a)$. Insgesamt erhalten wir dann die gewünschte Gleichung $(g \circ f)'(a) = (\tilde{g} \circ \tilde{f})'(0) = \tilde{g}'(0) \circ \tilde{f}'(0) = g'(f(a)) \circ f'(a)$.

Zu zeigen ist also, dass unter der Voraussetzung von $\tilde{f}(0) = 0$, $\tilde{g}(0) = 0$ und der Differenzierbarkeit von $\tilde{f}$ und $\tilde{g}$ im Nullpunkt auch die Funktion $\tilde{g} \circ \tilde{f}$ im Nullpunkt differenzierbar ist und $(\tilde{g} \circ \tilde{f})'(0) = \tilde{g}'(0) \circ \tilde{f}'(0)$ gilt. Zur Vereinfachung der Notation bezeichnen wir $\tilde{f}$ und $\tilde{g}$ wieder mit f und g. Auf Grund der Differenzierbarkeit im Nullpunkt können wir f und g in der Form

$$\begin{aligned} f(h) &= f'(0)(h) + \varphi(h) \\ g(k) &= g'(0)(k) + \psi(k) \end{aligned}$$

schreiben, wobei die Restterme $\lim\limits_{h\to 0} \varphi(h)/\|h\| = 0$ und $\lim\limits_{k\to 0} \psi(k)/\|k\| = 0$ erfüllen. Setzen wir den Ausdruck für f in g ein, so erhalten wir

$$(g \circ f)(h) = g'(0)(f(h)) + \psi(f(h)) = g'(0)(f'(0)(h)) + g'(0)(\varphi(h)) + \psi(f(h)).$$

Zu zeigen ist, dass der Restterm $\rho(h) = g'(0)(\varphi(h)) + \psi(f(h))$ auch nach Division durch $\|h\|$ für $h \to 0$ noch gegen Null konvergiert, denn dann zeigt die Gleichung, dass $g'(0) \circ f'(0)$ die Ableitung von $g \circ f$ im Nullpunkt ist. Für den ersten Summanden $g'(0)(\varphi(h))$ gilt dies wegen

$$g'(0)(\varphi(h)) = \|h\| g'(0)\left(\frac{\varphi(h)}{\|h\|}\right)$$

und auf Grund der Stetigkeit von $g'(0)$ im Nullpunkt. Den zweiten Summanden schreiben wir als $\psi(f'(0)(h) + \varphi(h))$. Nach Voraussetzung gibt es Funktionen φ_1 und ψ_1 mit den Eigenschaften $\varphi(h) = \|h\|\varphi_1(h)$, $\psi(k) = \|k\|\psi_1(k)$ und $\lim\limits_{h\to 0} \varphi_1(h) = \lim\limits_{k\to 0} \psi_1(k) = 0$. Ferner gibt es eine Konstante $\gamma \in \mathbb{R}^+$ mit $\|f'(0)(h)\| \le \gamma\|h\|$ für alle $h \in V$. Setzen wir dies ein, so erhalten wir

$$\|\psi(f'(0)(h) + \varphi(h))\| = \|\psi_1(f'(0)(h) + \varphi(h))\| \|f'(0)(h) + \varphi(h)\| =$$

$$\|\psi_1(f'(0)(h) + \varphi(h))\| \|f'(0)(h) + \|h\|\varphi_1(h)\| \le \|\psi_1(f'(0)(h) + \varphi(h))\| \cdot \|h\| \cdot (\gamma + \|\varphi_1(h)\|).$$

Der Ausdruck im linken Faktor konvergiert für $h \to 0$ gegen Null, weil $f'(0)(h)$ und $\varphi(h)$ gegen Null laufen. Der rechte Faktor ist beschränkt, also konvergiert das gesamte Produkt auch nach Division durch $\|h\|$ noch gegen Null. □

Sei $f : \mathbb{R}^3 \to \mathbb{R}$, $(x,y,z) \mapsto (x+y+z)$ und $g : \mathbb{R} \to \mathbb{R}$, $z \mapsto z^3$. Dann sind die Funktionalmatrizen von f und g in jedem Punkt des Definitionsbereichs gegeben durch

$$f'(x,y,z) = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \qquad \text{und} \qquad g'(z) = \begin{pmatrix} 3z^2 \end{pmatrix}.$$

Die Anwendung der Kettenregel liefert

$$(g \circ f)'(x,y,z) = g'(f(x,y,z)) \circ f'(x,y,z) = g'(x+y+z) \circ f'(x,y,z) =$$

$$\begin{pmatrix} 3(x+y+z)^2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 3(x+y+z)^2 & 3(x+y+z)^2 & 3(x+y+z)^2 \end{pmatrix}.$$

Für alle $v = (v_1, v_2, v_3) \in \mathbb{R}^3$ gilt also $(g \circ f)'(x,y,z)(v) = 3(x+y+z)^2(v_1 + v_2 + v_3)$.

Als weiteres Anwendungsbeispiel seien $m, n \in \mathbb{N}$ vorgegeben und $f_1, ..., f_m : \mathbb{R} \to \mathbb{R}$ differenzierbare Funktionen. Außerdem sei $F : \mathbb{R}^m \to \mathbb{R}^n$ in jedem Punkt des Definitionsbereichs differenzierbar. Betrachten wir die f_i als Komponenten einer Funktion $f : \mathbb{R} \to \mathbb{R}^m$, dann ist die Ableitung im Punkt $x \in \mathbb{R}$ nach (8.3) gegeben durch

$$f'(x) \quad = \quad \begin{pmatrix} f_1'(x) \\ \vdots \\ f_m'(x) \end{pmatrix}.$$

Auf Grund der Kettenregel git für die Ableitung von $F \circ f : \mathbb{R} \to \mathbb{R}^n$ im Punkt x die Gleichung

$$(F \circ f)'(x) \quad = \quad F'(f(x)) \circ f'(x) \quad = \quad \begin{pmatrix} \partial_1 F_1(f(x)) & \cdots & \partial_m F_1(f(x)) \\ \vdots & & \vdots \\ \partial_1 F_n(f(x)) & \cdots & \partial_m F_n(f(x)) \end{pmatrix} \begin{pmatrix} f_1'(x) \\ \vdots \\ f_m'(x) \end{pmatrix}$$

$$= \quad \begin{pmatrix} \sum_{j=1}^{m} \partial_j F_1(f(x)) f_j'(x) \\ \vdots \\ \sum_{j=1}^{m} \partial_j F_n(f(x)) f_j'(x) \end{pmatrix}.$$

Ist speziell $n = 1$ und $F(x_1, ..., x_m) = F_1(x_1, ..., x_m) = x_1 + ... + x_m$, dann ist $(F \circ f)(x) = f_1(x) + ... + f_m(x)$. Die Einträge in der Funktionalmatrix von F sind dann alle gleich 1, und somit gilt $(F \circ f)'(x) = f_1'(x) + ... + f_m'(x)$.

Sei nun $n = 1$, $m = 2$ und $F(x_1, x_2) = x_1 x_2$. Dann gilt $(F \circ f)(x) = f_1(x) f_2(x)$. Die Funktionalmatrix von F ist gegeben durch

$$F'(x_1, x_2) \quad = \quad \begin{pmatrix} x_2 & x_1 \end{pmatrix} \quad ,$$

und es gilt

$$(F \circ f)'(x) \quad = \quad \begin{pmatrix} f_2(x) & f_1(x) \end{pmatrix} \begin{pmatrix} f_1'(x) \\ f_2'(x) \end{pmatrix} \quad = \quad \Big(f_2(x) f_1'(x) + f_1(x) f_2'(x) \Big).$$

Dies ist die gewöhnliche Produktregel für reellwertige Funktionen in einer Variablen.

Um den Beweis der nächsten Ableitungsregel vorzubereiten, zeigen wir

> **(8.14) Lemma.** Sei $U \subseteq \mathbb{R}^n$ offen und $f : U \to \mathbb{R}^m$ eine in $a \in U$ differenzierbare Abbildung. Dann gibt es eine offene Umgebung $U' \subseteq \mathbb{R}^n$ von $0_{\mathbb{R}^n}$ und eine in $0_{\mathbb{R}^n}$ stetige Abbildungen $\phi : U' \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ mit $\phi(0_{\mathbb{R}^n}) = f'(a)$ und
>
> $$f(a+h) \quad = \quad f(a) + \phi(h)(h) \qquad \text{für alle} \quad h \in U'.$$

Beweis: Auf Grund der Differenzierbarkeit von f in a gibt es eine Umgebung U' von $0_{\mathbb{R}^n}$ und eine Funktion $\psi : U' \to \mathbb{R}^m$ mit

$$f(a+h) \quad = \quad f(a) + f'(a)(h) + \psi(h) \qquad \text{und} \qquad \lim_{h \to 0} \frac{\psi(h)}{\|h\|_2} = 0 \quad ,$$

wobei $\|\cdot\|_2$ die euklidische Norm bezeichnet. Seien $\psi_1, ..., \psi_m : U' \to \mathbb{R}$ die Komponentenfunktionen von ψ. In jedem Punkt $h = (h_1, ..., h_n)$ von U' definieren wir die lineare Abbildung $\varrho(h) \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ durch $\varrho(h) = (r_{ij}(h))$ mit

$$r_{ij}(h) \quad = \quad \frac{\psi_i(h)}{\|h\|_2^2} h_j \qquad \text{für} \quad 1 \le i \le m, 1 \le j \le n$$

für $h \neq 0_{\mathbb{R}^n}$ und $r_{ij}(0_{\mathbb{R}^n}) = 0$. Für $h \to 0_{\mathbb{R}^n}$ konvergieren die Einträge dieser Matrix wegen $|h_j| \le \|h\|_2$ gegen Null, also ist ϱ im Nullpunkt stetig. Außerdem ist $\varrho(h)$ so konstruiert, dass $\varrho(h)(h) = \psi(h)$ für alle $h \in U'$ gilt, denn die i-te Komponente von $\varrho(h)(h)$ ist jeweils gegeben durch

$$\sum_{j=1}^{n} r_{ij}(h) h_j \quad = \quad \sum_{j=1}^{n} \frac{\psi_i(h)}{\|h\|_2^2} h_j^2 \quad = \quad \psi_i(h) \sum_{j=1}^{n} \frac{h_j^2}{\|h\|_2^2} \quad = \quad \psi_i(h).$$

Die Abbildung ϕ erhalten wir nun durch die Definition $\phi(h) = f'(a) + \varrho(h)$ für $h \in U'$. □

(8.15) Satz. *(Umkehrregel)*

Sei $U \subseteq \mathbb{R}^n$ offen und $f : U \to \mathbb{R}^n$ eine Abbildung mit der Eigenschaft, dass auch $V = f(U)$ offen in $\mathbb{R}^n$ ist und eine Umkehrabbildung $g : V \to \mathbb{R}^n$ von f existiert. Sei f in $a \in U$ differenzierbar, und es gelte $\det f'(a) \neq 0$. Schließlich setzen wir noch voraus, dass g in $b = f(a)$ stetig ist. Dann ist g in b differenzierbar, und es gilt $g'(b) = f'(a)^{-1}$.

Beweis: Wie beim Beweis der Kettenregel können wir $a = b = 0_{\mathbb{R}^n}$ voraussetzen. Auf Grund des Lemmas existiert eine offene Umgebung U' von $0_{\mathbb{R}^n}$ und Abbildung $\phi : U' \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$ mit $f(h) = \phi(h)(h)$ für alle $h \in U'$ und $\lim\limits_{h \to 0} \phi(h) = f'(0_{\mathbb{R}^n})$. Wegen $\det f'(0_{\mathbb{R}^n}) \neq 0$ können wir nach eventueller Verkleinerung von U' voraussetzen, dass $\det \phi(h) \neq 0$ für alle $h \in U'$ gilt. Also ist die lineare Abbildung $\phi(h)$ für alle $h \in U'$ invertierbar.

In der Linearen Algebra haben wir gezeigt, dass die Inverse A^{-1} einer invertierbaren Matrix in der Form $\det(A)^{-1}\tilde{A}$ dargestellt werden kann, wobei $\tilde{A}$ die zu A ***adjunkte*** Matrix bezeichnet. Die Einträge von $\tilde{A}$ sind dabei Determinanten gewisser Teilmatrizen von A. Stellen wir nun $\phi(h)$ und $\phi(h)^{-1}$ als Matrizen dar, dann sind die Komponenten von $\phi(h)^{-1}$ als Polynomausdrücke in den Komponenten von $\phi(h)$, dividiert durch $\det \phi(h)$, darstellbar. Dies zeigt, dass mit ϕ auch die Funktion $h \mapsto \phi(h)^{-1}$ im Nullpunkt stetig ist. Es gilt also $\lim\limits_{h \to 0} \phi(h)^{-1} = f'(0_{\mathbb{R}^n})^{-1}$.

Sei nun $V' = f(U')$, außerdem $k \in V'$ beliebig und $h = g(k)$. Dann gilt $k = f(h) = \phi(h)(h)$, und die Anwendung von $\phi(h)^{-1}$ auf beide Seiten der Gleichung liefert $\phi(h)^{-1}(k) = h$, insgesamt also

$$g(k) \quad = \quad h \quad = \quad \phi(h)^{-1}(k) \quad = \quad \phi(g(k))^{-1}(k).$$

Lassen wir k gegen $0_{\mathbb{R}^n}$ laufen, dann läuft $g(k)$ auf Grund der Stetigkeit im Nullpunkt ebenfalls gegen $0_{\mathbb{R}^n}$, woraus wiederum $\lim\limits_{k \to 0} \phi(g(k))^{-1} = f'(0_{\mathbb{R}^n})^{-1}$ folgt. Also können wir g in der Form

$$g(k) \quad = \quad f'(0_{\mathbb{R}^n})^{-1}(k) + \rho(k)(k)$$

darstellen, wobei die Abbildung $\rho(k) = \phi(g(k))^{-1} - f'(0_{\mathbb{R}^n})^{-1}$ für $k \to 0_{\mathbb{R}^n}$ in $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$ gegen Null konvergiert. Setzen wir $\psi(k) = \rho(k)(k)$, dann gilt also $g(k) = g(0_{\mathbb{R}^n}) + f'(0_{\mathbb{R}^n})^{-1}(k) + \psi(k)$ für alle $k \in U'$ sowie

$$\lim_{k \to 0} \frac{\psi(k)}{\|k\|} \quad = \quad \lim_{k \to 0} \rho(k) \left(\frac{k}{\|k\|} \right) \quad = \quad 0 \quad ,$$

denn der Quotient $k/\|k\|$ ist für $k \to 0_{\mathbb{R}^n}$ beschränkt. Also ist die Funktion g in $0_{\mathbb{R}^n}$ differenzierbar, und es gilt die Gleichung $g'(0_{\mathbb{R}^n}) = f'(0_{\mathbb{R}^n})^{-1}$. □

Auch zu dieser Ableitungsregel sehen wir uns ein Beispiel an. Dazu betrachten wir die in § 6 eingeführte Polarkoordinaten-Abbildung

$$\rho_{\text{pol}} : \mathbb{R}^+ \times]0, 2\pi[\longrightarrow \mathbb{R}^2 \setminus \{0_{\mathbb{R}^2}\} \qquad , \qquad (r, \varphi) \mapsto (r\cos(\varphi), r\sin(\varphi)).$$

Ist (r, φ) ein beliebiger Punkt im Definitionsbereich und $(x, y) = \rho_{\text{pol}}(r, \varphi)$, dann gilt also $x = r\cos(\varphi)$ und $y = r\sin(\varphi)$. Es folgt

$$r^2 \quad = \quad r^2(\cos(\varphi)^2 + \sin(\varphi)^2) \quad = \quad (r\cos(\varphi))^2 + (r\sin(\varphi))^2 \quad = \quad x^2 + y^2 \quad ,$$

also $r = \sqrt{x^2 + y^2}$. Außerdem gilt $\cos(\varphi) = \frac{x}{r}$ und $\sin(\varphi) = \frac{y}{r}$. Dies können wir nun verwenden, um mit Hilfe der Umkehrregel die Jacobi-Matrix von ρ_{pol}^{-1} im Punkt (x, y) anzugeben. Es gilt

$$\rho'_{\text{pol}}(r, \varphi) \quad = \quad \begin{pmatrix} \cos(\varphi) & -r\sin(\varphi) \\ \sin(\varphi) & r\cos(\varphi) \end{pmatrix}$$

und somit

$$(\rho_{\text{pol}}^{-1})'(x, y) \quad = \quad \rho'_{\text{pol}}(r, \varphi)^{-1} \quad = \quad \begin{pmatrix} \cos(\varphi) & -r\sin(\varphi) \\ \sin(\varphi) & r\cos(\varphi) \end{pmatrix}^{-1} \quad =$$

$$\begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\frac{1}{r}\sin(\varphi) & \frac{1}{r}\cos(\varphi) \end{pmatrix} \quad = \quad \begin{pmatrix} \dfrac{x}{\sqrt{x^2+y^2}} & \dfrac{y}{\sqrt{x^2+y^2}} \\ -\dfrac{y}{x^2+y^2} & \dfrac{x}{x^2+y^2} \end{pmatrix}$$

§ 9. Höhere Ableitungen und lokale Extremstellen

Zusammenfassung. In diesem Kapitel untersuchen wir die höheren Ableitungen $f^{(n)}$ für eine mehrdimensionale Funktionen $f : U \to W$, wobei W einen endlich-dimensionalen $\mathbb{R}$-Vektorraum und U eine offene Teilmenge eines weiteren endlich-dimensionalen $\mathbb{R}$-Vektorraums bezeichnet. Wie wir aus § 8 wissen, ist die Ableitung $f'(p)$ von f in einem Punkt $p \in U$ keine Zahl, sondern eine lineare Abbildung $V \to W$. Entsprechend ist auch $f''(p)$ keine Zahl, sondern eine bilineare Abbildung $V \times V \to W$. Allgemein ist $f^{(n)}(p)$ durch eine n-fach lineare (multilineare) Abbildung $f : V^n \to W$ gegeben.

Mit den höheren Ableitungen können wir f an jeder Stelle $p \in V$ ein ***Taylorpolynom*** $\tau_n(f, p)$ zuordnen, dass die Funktion f in einer Umgebung von p mit wachsender Genauigkeit approximiert. Wir verwenden dies zur Herleitung von notwendigen und hinreichenden Kriterien für Extrema, in Analogie zu den Sätzen aus der Schulmathematik bzw. der Analysis einer Variablen.

Wichtige Begriffe und Sätze in diesem Kapitel

- Definition der mehrfach (total) differenzierbaren Funktionen
- Definition der Hesse-Matrix einer zweifach differenzierbaren Funktion
- Approximation durch mehrdimensionale Taylor-Polynome
- lokale Extremstellen im Mehrdimensionalen, notwendige und hinreichende Kriterien

Auch in diesem Abschnitt bezeichnen V, W wieder endlich-dimensionale, normierte $\mathbb{R}$-Vektorräume. Sei $U \subseteq V$ offen und $f : U \to W$ eine Funktion, die in jedem Punkt $x \in U$ differenzierbar ist. Mit V und W ist auch der Raum $\mathscr{L}(V, W)$ der (stetigen) linearen Abbildungen $V \to W$ endlich-dimensional, außerdem ist $\mathscr{L}(V, W)$ auf natürliche Weise mit einer Norm ausgestattet, nämlich der in Abschnitt § 6 definierten Operatornorm. Durch die Zuordnung

$$f' : U \longrightarrow \mathscr{L}(V, W) \quad , \quad x \mapsto f'(x)$$

ist eine Abbildung zwischen der offenen Teilmenge $U \subseteq V$ und dem $\mathbb{R}$-Vektorraum $\mathscr{L}(V, W)$ gegeben.

(9.1) Definition. Ist die Abbildung f' auf U stetig, dann bezeichnen wir f als ***stetig differenzierbare*** Funktion. Ist sie darüber hinaus in jedem Punkt $x \in U$ differenzierbar, dann sprechen wir von einer ***zweimal differenzierbaren*** Funktion.

Die zweite Ableitung ist dann eine Abbildung von U in die Menge der linearen Abbildungen von V nach $\mathscr{L}(V, W)$, also eine Abbildung $f'' : U \to \mathscr{L}(V, \mathscr{L}(V, W))$. Ist auch diese Abbildung differenzierbar, dann nennt man f ***dreimal differenzierbar***. Die dritte Ableitung ist dann eine Abbildung $f''' : U \to \mathscr{L}(V, \mathscr{L}(V, \mathscr{L}(V, W)))$. Offenbar lässt sich dies beliebig fortsetzen. Wir erhalten auf diese Weise für jedes $n \in \mathbb{N}$ den Begriff der *n-fach differenzierbaren Funktion* und der *n-fachen Ableitung*. Definieren wir eine Folge $\hat{\mathscr{L}}^n(V, W)$ von $\mathbb{R}$-Vektorräumen rekursiv durch

$$\hat{\mathscr{L}}^1(V, W) \quad = \quad \mathscr{L}(V, W) \qquad \text{und} \qquad \hat{\mathscr{L}}^{n+1}(V, W) \quad = \quad \mathscr{L}(V, \hat{\mathscr{L}}^n(V, W))$$

dann ist die n-te Ableitung also eine Abbildung $f^{(n)} : U \to \hat{\mathscr{L}}^n(V, W)$.

Auf Grund der rekursiven Definition ist die Handhabung der Vektorräume $\hat{\mathscr{L}}^n(V,W)$ recht mühsam. Man ersetzt sie deshalb durch isomorphe $\mathbb{R}$-Vektorräume, deren Elemente sich leichter definieren lassen. Sei dazu V^n das n-fache kartesische Produkt $V \times ... \times V$. Wir bezeichnen eine Abbildung $\phi : V^n \to W$ als ***n*-fach linear** oder auch ***multilinear***, wenn sie in jeder ihrer n Komponenten linear ist. Das bedeutet, dass für jedes $k \in \{1,...,n\}$ und jedes Tupel $(v_1,...,v_{k-1},v_{k+1},...,v_n)$ von Vektoren aus V die Abbildung $w \mapsto \phi(v_1,...,v_{k-1},w,v_{k+1},...,v_n)$ linear ist. Man überprüft leicht, dass die n-fach linearen Abbildung $V^n \to W$ einen $\mathbb{R}$-Vektorraum bilden. Dieser wird von uns mit $\mathscr{L}^n(V,W)$ bezeichnet.

(9.2) Satz. Jedem Element $\hat{\phi} \in \hat{\mathscr{L}}^n(V,W)$ kann durch die Definition

$$\phi(v_1,...,v_n) \quad = \quad \hat{\phi}(v_1)(v_2)...(v_n)$$

ein Element in $\mathscr{L}^n(V,W)$ zugeordnet werden, und die Abbildung $\Phi : \hat{\mathscr{L}}^n(V,W) \to \mathscr{L}^n(V,W)$, $\hat{\phi} \mapsto \phi$ ist ein Isomorphismus von $\mathbb{R}$-Vektorräumen.

Beweis: Der Beweis ist nicht schwierig, erforder aber viel Schreibarbeit. Aus Gründen der Übersichtlichkeit verschieben wir ihn in einen Anhang zu diesem Kapitel. □

Sei $p \in \mathbb{N}_0$, sei $D \subseteq \mathbb{R}$ ein offenes Intervall, $a \in I$ und $f : I \to \mathbb{R}$ eine mindestens p-mal differenzierbare Funktion. Bereits in der Analysis einer Variablen haben wir das ***p*-te Taylorpolynom**

$$\tau_p(f,a)(x) \quad = \quad \sum_{k=0}^{p} \frac{1}{k!} f^{(k)}(a)(x-a)^k.$$

von f an der Stelle a definiert. Der Punkt a wird auch als ***Entwicklungspunkt*** des Taylorpolynoms bezeichnet. Das Polynom $\tau_p(f,a)$ ist dadurch ausgezeichnet, dass für $0 \leq k \leq p$ die k-te Ableitung von $\tau_p(f,a)$ an der Stelle a jeweils mit $f^{(k)}(a)$ übereinstimmt. Tatsächlich ist die k-te Ableitung von $\tau_p(f,a)$ gegeben durch

$$\tau_p(f,a)^{(k)}(x) \quad = \quad \sum_{\ell=k}^{p} \frac{1}{\ell!} f^{(k)}(a) \cdot k \cdot (k-1) \cdot ... \cdot (k-\ell+1) \cdot (x-a)^{k-\ell}$$

und somit $\tau_p(f,a)^{(k)}(a) = \frac{1}{k!} f^{(k)}(a) \cdot k! \cdot (x-a)^0 = f^{(k)}(a)$. Die Funktion f kann in einer kleinen Umgebung des Punktes a durch die Taylorpolynome beliebig genau approximiert werden.

(9.3) Satz. Sei $p \in \mathbb{N}$ und $x \in \mathbb{R}$ ein Punkt mit $x > a$, so dass f auf dem offenen Intervall $]a,x[$ mindestens $(p-1)$-mal stetig differenzierbar und p-mal differenzierbar ist. Dann gibt es einen Punkt $\xi \in]a,x[$ mit

$$f(x) \quad = \quad \tau_{p-1}(f,a)(x) + \frac{1}{p!} f^{(p)}(\xi)(x-a)^p.$$

Beweis: Sei $F : [a,x] \to \mathbb{R}$ definiert durch $F(t) = f(t) - \tau_{p-1}(f,a)(t)$ und $G(t) = (t-a)^p$. Auf Grund der soeben beschriebenen charakteristischen Eigenschaft des Taylorpolynoms gilt $F^{(k)}(a) = 0$ für $0 \le k \le p-1$. Außerdem gilt für $0 \le k \le p-1$ sowohl $G^{(k)}(a) = 0$ als auch $G(k)(t) \neq 0$ für alle $t \in]a,x]$. Auf Grund des verallgemeinerten Mittelwertsatzes der Analysis einer Variablen, Satz (14.7), gibt es ein $x_1 \in]a,x[$ mit

$$\frac{F(x)}{G(x)} \quad = \quad \frac{F(x)-F(a)}{G(x)-G(a)} \quad = \quad \frac{F'(x_1)}{G'(x_1)}.$$

Eine erneute Anwendung dieses Satzes liefert ein $x_2 \in]a,x_2[$ mit

$$\frac{F'(x_1)}{G'(x_1)} \quad = \quad \frac{F'(x_1)-F'(a)}{G'(x_1)-G'(a)} \quad = \quad \frac{F''(x_2)}{G''(x_2)}.$$

Die $(p-1)$-fache Wiederholung dieses Vorgangs liefert schließlich ein $\xi = x_p \in]a,x[$ mit

$$\frac{F(x)}{G(x)} \quad = \quad \frac{F'(x_1)}{G'(x_1)} \quad = \quad \frac{F''(x_2)}{G''(x_2)} \quad = \quad \cdots \quad = \quad \frac{F^{(p)}(x_p)}{G^{(p)}(x_p)}.$$

Es ist $G(x) = (x-a)^p$, die p-te Ableitung $G^{(p)}$ von G ist konstant gleich $p!$, und die p-te Ableitung auf $\tau_{p-1}(f,a)$ ist gleich null, weil das Polynom vom Grad $p-1$ ist. Nach Definition von F erhalten wir

$$(x-a)^{-p}(f(x)-\tau_{p-1}(f,a)(x)) \quad = \quad \frac{1}{p!}F^{(p)}(\xi) \quad = \quad \frac{1}{p!}f^{(p)}(\xi).$$

Durch Multiplikation mit $(x-a)^p$ und Umstellen der Gleichung ergibt sich die gewünschte Aussage. □

Wir verallgemeinern die Definition des Taylorpolynoms nun auf höhere Dimension.

(9.4) Definition. Sei $p \in \mathbb{N}$, $U \subseteq V$ offen, $a \in U$ und $f : U \to \mathbb{R}$ eine im Punkt a p-mal differenzierbare Funktion. Dann bezeichnet man die Funktion gegeben durch

$$\tau_p(f,a)(x) \quad = \quad f(a) + \sum_{k=1}^{p} \frac{1}{k!} f^{(k)}(a)(\underbrace{x-a,\ldots,x-a}_{k\text{-mal}})$$

für alle $x \in V$ als ***Taylorpolynom p-ten Grades*** von f an der Stelle a.

Um die Taylorpolynome explizit ausrechnen zu können, zeigen wir, wie die höheren Ableitungen $f^{(p)}(a)$ mit Hilfe der Richtungsableitungen ausgedrückt werden können. Beim folgenden Satz handelt es sich um eine weitreichende Verallgemeinerung von Prop. (8.4).

(9.5) Proposition. Sei $U \subseteq V$ offen, $a \in U$ und $f : U \to \mathbb{R}$ eine p-mal in a differenzierbare Funktion. Dann gilt für alle $v_1,\ldots,v_p \in V$ jeweils

$$f^{(p)}(a)(v_1,\ldots,v_p) \quad = \quad \partial_{v_1} \cdots \partial_{v_p} f(a).$$

Beweis: Wir beweisen die Aussage durch vollständige Induktion über $p \in \mathbb{N}_0$. Im Fall $p = 0$ besteht die Aussage lediglich in der trivialen Gleichung $f(a) = f(a)$. Sei nun $p \geq 1$, und setzen wir die die Gleichung für $p-1$ voraus. Definieren wir die Hilfsfunktionen $\tilde{f} = \partial_{v_2} \cdots \partial_{v_p} f$ und $g = f^{(p-1)}$, dann gilt auf Grund der Induktionsvoraussetzung $\tilde{f}(a+h) = g(a+h)(v_2, ..., v_p)$ für alle h in einer hinreichend kleinen Umgebung des Nullpunkts. Weil g im Punkt a total differenzierbar ist, gilt für hinreichend kleine $h \in V$ eine Gleichung der Form

$$g(a+h) \quad = \quad g(a) + g'(a)(h) + \psi(h) \quad ,$$

wobei wie immer ψ eine Funktion mit $\lim\limits_{h \to 0} \psi(h)/\|h\| = 0$ bezeichnet. Es folgt

$$\tilde{f}(a+h) \quad = \quad g(a+h)(v_2, ..., v_p) \quad = \quad g(a)(v_2, ..., v_p) + g'(a)(h, v_2, ..., v_p) + \varphi(h)$$

wobei die Funktion $\varphi(h) = \psi(h)(v_2, ..., v_p)$ ebenfalls die Bedingung $\lim_{h \to 0} \varphi(h)/\|h\| = 0$ erfüllt. Die Funktion $\tilde{f}$ ist also in a total differenzierbar, und es gilt $\tilde{f}'(a)(v_1) = g'(a)(v_1, v_2, ..., v_p)$ für alle $v \in V$. Nach Prop. (8.4) gilt $\partial_{v_1} \tilde{f}(a) = \tilde{f}'(a)(v_1)$. Damit erhalten wir das gewünschte Resultat, denn es gilt nun

$$\partial_{v_1} \partial_{v_2} \cdots \partial_{v_p} f(a) \quad = \quad \partial_{v_1} \tilde{f}(a) \quad = \quad \tilde{f}'(a)(v_1) \quad = \quad g'(a)(v_1, ..., v_p) \quad = \quad f^{(p)}(a)(v_1, ..., v_p). \qquad \square$$

(9.6) Satz. *(mehrdimensionale Taylor-Approximation)*

Sei $U \subseteq \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ eine p-mal stetig differenzierbare Funktion. Dann gibt es zu jedem Punkt $a \in U$ eine Umgebung U_0 des Nullpunkts und eine Funktion $\psi : U_0 \to \mathbb{R}$ mit

$$f(a+h) \quad = \quad \tau_p(f, a)(a+h) + \psi(h) \quad \text{für alle } h \in U_0 \quad \text{und} \quad \lim_{h \to 0} \frac{\psi(h)}{\|h\|^p} = 0.$$

Beweis: Seien $a \in U$ und $h \in \mathbb{R}^n$ so gewählt, dass die Verbindungsstrecke $[a, a+h]$ ganz in U liegt und $\phi : \mathbb{R} \to \mathbb{R}^n$, $t \mapsto a + th$. Setzen wir $I = \phi^{-1}(U)$, dann ist $I \subseteq \mathbb{R}$ eine offene Teilmenge mit $I \supseteq [0, 1]$. Außerdem sei $g : I \to U$ definiert durch $g = f \circ \phi$, also $g(t) = f(a + th)$ für alle $t \in I$. Es gilt $\phi'(t) = h$ für alle $t \in]0, 1[$, was wir mit der linearen Abbildung $\mathbb{R} \to \mathbb{R}^n$ gegeben durch $\lambda \mapsto \lambda h$ identifizieren. Mit der Kettenregel erhalten wir $g'(t) = (f \circ \phi)'(t) = f'(\phi(t)) \circ \phi'(t) = f'(a + th)(h) = \sum_{i=1}^n \partial_i f(a + th) h_i$. Wir zeigen nun durch vollständige Induktion über k, dass für $1 \leq k \leq p$ und alle $t \in]0, 1[$ jeweils

$$g^{(k)}(t) \quad = \quad \sum_{i_1}^n \cdots \sum_{i_k}^n \partial_{i_1 \ldots i_k} f(a + th)\, h_{i_1} \cdots h_{i_k} \qquad \text{gilt.}$$

Für $k = 1$ ist dies soeben geschehen. Ist $1 \leq k < p$, und setzen wir die Aussage für $1 \leq \ell \leq k$ voraus, so können wir $\tilde{f} = \sum_{i_1=1}^n \ldots \sum_{i_k=1}^n h_{i_1} \cdots h_{i_k} \partial_{i_1 \ldots i_k} f$ setzen. Nach Induktionsvoraussetzung gilt dann $g^{(k)}(t) = (\tilde{f} \circ \phi)(t)$. Durch erneute Anwendung der Kettenregel erhalten wir

$$g^{(k+1)}(t) \quad = \quad (\tilde{f} \circ \phi)'(t) \quad = \quad \tilde{f}'(a + th) \circ \phi'(t) \quad = \quad \sum_{i_0=1}^n \partial_{i_0} \tilde{f}(a + th)\, h_{i_0} \quad =$$

$$\sum_{i_0=1}^n \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n \partial_{i_0 i_1 \ldots i_k} f(a + th)\, h_{i_0} h_{i_1} \cdots h_{i_k} \quad = \quad \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_{k+1}=1}^n \partial_{i_1 \ldots i_{k+1}} f(a + th)\, h_{i_1} \cdots h_{i_{k+1}}$$

wobei im letzten Schritt lediglich umparametrisiert wurde. Damit ist der Induktionsbeweis abgeschlossen.

Sei nun $k \in \{1,...,p\}$. Auf Grund von Prop. (9.5), und weil die Abbildung $f^{(k)}(a) \in \mathscr{L}^k(V,\mathbb{R})$ in jedem ihrer Argumente linear ist, gilt weiter

$$\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n}\partial_{i_1\ldots i_k}f(a+th)\,h_{i_1}\cdots h_{i_k} \quad = \quad \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n}h_{i_1}\cdots h_{i_k}f^{(k)}(a+th)(e_{i_1},...,e_{i_k}) \quad =$$

$$f^{(k)}(a+th)\left(\sum_{i_1=1}^{n}h_{i_1}e_{i_1},...,\sum_{i_1=1}^{n}h_{i_k}e_{i_k}\right) \quad = \quad f^{(k)}(a+th)(h,...,h).$$

Damit haben wir für $1 \le k \le p$ die Gleichung $g^{(k)}(t) = f^{(k)}(a+th)(h,...,h)$ bewiesen.

Mit Satz (9.3), der eindimensionalen Taylorapproximation, erhalten wir nun ein $\xi(h) \in]0,1[$, so dass die Gleichung $g(1) = \sum_{k=0}^{p-1}\frac{1}{k!}g^{(k)}(0) + \frac{1}{p!}g^{(p)}(\xi(h))$ erfüllt ist; die Summe erhält man dabei durch Einsetzen von 1 in das $(p-1)$-te Taylorpolynom $\tau_{p-1}(g,0)$ von g an der Stelle 0. Es folgt

$$f(a+h) = g(1) = \sum_{k=0}^{p-1}\frac{1}{k!}g^{(k)}(0) + \frac{1}{p!}g^{(p)}(\xi(h)) = f(a) + \sum_{k=1}^{p-1}\frac{1}{k!}f^{(k)}(a)(h,...,h) + \frac{1}{p!}f^{(p)}(a+\xi(h)h)(h,...,h).$$

Das mehrdimensionale Taylorpolynom an der Stelle a, ausgewertet im Punkt $a+h$, hat die Form

$$\tau_p(f,a)(x) \quad = \quad f(a) + \sum_{k=1}^{p}\frac{1}{k!}f^{(k)}(a)(h,...,h) \quad = \quad f(a) + \sum_{k=1}^{p-1}\frac{1}{k!}f^{(k)}(a)(h,...,h) + \frac{1}{p!}f^{(p)}(a)(h,...,h).$$

Der Fehlerterm in der Aussage unseres Satzes ist also gegeben durch die Differenz

$$\psi(h) \quad = \quad \frac{1}{p!}f^{(p)}(a+\xi(h)h)(h,...,h) - \frac{1}{p!}f^{(p)}(a)(h,...,h) \quad =$$

$$\frac{1}{p!}\sum_{i_1=1}^{n}\cdots\sum_{i_p=1}^{n}\left(\partial_{i_1\ldots i_p}f(a+\xi(h)h) - \partial_{i_1\ldots i_p}f(a)\right)h_{i_1}...h_{i_p}.$$

Legen wir die $\|\cdot\|_\infty$-Norm zu Grunde, so gilt die Abschätzung $|h_{i_1}...h_{i_p}| \le \|h\|^p$. Weil unsere Funktion p-mal stetig differenzierbar ist, sind die p-fachen partiellen Ableitungen $\partial_{i_1\ldots i_p}f$ stetig. Für $h \to 0$ gilt $\xi(h)h \to 0$ wegen $\xi(h) \in]0,1[$ für alle h; somit läuft $\partial_{i_1\ldots i_p}(f(a+\xi(h)h) - f(a))$ auch gegen Null. Insgesamt ist damit $\|h\|^{-p}\psi(h) \to 0$ für $h \to 0$ nachgewiesen. □

Im weiteren Verlauf werden wir uns vor allem die Fälle $p = 1,2$ konzentrieren. Wie wir bereits in § 11 gesehen haben, kann die Ableitung $f'(a)$ als $(n \times 1)$-Matrix dargestellt werden, mit den Einträgen $(\partial_1 f(a) \cdots \partial_n f(a))$. Die zweite Ableitung $f''(a)$ ist eine bilineare Abbildung $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, also eine Bilinearform auf dem $\mathbb{R}^n$.

Die Darstellungsmatrix $A \in \mathscr{M}_{n,\mathbb{R}}$ dieser Bilinearform hat nach (9.5) die Einträge $a_{ij} = f''(a)(e_i,e_j) = \partial_{ij}f(a)$ für $1 \le i,j \le n$ und wird die ***Hesse-Matrix*** von f an der Stelle a genannt. Wir bezeichnen sie auch mit $\mathscr{H}(f)(a)$. Die Taylorpolynome für $p = 1,2$ sind also gegeben durch

$$\tau_1(f,a)(x) \quad = \quad f(a) + f'(a)(x-a) \quad = \quad f(a) + f'(a)\cdot(x-a) \quad = \quad f(a) + \sum_{i=1}^{n}\partial_i f(a)\cdot(x_i - a_i)$$

und

$$\tau_2(f,a)(x) \quad = \quad f(a) + f'(a)(x-a) + f''(a)(x-a,x-a) \quad =$$

$$f(a) + f'(a)\cdot(x-a) + \tfrac{1}{2}\,{}^{\mathrm{t}}(x-a)\cdot\mathscr{H}(f)(a)\cdot(x-a) \quad =$$

$$f(a) + \sum_{i=1}^{n}\partial_i f(a)\cdot(x_i-a_i) + \tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\partial_{ij}f(a)(x_i-a_i)(x_j-a_j).$$

Als für die Anwendungen wichtigen Spezialfall von Satz (9.6) notieren wir

(9.7) Folgerung. Sei $U \subseteq \mathbb{R}^n$ offen, $f : U \to \mathbb{R}$ eine Funktion und $a \in U$ ein Punkt, in dem f zweimal differenzierbar ist. Dann gibt es eine Funktion $\psi : U_a \to \mathbb{R}$ mit

$$f(a+h) = f(a) + f'(a)\cdot h + \tfrac{1}{2}\,{}^t h \cdot \mathscr{H}(f)(a)\cdot h + \psi(h) \qquad \text{und} \qquad \lim_{h\to 0} \|h\|^{-2}\,\psi(h) = 0.$$

Zur Illustration der bisher behandelten Konzepte berechnen wir das zweite Taylorpolynom der Polynomfunktion $f(x,y) = x^2+3y^2-7xy+5x-3y+8$ an der Stelle $(3,5)$. Die partiellen Ableitungen von f bis zur zweiten Ordnung sind gegeben durch

$$\begin{aligned} \partial_1 f(x,y) &= 2x-7y+5 \\ \partial_2 f(x,y) &= -7x+6y-3 \\ \partial_{11} f(x,y) = 2 \quad &, \quad \partial_{12} f(x,y) = \partial_{21} f(x,y) = -7 \quad , \quad \partial_{22} f(x,y) = 6. \end{aligned}$$

Funktionswert, erste Ableitung (Jacobi-Matrix) und zweite Ableitung (Hesse-Matrix) an der Stelle $(3,5)$ sind gegeben durch

$$f(3,5) = -13 \quad , \quad f'(3,5) = \begin{pmatrix} -24 & 6 \end{pmatrix} \quad , \quad \mathscr{H}(f)(3,5) = \begin{pmatrix} 2 & -7 \\ -7 & 6 \end{pmatrix}.$$

Das zweite Taylorpolynom an der Stelle $(3,5)$ ist also gegeben durch

$$\begin{aligned} \tau_2(f,(3,5))(x,y) &= f(3,5) + f'(3,5)\cdot \begin{pmatrix} x-3 \\ y-5 \end{pmatrix} + \tfrac{1}{2}\begin{pmatrix} x-3 & y-5 \end{pmatrix} \mathscr{H}(f)(3,5)\cdot \begin{pmatrix} x-3 \\ y-5 \end{pmatrix} \\ &= (-13) + \begin{pmatrix} -24 & 6 \end{pmatrix}\begin{pmatrix} x-3 \\ y-5 \end{pmatrix} + \tfrac{1}{2}\begin{pmatrix} x-3 & y-5 \end{pmatrix}\begin{pmatrix} 2 & -7 \\ -7 & 6 \end{pmatrix}\begin{pmatrix} x-3 \\ y-5 \end{pmatrix} \\ &= (-13) + (-24x+6y+42) + \tfrac{1}{2}(2x^2-14xy+58x+6y^2-18y-42) \quad = \quad x^2+3y^2-7xy+5x-3y+8. \end{aligned}$$

Die Funktion f stimmt also mit ihrem zweiten Taylorpolynom an der Stelle $(3,5)$ überein. Dies bedeutet, dass (9.7) für $a = (3,5)$ mit dem Fehlerterm $\psi = 0$ erfüllt ist. Man kann zeigen, dass allgemein jede Polynomfunktion auf dem $\mathbb{R}^n$ für hinreichend großes p gleich ihrem Taylorpolynom $\tau_p(f,a)$ ist, wobei der Entwicklungspunkt $a \in \mathbb{R}^n$ beliebig gewählt werden kann.

(9.8) Definition. Sei $U \subseteq \mathbb{R}^n$ eine beliebige Teilmenge, $a \in U$ und $f : U \to \mathbb{R}$ eine reellwertige Funktion. Man sagt, f hat im Punkt a ein

(i) ***lokales Maximum***, wenn eine Umgebung $U' \subseteq \mathbb{R}^n$ von a existiert, so dass $f(a) \geq f(x)$ für alle $x \in U \cap U'$ gilt,

(ii) ***isoliertes lokales Maximum***, wenn U so gewählt werden kann, dass sogar $f(a) > f(x)$ für alle $x \in (U \cap U') \setminus \{a\}$ erfüllt ist.

Entsprechend definiert man lokale Minima und isolierte lokale Minima. Wie in der Analysis einer Variablen verwenden wir den Begriff ***Extremum*** als Oberbegriff für Minima und Maxima.

In diesem Abschnitt betrachten wir nur Extrema von Funktionen, die auf *offenen* Teilmengen des $\mathbb{R}^n$ definiert sind. Wie in der Analysis einer Variablen zeigen wir

(9.9) Satz. *(notwendiges Kriterium für Extrema)*

Sei $U \subseteq \mathbb{R}^n$ offen, $f : U \to \mathbb{R}$ differenzierbar und $a \in U$ ein lokales Extremum von f.
Dann gilt $f'(a) = 0$.

Beweis: Wir behandeln nur den Fall, dass f an der Stelle a ein lokales Maximum besitzt. Nach eventueller Verkleinerung von U können wir $f(a) \geq f(x)$ für alle $x \in U$ annehmen. Sei $k \in \{1, ..., n\}$ und die Funktion g für alle $t \in \mathbb{R}$ mit hinreichend kleinem Betrag durch $g(t) = f(a + te_k)$ definiert. Dann gilt $g(0) \geq g(t)$ für diese t und außerdem $g'(0) = \partial_k f(a)$. Die Funktion g besitzt also in 0 ein lokales Extremum. Wie in der Analysis einer Variablen gezeigt wurde, folgt daraus $\partial_k f(a) = g'(0) = 0$. Es gilt also $\partial_k f(a) = 0$ für $1 \leq k \leq n$, und daraus folgt $f'(a) = 0$. □

Einen Punkt, in dem die erste Ableitung einer Funktion f verschwindet, bezeichnet man als ***kritische Stelle*** von f. Wie in der Analysis einer Variablen gibt es auch hinreichende Kriterien für Extrema, die mit der zweiten Ableitung der Funktion zusammenhängen. Nach Definition ist eine Matrix positiv definit, wenn ${}^t vAv > 0$ für alle $v \in \mathbb{R}^n \setminus \{0\}$ gilt.

(9.10) Definition. Wir bezeichnen eine Matrix $A \in \mathscr{M}_{n,\mathbb{R}}$ als

(i) ***negativ definit***, wenn ${}^t vAv < 0$ für alle $v \in \mathbb{R}^n \setminus \{0\}$, als

(ii) ***positiv semidefinit***, wenn ${}^t vAv \geq 0$ für alle $v \in \mathbb{R}^n$, und als

(iii) ***negativ semidefinit***, wenn ${}^t vAv \leq 0$ für alle $v \in \mathbb{R}^n$ erfüllt ist.

Eine Matrix, die weder positiv noch negativ semidefinit ist, bezeichnen wir als ***indefinit***.

Die indefiniten Matrizen A sind dadurch gekennzeichnet, dass es Vektoren $v, w \in \mathbb{R}^n$ mit ${}^t vAv > 0$ und ${}^t wAw < 0$ gibt. Offenbar ist A genau dann negativ definit, wenn $-A$ positiv definit ist. Deshalb kann auch die Eigenschaft „negativ definit“ mit dem Hurwitz-Kriterium getestet werden. Für positive oder negative Semidefinitheit gibt es leider kein so einfaches Kriterium.

(9.11) Satz. *(hinreichende Kriterien für Extrema)*

Sei $U \subseteq \mathbb{R}^n$ offen, $f : U \to \mathbb{R}$ eine zweimal stetig differenzierbare Funktion und $a \in U$ eine kritische Stelle von f.

(i) Ist $\mathscr{H}(f)(a)$ positiv definit, dann besitzt f in a ein isoliertes lokales Minimum.

(ii) Ist $\mathscr{H}(f)(a)$ negativ definit, dann besitzt f in a ein isoliertes lokales Maximum.

(iii) Ist $\mathscr{H}(f)(a)$ indefinit, dann hat f in a kein lokales Extremum.

Beweis: Auf Grund der Voraussetzungen ist (9.7) anwendbar. Wie dort bezeichnen wir den Fehlerterm mit ψ, und die Hesse-Matrix $\mathscr{H}(f)(a) \in \mathscr{M}_{n,\mathbb{R}}$ mit A.

zu (i) Weil A positiv definit ist, gilt ${}^t hAh > 0$ für alle $h \in \mathbb{R}^n$ mit $v \neq 0_{\mathbb{R}^n}$. Nach dem Maximumsprinzip (5.17), angewendet auf die kompakte Menge $S = \{h \in \mathbb{R}^n \mid \|h\| = 1\}$, besitzt die Abbildung $S \to \mathbb{R}$, $h \mapsto {}^t hAh$ ein positives Minimum m. Wegen $\lim_{h\to 0} \|h\|^{-2}\psi(h) = 0$ gibt es ein $\varepsilon \in \mathbb{R}^+$, so dass $|\psi(h)| \leq \frac{1}{4}m\|h\|^2$ für alle $h \in \mathbb{R}^n$ mit $0 < \|h\| < \varepsilon$ gilt. Definieren wir $h_0 = \|h\|^{-1}h$ für ein solches h, dann ist h_0 in S enthalten. Sofern $a+h$ in U liegt, gilt somit

$$\begin{aligned} f(a+h) &= f(a) + \tfrac{1}{2}\,{}^t hAh + \psi(h) = f(a) + \tfrac{1}{2}\|h\|^2\,{}^t h_0 Ah_0 + \psi(h) \\ &\geq f(a) + \tfrac{1}{2}m\|h\|^2 - \tfrac{1}{4}m\|h\|^2 = f(a) + \tfrac{1}{4}m\|h\|^2 > f(a). \end{aligned}$$

Es gilt also $f(x) > f(a)$ für alle $x \in B_\varepsilon(a) \cap U$. Dies zeigt, dass f in a ein isoliertes lokales Minimum besitzt.

zu (ii) Der Beweis dieser Aussage kann durch Betrachtung der Funktion $-f$ auf (i) zurückgeführt werden. Ist nämlich $A = \mathscr{H}(f)(a)$ negativ definit, dann ist $\mathscr{H}(-f)(a) = -A$ positiv definit, und ein isoliertes lokales Minimum von $-f$ ist ein isoliertes lokales Maximum von f.

zu (iii) Auf Grund der Voraussetzung gibt es Vektoren $v, w \in \mathbb{R}^n$ mit ${}^t vAv > 0$ und ${}^t wAw < 0$. Wegen der Eigenschaft $\lim_{h\to 0} \|h\|^{-2}\psi(h) = 0$ des Fehlerterms und wegen $\|tv\|^{-2} = t^{-2}\|v\|^{-2}$ und $\|tw\|^{-2} = t^{-2}\|w\|^{-2}$ für $t \neq 0$ gilt insbesondere auch $\lim_{t\to 0} t^{-2}\psi(tv) = 0$ und $\lim_{t\to 0} t^{-2}\psi(tw) = 0$. Für jede Konstante $\gamma \in \mathbb{R}^+$ existiert somit ein $\varepsilon \in \mathbb{R}^+$ mit der Eigenschaft, dass $|\psi(th)| < \gamma t^2$ für $h \in \{v, w\}$ und alle $t \in \mathbb{R}$ mit $|t| < \varepsilon$ gilt. Insbesondere können wir $\varepsilon \in \mathbb{R}^+$ so klein wählen, $|\psi(th)| < \frac{1}{4}|{}^t(th)A(th)| = \frac{1}{4}t^2|{}^t hAh|$ für $h \in \{v, w\}$ und alle $t \in \mathbb{R}$ mit $0 < |t| < \varepsilon$ erfüllt ist. Ist nun $t \in \mathbb{R}$ so gewählt, neben dieser Bedingung auch $a + tv \in U$ und $a + tw \in U$ gilt, dann erhalten wir sowohl

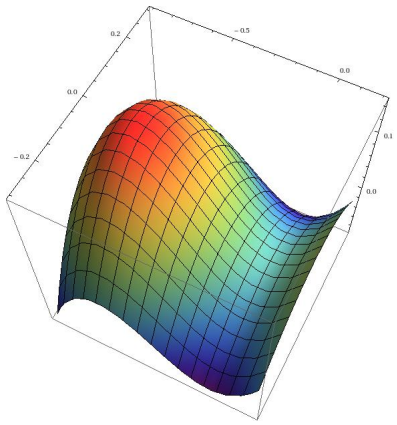
$$f(a+tv) = f(a) + \tfrac{1}{2}\,{}^t(tv)A(tv) + \psi(tv) \geq f(a) + \tfrac{1}{2}t^2\,{}^t vAv - \tfrac{1}{4}t^2\,{}^t vAv = f(a) + \tfrac{1}{4}t^2\,{}^t vAv > f(a)$$

als auch

$$\begin{aligned} f(a+tw) &= f(a) + \tfrac{1}{2}\,{}^t(tw)A(tw) + \psi(tw) \leq f(a) + \tfrac{1}{2}t^2\,{}^t wAw - \tfrac{1}{4}t^2\,{}^t wAw \\ &= f(a) + \tfrac{1}{4}t^2\,{}^t wAw < f(a). \end{aligned}$$

Dies zeigt, dass in jeder beliebig kleinen Umgebung von a Punkte x, y existieren mit $f(x) > f(a)$ und $f(y) < f(a)$. Dies zeigt, dass f an der Stelle a kein lokales Extremum besitzt. □

Wir betrachten die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch $f(x, y) = y^2(x-1) + x^2(x+1)$.



Graph der Funktion f
Man erkennt deutlich das lokale Maximum (rot) an der Stelle $(-\frac{2}{3}, 0)$ und die kritische Stelle $(0, 0)$ im blaugrünen Bereich, an der kein lokales Extremum existiert.

Die erste Ableitung und die Hesse-Matrix von f sind gegeben durch

$$f'(x,y) = (y^2+3x^2+2x \;\; 2(x-1)y) \qquad \text{und} \qquad \mathscr{H}(f)(x,y) = \begin{pmatrix} 6x+2 & 2y \\ 2y & 2(x-1) \end{pmatrix}.$$

Zunächst bestimmen wir die kritischen Punkte der Funktion, also die Punkte $(x,y) \in \mathbb{R}^2$ mit

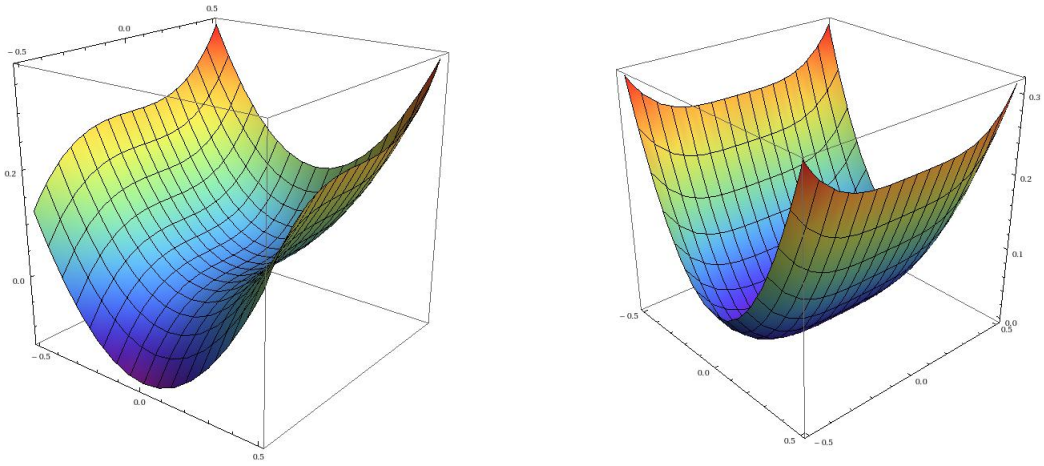
$$f'(x,y) \quad = \quad (y^2+3x^2+2x \;\; 2(x-1)y) \quad = \quad (0 \;\; 0).$$

Wegen $2(x-1)y = 0$ muss $x = 1$ oder $y = 0$ sein. Ist $x = 1$, dann ist die zweite Komponente gleich y^2+5 und kann somit nicht Null sein. Also bleibt nur die Möglichkeit $y = 0$ übrig, und wir erhalten für die erste Komponente den Ausdruck $3x^2+2x = 3x(x+\frac{2}{3})$. Dieser ist genau dann gleich Null, wenn $x = 0$ oder $x = -\frac{2}{3}$ ist. Die beiden kritischen Punkte sind also $(0,0)$ und $(-\frac{2}{3},0)$. Die Hesse-Matrizen in diesen beiden Punkten sind gegeben durch

$$\mathscr{H}(f)(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \qquad \text{und} \qquad \mathscr{H}(f)(-\tfrac{2}{3},0) = \begin{pmatrix} -2 & 0 \\ 0 & -\frac{10}{3} \end{pmatrix}.$$

Die Matrix $\mathscr{H}(f)(0,0)$ ist indefinit, denn für die Einheitsvektor $e_1 = (1,0)$, $e_2 = (0,1)$ gilt ${}^t e_1 \mathscr{H}(f)(0,0) e_1 = 2 > 0$ und ${}^t e_2 \mathscr{H}(f)(0,0) e_2 = -2 < 0$. Also liegt nach (9.11) (iii) an der Stelle $(0,0)$ kein lokales Extremum vor. Andererseits ist $\mathscr{H}(f)(-\frac{2}{3},0)$ negativ definit, denn für einen beliebigen Vektor $v = (x,y) \neq (0,0)$ gilt ${}^t v \mathscr{H}(f)(-\frac{2}{3},0) v = -2x^2 - \frac{10}{3}y^2 < 0$. Nach (9.11) (ii) besitzt f in $(-\frac{2}{3},0)$ also ein isoliertes lokales Maximum.

Ist die Hesse-Matrix an einer kritischen Stelle a positiv oder negativ ***semidefinit***, erfüllt die zweite Ableitung also nur $f''(a)(v,v) \geq 0$ für alle $v \in \mathbb{R}^n$ oder $f''(a)(v,v) \leq 0$ für alle $v \in \mathbb{R}^n$, dann ist keine allgemeine Aussage über das Auftreten von Extrema möglich. Als Beispiel betrachten wir die Funktionen f und g definiert durch $f(x,y) = x^2+y^3$ und $g(x,y) = x^2+y^4$.



Funktionsgraphen von f und g
Beide Graphen besitzen an der Stelle $(0,0)$ einen kritischen Punkt, und die Hessematrix ist in beiden Fällen positiv semidefinit. Aber nur die Funktion g besitzt in $(0,0)$ ein lokales Minimum.

Es gilt $f'(x,y) = (2x \;\; 3y^2)$ und $g'(x,y) = (2x \;\; 4y^3)$, also ist $(0,0)$ für beide Funktionen eine kritische Stelle. Es gilt

$$\mathscr{H}(f)(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 6y \end{pmatrix} \qquad \text{und} \qquad \mathscr{H}(g)(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 12y^2 \end{pmatrix} \quad ,$$

also insbesondere

$$\mathscr{H}(f)(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{und} \qquad \mathscr{H}(g)(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \quad ,$$

d.h. die Hesse-Matrix ist bei $(0,0)$ jeweils positiv semidefinit. Die Funktion g besitzt bei a offenbar ein isoliertes lokales Minimum, denn an allen Stellen $\neq (0,0)$ nimmt g nur positive Werte an. Dagegen ist $f(\varepsilon, 0) = \varepsilon^2$ für beliebig kleine $\varepsilon \in \mathbb{R}^+$ positiv, während $f(0,-\varepsilon) = -\varepsilon^3$ negativ ist. Also besitzt f bei $(0,0)$ kein lokales Extremum.

Wenigstens ist die positive Semidefinitheit ein *notwendiges* Kriterium für das Auftreten lokaler Minima.

(9.12) Satz. Sei $U \subseteq \mathbb{R}^n$ offen, $f : U \to \mathbb{R}$ zweimal stetig differenzierbar, $a \in U$ ein kritischer Punkt von f und $A = \mathscr{H}(f)(a)$.

(i) Besitzt f bei a ein lokales Minimum, dann ist A positiv semidefinit.

(ii) Besitzt f bei a ein lokales Maximum, dann ist A negativ semidefinit.

Beweis: Wieder genügt es, die Aussage (i) zu beweisen, weil (ii) durch Übergang zu $-f$ auf (i) zurückgeführt werden kann. Nehmen wir nun an, dass a ein lokales Minimum von f, die Matrix A aber nicht positiv semidefinit ist. Dann gibt es eine Vektor $v \in \mathbb{R}^n$ mit ${}^t vAv < 0$. Wie im Beweis von (9.11) (iii) zeigt man, dass in jeder beliebig kleinen Umgebung von a ein Punkt x mit $f(x) < f(a)$ existiert. Aber dies widerspricht der Annahme, dass f in a ein lokales Minimum besitzt. □

Anhang: Der Beweis von Satz (9.2)

Wir zeigen durch vollständige Induktion über n, dass die angegebene Abbildung Φ ein Homomorphismus von $\mathbb{R}$-Vektorräumen ist, und dass eine Abbildung $\mathscr{L}^n(V,W) \to \hat{\mathscr{L}}^n(V,W)$ existiert, die jedem $\phi \in \mathscr{L}^n(V,W)$ ein $\hat{\phi} \in \hat{\mathscr{L}}^n(V,W)$ mit $\hat{\phi}(v_1)\dots(v_n) = \phi(v_1,\dots,v_n)$ für alle $v_1,\dots,v_n \in V$ zuordnet. Daraus ergibt sich, dass die beiden Abbildungen zueinander invers sind und insgesamt ein Isomorphismus von $\mathbb{R}$-Vektorräumen vorliegt.

Für $n = 1$ gilt $\hat{\mathscr{L}}^n(V,W) = \mathscr{L}^n(V,W)$, und Φ ist die Identität. In diesem Fall braucht also nichts gezeigt werden. Sei nun $n \in \mathbb{N}$, und setzen wir die Aussagen für dieses n voraus. Sei $\hat{\phi}$ ein Element aus $\hat{\mathscr{L}}^{n+1}(V,W) = \mathscr{L}(V,\hat{\mathscr{L}}^n(V,W))$, und sei $\phi = \Phi(\hat{\phi})$ wie angegeben definiert. Für jedes $v_1 \in V$ definieren wir $\hat{\phi}_{v_1} \in \hat{\mathscr{L}}^n(V,W)$ durch $\hat{\phi}_{v_1} = \hat{\phi}(v_1)$. Nach Induktionsvoraussetzung ist die Abbildung $\phi_{v_1} : V^n \to W$ gegeben durch $\phi_{v_1}(v_2,\dots,v_{n+1}) = \hat{\phi}_{v_1}(v_2)\dots(v_{n+1})$ in $\mathscr{L}^n(V,W)$ enthalten, und nach Definition gilt

$$\phi(v_1,\dots,v_{n+1}) \quad = \quad \hat{\phi}(v_1)\dots(v_{n+1}) \quad = \quad \hat{\phi}_{v_1}(v_2)\dots(v_{n+1}) \quad = \quad \phi_{v_1}(v_2,\dots,v_{n+1}).$$

Wegen $\phi_{v_1} \in \mathscr{L}^n(V,W)$ ist ϕ also für $2 \le k \le n+1$ jeweils in der k-ten Komponente linear. Auch in der ersten Komponente ist ϕ linear. Weil nämlich $\hat{\phi}$ in $\hat{\mathscr{L}}^{n+1}(V,W) = \mathscr{L}(V,\hat{\mathscr{L}}^n(V,W))$ liegt, ist die Abbildung $v_1 \mapsto \hat{\phi}_{v_1}$ linear. Daraus folgt für alle $v_1, v_1' \in V$, $v_2,\dots,v_{n+1} \in V$ und $\lambda \in \mathbb{R}$ sowohl

$$\begin{gathered}\phi(v_1+v_1',v_2,\dots,v_{n+1}) \quad = \quad \phi_{v_1+v_1'}(v_2,\dots,v_{n+1}) \quad = \quad \hat{\phi}_{v_1+v_1'}(v_2)\dots(v_{n+1}) \quad = \quad (\hat{\phi}_{v_1}+\hat{\phi}_{v_1'})(v_2)\dots(v_{n+1}) \quad = \\ \hat{\phi}_{v_1}(v_2)\dots(v_{n+1}) + \hat{\phi}_{v_1'}(v_2)\dots(v_{n+1}) \quad = \quad \phi_{v_1}(v_2,\dots,v_{n+1}) + \phi_{v_1'}(v_2,\dots,v_{n+1}) \\ = \quad \phi(v_1,v_2,\dots,v_{n+1}) + \phi(v_1,v_2,\dots,v_{n+1})\end{gathered}$$

als auch

$$\begin{gathered}\phi(\lambda v_1,v_2,\dots,v_{n+1}) \quad = \quad \phi_{\lambda v_1}(v_2,\dots,v_{n+1}) \quad = \quad \hat{\phi}_{\lambda v_1}(v_2)\dots(v_{n+1}) \quad = \quad (\lambda\hat{\phi}_{v_1})(v_2)\dots(v_{n+1}) \\ = \quad \lambda\hat{\phi}_{v_1}(v_2)\dots(v_{n+1}) \quad = \quad \lambda\phi_{v_1}(v_2,\dots,v_{n+1}) \quad = \quad \lambda\phi(v_1,v_2,\dots,v_{n+1}).\end{gathered}$$

Insgesamt ist ϕ also in allen $n+1$ Komponenten linear und damit ein Element von $\mathscr{L}^{n+1}(V,W)$. Darüber hinaus ist die Abbildung Φ ist linear. Denn für beliebige $\hat{\phi},\hat{\psi} \in \hat{\mathscr{L}}^{n+1}(V,W)$ und $\lambda \in \mathbb{R}$ gilt jeweils

$$\begin{gathered}(\phi+\psi)(v_1,v_2,\dots,v_{n+1}) \quad = \quad \phi(v_1,v_2,\dots,v_{n+1}) + \psi(v_1,v_2,\dots,v_{n+1}) \quad = \\ \hat{\phi}(v_1)(v_2)\dots(v_{n+1}) + \hat{\psi}(v_1)(v_2)\dots(v_{n+1}) \quad = \quad (\hat{\phi}+\hat{\psi})(v_1)(v_2)\dots(v_{n+1})\end{gathered}$$

und

$$(\lambda\phi)(v_1,v_2,\dots,v_{n+1}) \quad = \quad \lambda\phi(v_1,v_2,\dots,v_{n+1}) \quad = \quad \lambda\hat{\phi}(v_1)(v_2)\dots(v_{n+1}) \quad = \quad (\lambda\hat{\phi})(v_1)(v_2)\dots(v_{n+1})$$

für beliebige $v_1,\dots,v_{n+1} \in V$. Diese Gleichungen zeigen, dass tatsächlich $\Phi(\hat{\phi}+\hat{\psi}) = \phi+\psi = \Phi(\hat{\phi})+\Phi(\hat{\psi})$ und $\Phi(\lambda\hat{\phi}) = \lambda\phi = \lambda\Phi(\hat{\phi})$ erfüllt ist.

Nun beweisen wir noch die Aussage zur Umkehrabbildung. Sei $\phi \in \mathscr{L}^{n+1}(V,W)$. Dann ist für jedes $v_1 \in V$ die Abbildung $\phi_{v_1} : V^n \to W$ gegeben durch $\phi_{v_1}(v_2,\dots,v_{n+1}) = \phi(v_1,v_2,\dots,v_{n+1})$ in $\mathscr{L}^n(V,W)$ enthalten, denn weil ϕ in jeder ihrer Komponenten linear ist, gilt dasselbe auch für die Komponenten von ϕ_{v_1}. Nach Induktionsvoraussetzung gibt es ein Element $\hat{\phi}_{v_1} \in \hat{\mathscr{L}}^n(V,W)$ mit $\hat{\phi}_{v_1}(v_2)\dots(v_{n+1}) = \phi_{v_1}(v_2,\dots,v_{n+1})$ für alle $v_2,\dots,v_{n+1}$. Definieren wir eine

Abbildung $\hat{\phi} : V \to \mathscr{L}^n(V,W)$ durch $\hat{\phi}(v_1) = \hat{\phi}_{v_1}$, so ist dies ein Element von $\hat{\mathscr{L}}^{n+1}(V,W) = \mathscr{L}(V, \hat{\mathscr{L}}^n(V,W))$, denn auf Grund der Linearität von ϕ in der ersten Komponente gilt für alle $v_1, v_1' \in V$, $v_2, ..., v_{n+1} \in V$ und $\lambda \in \mathbb{R}$ jeweils

$$\begin{aligned}
\hat{\phi}(v_1 + v_1')(v_2)...(v_{n+1}) &= \hat{\phi}_{v_1+v_1'}(v_2)...(v_{n+1}) = \phi_{v_1+v_1'}(v_2, ..., v_{n+1}) = \phi(v_1 + v_1', v_2, ..., v_{n+1}) \\
&= \phi(v_1, v_2, ..., v_{n+1}) + \phi(v_1', v_2, ..., v_{n+1}) = \phi_{v_1}(v_2, ..., v_{n+1}) + \phi_{v_1'}(v_2, ..., v_{n+1}) \\
&= \hat{\phi}_{v_1}(v_2)...(v_{n+1}) + \hat{\phi}_{v_1'}(v_2)...(v_{n+1}) = \hat{\phi}(v_1)(v_2)...(v_{n+1}) + \hat{\phi}(v_1')(v_2)...(v_{n+1}) \\
&= (\hat{\phi}(v_1) + \hat{\phi}(v_1'))(v_2)...(v_{n+1})
\end{aligned}$$

und

$$\begin{aligned}
\hat{\phi}(\lambda v_1)(v_2)...(v_{n+1}) &= \hat{\phi}_{\lambda v_1}(v_2)...(v_{n+1}) = \phi_{v_1}(v_2, ..., v_{n+1}) = \phi(\lambda v_1, v_2, ..., v_{n+1}) \\
= \lambda\phi(v_1, v_2, ..., v_{n+1}) &= \lambda\phi_{v_1}(v_2, ..., v_{n+1}) = \lambda\hat{\phi}_{v_1}(v_2)...(v_{n+1}) = \lambda\hat{\phi}(v_1)(v_2)...(v_{n+1}) \\
&= (\lambda\hat{\phi})(v_1)(v_2)...(v_{n+1})
\end{aligned}$$

also $\hat{\phi}(v_1 + v_1') = \hat{\phi}(v_1) + \hat{\phi}(v_1')$ und $\hat{\phi}(\lambda v_1) = \lambda\hat{\phi}(v_1)$. Insgesamt existiert also für jedes $\phi \in \mathscr{L}^{n+1}(V,W)$ ein $\hat{\phi} \in \hat{\mathscr{L}}^{n+1}(V,W)$ mit $\hat{\phi}(v_1)(v_2)...(v_{n+1}) = \hat{\phi}_{v_1}(v_2)...(v_{n+1}) = \phi_{v_1}(v_2, ..., v_{n+1}) = \phi(v_1, v_2, ..., v_{n+1})$ für alle $v_1, v_2, ..., v_{n+1} \in V$. Damit ist der Induktionsschritt abgeschlossen. □

Der Beweis lässt sich bedeutend übersichtlicher führen, wenn das Konzept der ***Tensorprodukte*** zur Verfügung steht. Aus Zeitgründen sind wir hier nicht der Lage, die genaue Definition des Tensorprodukts $V \otimes W$ zweier $\mathbb{R}$-Vektorräume V, W anzugeben. Wir erwähnen lediglich, dass es sich dabei um einen $\mathbb{R}$-Vektorraum handelt, dessen Elemente, die auch als ***Tensoren*** bezeichnet werden, in der Form $v_1 \otimes w_1 + ... + v_r \otimes w_r$ geschrieben werden können. Dabei gelten für $\lambda \in \mathbb{R}$, $v, v' \in V$ und $w, w' \in W$ jeweils die Rechenregeln $(v + v') \otimes w = v \otimes w + v' \otimes w$, $v \otimes (w + w') = v \otimes w + v \otimes w'$ und $(\lambda v) \otimes w = v \otimes (\lambda w) = \lambda(v \otimes w)$.

Sei V^* der $\mathbb{R}$-Vektorraum der linearen Abbildungen $V \to \mathbb{R}$, auch ***Dualraum*** von V genannt. Dann kann für $\varphi_1, ..., \varphi_n \in V^*$ und $w \in W$ das Element $\varphi_1 \otimes ... \otimes \varphi_n \otimes w$ über die Definition

$$(\varphi_1 \otimes ... \otimes \varphi_n \otimes w)(v_1, ..., v_n) = \varphi_1(v_1) \cdots \varphi_n(v_n) \cdot w$$

als multilineare Abbildung $V^n \to W$ betrachtet werden; die n-fache Linearität ergibt sich direkt aus der Linearität von $\varphi_1, ..., \varphi_n$. Sind V und W endlich-dimensional, so erhält man auf diese Weise einen Isomorphismus

$$\underbrace{V^* \otimes ... \otimes V^*}_{n\text{-mal}} \otimes W \cong \mathscr{L}^n(V,W)$$

von $\mathbb{R}$-Vektorräumen. Auf Grund der rekursiven Definition zu Beginn dieses Kapitels gilt andererseits $\hat{\mathscr{L}}^1(V,W) = \mathscr{L}(V,W) \cong V^* \otimes W$, $\hat{\mathscr{L}}^2(V,W) = \mathscr{L}(\hat{\mathscr{L}}(V,W)) \cong V^* \otimes \hat{\mathscr{L}}(V,W) \cong V^* \otimes V^* \otimes W$ und allgemein für jedes $n \in \mathbb{N}$ jeweils $\hat{\mathscr{L}}^n(V,W) \cong V^* \otimes ... \otimes V^* \otimes W$, auch hier mit n Faktoren V^*. Dies zeigt, dass die $\mathbb{R}$-Vektorräume $\mathscr{L}^n(V,W)$ und $\hat{\mathscr{L}}^n(V,W)$ auf natürliche Weise isomorph zueinander sind.

§ 10. Lokale Umkehrbarkeit und implizit definierte Funktionen

Zusammenfassung. Seien V, W normierte $\mathbb{R}$-Vektorräume derselben endlichen Dimension und $U \subseteq V$ eine offene Teilmenge. Eine Abbildung $f : U \to W$ wird als ***lokal umkehrbar*** an einer Stelle a bezeichnet, wenn f eine offene Umgebung $\tilde{U} \subseteq U$ von a bijektiv auf eine offene Umgebung $\tilde{W} \subseteq W$ von $b = f(a)$ abbildet. Wir werden sehen, dass dies für eine stetig differenzierbare Abbildung der Fall ist, wenn es sich bei der Ableitung $f'(a)$ um eine *invertierbare* lineare Abbildung $V \to W$ handelt. Die Existenz einer lokalen Umkehrfunktion $g : \tilde{W} \to \tilde{U}$ lässt sich dadurch interpretieren, dass die Gleichung $f(x) = y$ für jedes $x \in \tilde{U}$ durch $x = g(y)$ nach x aufgelöst werden kann. Daran erkennt man die Wichtigkeit des Kriteriums, denn die Lösbarkeit von Gleichungen kann als eines der Grundprobleme der Mathematik angesehen werden.

Man sagt, eine Funktion $f : D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}$ wird ***implizit*** durch eine Gleichung $g(x, y) = 0$ definiert, wenn $g(x, f(x)) = 0$ für alle $x \in D$ erüllt ist. Mit Hilfe der lokalen Umkehrbarkeit werden wir ein Kriterium herleiten, dass angibt, wann eine Gleichung gegeben durch eine Funktion $g : \mathbb{R}^2 \to \mathbb{R}$ in der Umgebung eines Punktes (a, b) implizit eine Funktion definiert. Natürlich lassen sich nicht nur ein-, sondern auch höherdimensionale Funktionen implizit definieren. Häufig sind Funktionen in physikalischen Anwendungen implizit definiert, z.B. wenn die Bahnkurven von Objekten durch Erhaltungsgrößen wie Energie, Impuls oder Drehimpuls festgelegt sind.

Wichtige Begriffe und Sätze in diesem Kapitel

- $\mathscr{C}^1$-Abbildung und $\mathscr{C}^1$-Diffeomorphismus
- Schrankensatz
- Satz über die lokale Umkehrbarkeit
- Satz über implizit definierte Funktionen

Bekanntlich ist die Funktion $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$ nicht injektiv und besitzt demzufolge auch keine Umkehrfunktion. Schränkt man f aber auf die Teilmenge $\mathbb{R}^+$ der positiven Zahlen oder auf die negativen Zahlen ein, so erhält man eine Bijektion auf das jeweilige Bild, deren Umkehrfunktion durch $y \mapsto \sqrt{y}$ bzw. $y \mapsto -\sqrt{y}$ gegeben ist. Es fällt auf, dass eine Umkehrfunktion immer dann existiert, wenn die einzige Nullstelle der Ableitung $f'(x) = 2x$, die Null, nicht im Definitionsbereich enthalten ist. Eines unserer Ziele in diesem Abschnitt besteht darin, diese Beobachtung auf mehrdimensionale Funktionen zu verallgemeinern.

Wie in den vorherigen Kapiteln bezeichnen V und W stets endlich-dimensionale normierte $\mathbb{R}$-Vektorräume.

(10.1) Definition. Sei $U \subseteq V$ eine offene Teilmenge. Eine stetig (total) differenzierbare Abbildung $f : U \to W$ wird auch ***$\mathscr{C}^1$-Abbildung*** genannt. Ist f eine Bijektion auf eine offene Teilmenge $\tilde{W} \subseteq W$, und ist die Umkehrabbildung $f^{-1} : \tilde{W} \to U$ ebenfalls ein $\mathscr{C}^1$-Abbildung, dann spricht man von einem ***$\mathscr{C}^1$-Diffeomorphismus*** zwischen U und $\tilde{W}$.

Der folgende Satz wird für die weiteren Beweise eine wichtige Rolle spielen.

(10.2) Satz. *(Schrankensatz)*
Seien $m, n \in \mathbb{N}$. Ist $U \subseteq \mathbb{R}^n$ offen, $f : \mathbb{R}^n \to \mathbb{R}^m$ eine $\mathscr{C}^1$-Abbildung und die Konstante $\gamma_U \in \mathbb{R}^+$ gegeben durch $\gamma_U = \sup\{\|f'(x)\| \mid x \in U\}$, dann gilt $\|f(x_1) - f(x_2)\| \leq \gamma_U \|x_1 - x_2\|$ für alle $x_1, x_2 \in U$ mit der Eigenschaft, dass die Verbindungsstrecke $[x_1, x_2]$ in U enthalten ist. Dabei wird auf $\mathbb{R}^n$ und $\mathbb{R}^m$ die Supremumsnorm zu Grunde gelegt, und $\|f'(x)\|$ bezeichnet die entsprechende Operatornorm von $f'(x) \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$.

Beweis: Seien $f_1, ..., f_m$ die einzelnen Komponenten $f_i : U \to \mathbb{R}$ von f. Für jedes $a \in U$ und jedes $v \in \mathbb{R}^n$ gilt jeweils $f'(a) = (f_1'(a), ..., f_m'(a))$ und somit $|f_i'(a)(v)| \leq \max\{|f_j'(a)(v)| \mid 1 \leq j \leq m\} = \|f'(a)(v)\| \leq \gamma_U \|v\|$, nach Definition der Operatornorm. Seien nun $x_1, x_2 \in U$ mit $[x_1, x_2] \subseteq U$ vorgegeben, und sei $v = x_2 - x_1$. Nach dem Mittelwertsatz (7.5) für Richtungsableitungen gibt es für $1 \leq j \leq m$ jeweils ein $a_j \in [x_1, x_2]$ mit $f_j'(a_j)(v) = \partial_v f_j(a_j) = f_j(x_2) - f_j(x_1)$, wobei $v = x_2 - x_1$ ist. Es folgt $|f_j(x_1) - f_j(x_2)| \leq |f_j'(a_i)| \leq \gamma_U \|v\|$ und somit $\|f(x_1) - f(x_2)\| \leq \gamma_U \|v\|$. □

(10.3) Satz. *(Satz über die lokale Umkehrbarkeit)*
Sei $f : U \to W$ eine $\mathscr{C}^1$-Abbildung. Ist $a \in U$ ein Punkt mit der Eigenschaft, dass $f'(a) \in \mathscr{L}(V, W)$ bijektiv ist, dann gibt es eine offene Umgebung $\tilde{U} \subseteq U$ von a und eine offene Umgebung $\tilde{W} \subseteq W$ von $b = f(a)$ mit der Eigenschaft, dass durch $f|_{\tilde{U}}$ $\mathscr{C}^1$-Diffeomorphismus zwischen $\tilde{U}$ und $\tilde{W}$ definiert ist.

Beweis: Ein Punkt a mit bijektiver Ableitung $f'(a)$ kann nur existieren, wenn die Dimensionen von V und W übereinstimmen, wovon wir von nun an ausgehen; sei $n = \dim V = \dim W$. Es gibt dann Isomorphismen $\kappa : V \to \mathbb{R}^n$ und $\kappa' : W \to \mathbb{R}^n$ von $\mathbb{R}$-Vektorräumen, die wegen (3.20) auch Homöomorphismen sind, unter denen also insbesondere offene Teilmengen erhalten bleiben. In den Übungen wurde gezeigt, dass lineare Abbildungen mit ihrer eigenen totalen Ableitung in einem beliebigen Punkt des Definitionsbereichs übereinstimmen. Definieren wir die Abbildung $\tilde{f} = \kappa' \circ f \circ \kappa^{-1}$ von $\kappa(U) \subseteq \mathbb{R}^n$ nach $\mathbb{R}^n$, dann ist auf Grund der Kettenregel auch $\tilde{f}$ total differenzierbar, und für jedes $a \in U$ gilt

$$\begin{aligned}\tilde{f}'(\kappa(a)) &= (\kappa \circ f \circ \kappa^{-1})'(\kappa(a)) = \kappa'((f \circ \kappa^{-1})(\kappa(a))) \circ (f \circ \kappa^{-1})'(\kappa(a)) = \\ &\kappa'(f(a)) \circ f'(\kappa^{-1}(\kappa(a))) \circ (\kappa^{-1})'(\kappa(a)) = \kappa \circ f'(a) \circ \kappa^{-1}.\end{aligned}$$

Mit $a \mapsto f'(a)$ ist auch die Abbildung $\kappa(a) \mapsto \kappa \circ f'(a) \circ \kappa^{-1}$ stetig, also ist auch $\tilde{f}$ eine $\mathscr{C}^1$-Abbildung. Ist die Aussage für $\tilde{f}$ bewiesen, dann gilt sie auf Grund des bisher Gesagten offenbar auch für f. Deshalb können wir uns von nun an auf den Fall $V = W = \mathbb{R}^n$ beschränken. Sei nun $a \in U$ ein Punkt mit der Eigenschaft, dass die Jacobi-Matrix $f'(a)$ invertierbar ist. Indem wir f durch $x \mapsto f(x+a) - f(a)$ ersetzen, können wir annehmen, dass $a = f(a) = 0_{\mathbb{R}^n}$ gilt. Weil außerdem f durch $f'(0_{\mathbb{R}^n})^{-1} \circ f$ ersetzt werden kann, haben wir die Aussage auf den Fall $f'(0_{\mathbb{R}^n}) = \mathrm{id}_{\mathbb{R}^n}$ reduziert.

Um die lokale Umkehrbarkeit zu beweisen, müssen wir für jedes w in der Nähe von $0_{\mathbb{R}^n}$ die Auflösbarkeit der Gleichung $w = f(v)$ nach v nachweisen. Offenbar ist die Gleichung äquivalent zu $w + v - f(v) = v$. Definieren wir die Hilfsfunktion $\varphi_w(x) = w + x - f(x)$, dann ist $w = f(v)$ also äquivalent zur Fixpunktgleichung $\varphi_w(v) = v$. Die entscheidende Idee besteht nun darin, den Banachschen Fixpunktsatz (2.11) auf diese Situation anzuwenden. Sei dazu $r \in \mathbb{R}^+$ so gewählt, dass der abgeschlossene Ball $\bar{B}_{2r}(0_{\mathbb{R}^n})$ in U enthalten ist und außerdem $\|\mathrm{id}_{\mathbb{R}^n} - f'(x)\| \leq \frac{1}{2}$ für alle $x \in \bar{B}_{2r}(0_{\mathbb{R}^n})$ gilt; dies ist wegen $f'(0_{\mathbb{R}^n}) = \mathrm{id}_{\mathbb{R}^n}$ und auf Grund der Stetigkeit der Ableitungsfunktion f' möglich. Für alle x in diesem abgeschlossenen Ball gilt dann $\varphi_w'(x) = \mathrm{id}_{\mathbb{R}^n} - f'(x)$, und mit dem Schrankensatz (10.2) erhalten wir $\|\varphi_w(x_1) - \varphi_w(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|$ für alle $x_1, x_2 \in \bar{B}_{2r}(0_{\mathbb{R}^n})$. Sind nun $w, x \in \mathbb{R}^n$ mit $\|w\| < r$ und $\|x\| \leq 2r$ vorgegeben, dann gilt die Abschätzung

$$\|\varphi_w(x)\| = \|\varphi_w(x) - \varphi_w(0_{\mathbb{R}^n}) + \varphi_w(0_{\mathbb{R}^n})\| \leq \|\varphi_w(x) - \varphi_w(0_{\mathbb{R}^n})\| + \|\varphi_w(0_{\mathbb{R}^n})\| \leq \tfrac{1}{2}\|x\| + \|w\| < 2r.$$

Dies alles zeigt, dass $\bar{B}_{2r}(0_{\mathbb{R}^n})$ durch φ_w in sich selbst abgebildet wird und eine Kontraktion mit der Konstanten $\gamma = \frac{1}{2}$ ist. Als abgeschlossene Teilmenge von $\mathbb{R}^n$ ist $\bar{B}_{2r}(0_{\mathbb{R}^n})$ ein vollständiger metrischer Raum, denn jede Cauchyfolge in $\bar{B}_{2r}(0_{\mathbb{R}^n})$ ist (auf Grund der Vollständigkeit von $\mathbb{R}^n$) eine in $\mathbb{R}^n$ konvergente Folge, deren Grenzwert nach Satz (4.16) wiederum in $\bar{B}_{2r}(0_{\mathbb{R}^n})$ enthalten ist. Insgesamt sind damit alle Voraussetzungen des Banachschen Fixpunktsatzes erfüllt. Für jedes $w \in B_r(0_{\mathbb{R}^n})$ gibt es demzufolge genau ein $v \in \bar{B}_{2r}(0_{\mathbb{R}^n})$ mit $\varphi_v(w) = w$ was, wie oben bemerkt, zu $f(v) = w$ äquivalent ist. Ferner zeigt die Ungleichung von oben, dass kein Fixpunkt von φ_w auf dem Rand der Kreisscheibe $\bar{B}_{2r}(0_{\mathbb{R}^n})$ liegen kann; der Fixpunkt v ist also sogar im offenen Ball $B_{2r}(0_{\mathbb{R}^n})$ enthalten. Jedes $w \in B_r(0_{\mathbb{R}^n})$ hat also genau ein Urbild in $B_{2r}(0_{\mathbb{R}^n})$. Definieren wir also $\tilde{W} = B_r(0_{\mathbb{R}^n})$ und $\tilde{U} = f^{-1}(\tilde{W}) \cap B_{2r}(0_{\mathbb{R}^n})$, dann ist $f|_{\tilde{U}}$ eine Bijektion zwischen $\tilde{U}$ und $\tilde{W}$, deren Umkehrabbildung $g : \tilde{W} \to \tilde{U}$ jedem $w \in \tilde{W}$ den eindeutig bestimmten Fixpunkt von φ_w in $B_{2r}(0_{\mathbb{R}^n})$ zuordnet.

Zum Schluss beweisen wir noch die Stetigkeit von g. Seien dazu $w_1, w_2 \in \tilde{W}$ vorgegeben und $x_1 = g(w_1)$, $x_2 = g(w_2)$. Mit Hilfe der Gleichung $\varphi_{0_{\mathbb{R}^n}}(x) = x - f(x)$ erhalten wir $x_2 - x_1 = \varphi_{0_{\mathbb{R}^n}}(x_2) - \varphi_{0_{\mathbb{R}^n}}(x_1) + f(x_2) - f(x_1)$ und somit

$$\begin{aligned} \|x_2 - x_1\| &= \|\varphi_{0_{\mathbb{R}^n}}(x_2) - \varphi_{0_{\mathbb{R}^n}}(x_1) + f(x_2) - f(x_1)\| \leq \|\varphi_{0_{\mathbb{R}^n}}(x_2) - \varphi_{0_{\mathbb{R}^n}}(x_1)\| + \|f(x_2) - f(x_1)\| \\ &\leq \tfrac{1}{2}\|x_2 - x_1\| + \|f(x_2) - f(x_1)\| \end{aligned}$$

was zu $\|g(w_2) - g(w_1)\| \leq \frac{1}{2}\|g(w_2) - g(x_1)\| + \|w_2 - w_1\|$ und somit zu $\|g(w_2) - g(w_1)\| \leq 2\|w_2 - w_1\|$ äquivalent ist. Damit ist die Stetigkeit von g auf $\tilde{W}$ bewiesen. Die Umkehrregel (8.15) zeigt nun, dass g auch differenzierbar ist, mit der Ableitungsfunktion gegeben durch $g'(w) = f'((f|_{\tilde{U}})^{-1}(w))^{-1}$. Weil die Invertierung von Matrizen stetig ist (wie wir schon im Beweis der Umkehrregel bemerkt haben), handelt es sich bei g sogar um eine $\mathscr{C}^1$-Abbildung. Insgesamt ist damit bewiesen, dass durch $f|_{\tilde{U}}$ ein $\mathscr{C}^1$-Diffeomorphismus zwischen $\tilde{U}$ und $\tilde{W}$ gegeben ist. □

Wir notieren uns noch eine wichtige Konsequenz aus dem Satz über die lokale Umkehrbarkeit.

(10.4) Folgerung. Sei $U \subseteq \mathbb{R}^n$ offen und $f : U \to \mathbb{R}^n$ eine injektive, stetig differenzierbare Abbildung mit der Eigenschaft, dass $f'(x)$ für jedes $x \in U$ invertierbar ist. Dann ist $V = f(U)$ eine offene Teilmenge von $\mathbb{R}^n$, und f ist ein Diffeomorphismus zwischen U und V.

Beweis: Zunächst überprüfen wir, dass V offen ist. Sei $b \in V$ vorgegeben und $a \in U$ mit $f(a) = b$. Auf Grund des Satzes über die lokale Umkehrbarkeit gibt es offene Umgebungen $\tilde{U} \subseteq U$ von a und $\tilde{V} \subseteq \mathbb{R}^n$ von b, so dass $f|_{\tilde{U}}$ ein Diffeomorphismus zwischen $\tilde{U}$ und $\tilde{V}$ ist. Insbesondere liegt $\tilde{V}$ in $V = f(U)$, was die Offenheit von V beweist.

Außerdem ist $f^{-1}|_{\tilde{V}} = (f|_{\tilde{U}})^{-1}$ und auf Grund der Diffeomorphismus-Eigenschaft von $f|_{\tilde{U}}$ ist f^{-1} in der Umgebung $\tilde{V}$ von b stetig differenzierbar. Nach Definition ist f als Abbildung $U \to V$ bijektiv, und wir haben soeben gezeigt, dass die Umkehrabbildung $f^{-1} : V \to U$ in jedem Punkt ihres Definitionsbereichs stetig differenzierbar ist. □

Diese Folgerung zeigt, dass insbesondere die Polarkoordinaten-Abbildung ρ_{pol} aus Satz (3.14) zu einem Diffeomorphismus auf ihre Bildmenge wird, sobald man sie auf $\mathbb{R}^+ \times I$ einschränkt, wobei $I \subseteq \mathbb{R}$ ein offenes Intervall mit Länge 2π bezeichnet. Dasselbe gilt auch für die Zylinderkoordinaten-Abbildung ρ_{zyl} eingeschränkt auf einen Bereich der Form $M_I = \mathbb{R}^+ \times I \times R$ und für die Kugelkoordinaten-Abbildung ρ_{kug} eingeschränkt auf $N_I = \mathbb{R}^+ \times]0, \pi[\times I$, mit einem ebensolchen Intervall I; diese Abbildungen wurden in Satz (3.15) definiert. Beispielsweise ist die Jacobi-Matix

von ρ_{kug} in jeden Punkt des Definitionsbereichs gegeben durch die Matrix

$$\rho'_{\text{kug}}(r,\vartheta,\varphi) \quad = \quad \begin{pmatrix} \sin(\vartheta)\cos(\varphi) & r\cos(\vartheta)\cos(\varphi) & -r\sin(\vartheta)\sin(\varphi) \\ \sin(\vartheta)\sin(\varphi) & r\cos(\vartheta)\sin(\varphi) & r\sin(\vartheta)\cos(\varphi) \\ \cos(\vartheta) & -r\sin(\vartheta) & 0 \end{pmatrix}$$

mit der Determinante $\det \rho'_{\text{kug}}(r,\vartheta,\varphi) = r^2\sin(\vartheta)$. Weil $\sin(\vartheta) > 0$ für $\vartheta \in]0,\pi[$ gilt, ist die Ableitung in jedem Punkt des Definitionsbereichs invertierbar.

(10.5) Definition. Seien $f : \mathbb{R}^2 \to \mathbb{R}$ eine Abbildung und $I', I'' \subseteq \mathbb{R}$ offene Intervalle. Wir sagen, eine Funktion $g : I' \to I''$ werde durch f ***implizit definiert***, wenn für alle $(x,y) \in I' \times I''$ die Äquivalenz

$$f(x,y) = 0 \quad \Leftrightarrow \quad y = g(x) \quad \text{erfüllt ist.}$$

Sei zum Beispiel $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch $f(x,y) = y^2 - x$. Dann definiert f implizit die Funktionen $g : \mathbb{R}^+ \to \mathbb{R}$, $x \mapsto \sqrt{x}$ und $h : \mathbb{R}^+ \to \mathbb{R}$, $x \mapsto -\sqrt{x}$. Dagegen gibt es keine durch f implizit definierte Funktion $k : I' \to I''$, die 0 oder eine negative Zahl in ihrem Definitionsbereich I' enthält.

Beweis: Zum Nachweis der impliziten Definition von g setzen wir $I' = I'' = \mathbb{R}^+$. Dann gilt für alle $(x,y) \in I' \times I''$ die Äquivalenz

$$f(x,y) = 0 \quad \Leftrightarrow \quad y^2 - x = 0 \quad \Leftrightarrow \quad y = \sqrt{x} \quad \Leftrightarrow \quad y = g(x).$$

Zum entsprechenden Nachweis für h setzen wir $I' = \mathbb{R}^+$ und $I'' = \mathbb{R}^- = \{y \in \mathbb{R} \mid y < 0\}$. Für alle $(x,y) \in I' \times I''$ gilt dann $y < 0$ und somit

$$f(x,y) = 0 \quad \Leftrightarrow \quad y^2 - x = 0 \quad \Leftrightarrow \quad y = -\sqrt{x} \quad \Leftrightarrow \quad y = h(x).$$

Nehmen wir an, es gibt eine durch f implizit definierte Funktion $k : I' \to I''$ mit $0 \in I'$. Dann gilt $f(0,k(0)) = 0$, also $k(0)^2 - 0 = 0$ und somit $k(0) = 0 \in I''$. Für hinreichend kleines $\varepsilon \in \mathbb{R}^+$ gilt $\varepsilon \in I'$ und $\pm\sqrt{\varepsilon} \in I''$. Aus $(\varepsilon, \sqrt{\varepsilon}) \in I' \times I''$ und $f(\varepsilon,\sqrt{\varepsilon}) = 0$ folgt nach Definition der implizit definierten Funktion $k(\varepsilon) = \sqrt{\varepsilon}$. Aus $(\varepsilon, -\sqrt{\varepsilon}) \in I' \times I'')$ und $f(\varepsilon,-\sqrt{\varepsilon}) = 0$ folgt ebenso $k(\varepsilon) = -\sqrt{\varepsilon}$. Der Widerspruch $\sqrt{\varepsilon} = k(\varepsilon) = -\sqrt{\varepsilon}$ zeigt nun, dass die Annahme falsch war und keine implizit definierte Funktion existiert.

Betrachten wir noch den Fall, dass der Definitionsbereich I' einer durch f implizit definierten Funktion $k : I' \to I''$ eine negative reelle Zahl a enthält. Dann müsste $k(a)^2 - a = f(a,k(a)) = 0$ gelten. Aber dies ist wegen $k(a)^2 - a \geq -a > 0$ unmöglich. □

Wie wir gleich sehen werden, hängt die Existenz von implizit definierten Funktionen mit den partiellen Ableitungen zusammen. Für $f(x,y) = y^2 - x$ gilt $\partial_2 f(x,y) = 2y$. Für $a > 0$ und $y = \pm\sqrt{a}$ gilt $f(a,y) = 0$ und $\partial_2 f(a,y) = 2y \neq 0$. Im Fall $a = 0$ ist $y = 0$ der einzige Wert mit $f(a,y) = 0$, und es gilt $\partial_2 f(0,0) = 0$.

Als weiteres Beispiel betrachten wir $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch $f(x,y) = x^2(1-x^2)-y^2$. Die Nullstellenmenge von f ist eine sog. ***Lemniskate***.

Sei nun $a \in]-1,1[\setminus \{0\}$. Dann definiert f in einer Umgebung von a implizit die Funktionen $g_\pm(x) = \pm\sqrt{x^2(1-x^2)}$. Dagegen gibt es in einer Umgebung von $a \in \{-1,0,1\}$ keine implizit definierten Funktionen.

Beweis: Sei zunächst $a \in]-1,1[$ mit $a \neq 0$. Dann gilt $a^2(1-a^2) > 0$. Im Fall $-1 < a < 0$ setzen wir $I' =]-1,0[$ und $I'' = \mathbb{R}^+$. Für alle $x \in I'$ gilt $x^2(1-x^2) > 0$, und für alle $(x,y) \in I' \times I''$ gilt folglich

$$f(x,y) = 0 \quad \Leftrightarrow \quad y^2 = x^2(1-x^2) \quad \Leftrightarrow \quad y = \sqrt{x^2(1-x^2)} \quad \Leftrightarrow \quad y = g_+(x).$$

Setzen wir $I'' = \mathbb{R}^-$, dann gilt für alle $(x,y) \in I' \times I''$ jeweils

$$f(x,y) = 0 \quad \Leftrightarrow \quad y^2 = x^2(1-x^2) \quad \Leftrightarrow \quad y = -\sqrt{x^2(1-x^2)} \quad \Leftrightarrow \quad y = g_-(x).$$

Den Fall $0 < a < 1$ behandelt man analog.

Sei nun $a = 1$, und nehmen wir an, dass $h : I' \to I''$ eine in der Nähe von a implizit definierte Funktion ist. Wegen $a^2(1-a^2) = 0$ ist $y = 0$ der einzige Wert mit $f(a,y) = 0$, und folglich gilt $(1,0) \in I' \times I''$. Sei $\varepsilon \in \mathbb{R}^+$ so klein gewählt, dass $x = 1+\varepsilon \in I'$ gilt. Wegen $x^2(1-x^2) < 0$ gibt es kein y mit $f(x,y) = x^2(1-x^2)-y^2 = 0$. Dies widerspricht der Annahme $f(x,h(x)) = 0$. Im Fall $a = -1$ läuft der Beweis analog.

Nehmen wir nun an, dass $h : I' \to I''$ eine in der Nähe von $a = 0$ implizit definierte Funktion ist. Wegen $a^2(1-a^2) = 0$ ist auch hier $y = 0$ der einzige Wert mit $f(a,y) = 0$, es gilt also $(0,0) \in I' \times I''$. Da $I' \times I''$ offen ist, gilt $\varepsilon \in I'$ und $\pm\sqrt{\varepsilon^2(1-\varepsilon^2)} \in I''$ für hinreichend kleines $\varepsilon \in \mathbb{R}^+$. Es ist $f(\varepsilon, \sqrt{\varepsilon^2(1-\varepsilon^2)}) = \varepsilon^2(1-\varepsilon^2)-\varepsilon^2(1-\varepsilon^2) = 0$ und ebenso $f(\varepsilon, -\sqrt{\varepsilon(1-\varepsilon^2)}) = 0$. Dies widerspricht der Annahme, dass für jedes $x \in I'$ nur ein $y \in I''$ mit $f(x,y) = 0$ existiert. □

Man kann bei der vorherigen Funktion auch x- und y-Wert vertauschen und die Abbildung $f(y,x) = x^2(1-x^2)-y^2$ betrachten. Durch Umformung der Gleichung $f(y,x) = 0$ nach x sieht man, dass f einer Umgebung jedes $b \in \mathbb{R}$ mit $-\frac{1}{2} < b < \frac{1}{2}$ implizit die Funktionen

$$g_\pm \quad = \quad \pm\tfrac{1}{2}\sqrt{2+2\sqrt{1-4y^2}}.$$

Für $b \neq 0$ kommen sogar noch die beiden Funktionen $h_\pm = \pm\frac{1}{2}\sqrt{2-2\sqrt{1-4y^2}}$ hinzu. Dagegen gibt es in einer Umgebung von $b = \pm\frac{1}{2}$ keine durch f implizit definierten Funktion.

(10.6) Definition. *(mehrdimensionale partielle Ableitungen)*

Seien X, Y, Z endlich-dimensionale $\mathbb{R}$-Vektorräume, $U \subseteq X \times Y$ eine offene Teilmenge und $f : U \to Z$ eine differenzierbare Abbildung. Ist $(x,y) \in U$, dann definieren wir die ***partielle Ableitung*** von f in X- bzw. Y-Richtung durch

$$\partial_X f(x,y) : X \to Z\,,\ v \mapsto f'(x,y)(v,0) \qquad \text{und} \qquad \partial_Y f(x,y) : Y \to Z\,,\ w \mapsto f'(x,y)(0,w).$$

Nach Definition gilt für alle $(x,y) \in U$, $v \in X$ und $w \in Y$ jeweils

$$\partial_X f(x,y)(v) + \partial_Y f(x,y)(w) \quad = \quad f'(x,y)(v,0) + f'(x,y)(0,w) \quad = \quad f'(x,y)(v,w).$$

Im Fall $X = \mathbb{R}^m$, $Y = \mathbb{R}^n$ können wir $X \times Y$ mit $\mathbb{R}^{m+n}$ identifizieren. Ist dann $U \subseteq \mathbb{R}^{m+n}$ offen und $f : U \to \mathbb{R}^k$ eine differenzierbare Abbildung, so besteht für jeden Punkt $(x,y) \in X \times Y$ die Matrix $\partial_X f(x,y)$ aus den vorderen m und $\partial_Y f(x,y)$ aus den hinteren n Spalten von $f'(x,y)$. Dies folgt ganz einfach aus der Tatsache, dass die Spalten einer Matrix $A \in \mathscr{M}_{m\times n,\mathbb{R}}$ die Bilder der Einheitsvektoren unter der Abbildung $\mathbb{R}^n \to \mathbb{R}^n$, $v \mapsto Av$ sind.

(10.7) Lemma. Seien X, Y, Z endlich-dimensionale $\mathbb{R}$-Vektorräume mit $\dim Y = \dim Z$, und seien $\phi : X \to Z$ und $\psi : Y \to Z$ lineare Abbildungen, wobei ψ invertierbar ist. Dann ist auch die lineare Abbildung $\Phi : X \times Y \to X \times Z$, $(u,v) \mapsto (u, \phi(u) + \psi(v))$ invertierbar, mit der Zuordnung $X \times Z \to X \times Y$, $(u,w) \mapsto (u, \psi^{-1}(w) - (\psi^{-1} \circ \phi)(u))$ als Umkehrabbildung.

Beweis: Seien $(u,v) \in X \times Y$ und $(u_1,w) \in X \times Y$. Dann gilt die Äquivalenz

$$\begin{gathered}(u_1,w) = \Phi(u,v) \quad \Leftrightarrow \quad (u_1,w) = (u, \phi(u)+\psi(v)) \quad \Leftrightarrow \quad (u_1 = u) \wedge (w = \phi(u)+\psi(v)) \quad \Leftrightarrow \\ (u_1 = u) \wedge (w = \phi(u_1)+\psi(v)) \quad \Leftrightarrow \quad (u = u_1) \wedge (\psi(v) = w - \phi(u_1)) \quad \Leftrightarrow \\ (u = u_1) \wedge (v = \psi^{-1}(w) - \psi^{-1}(\phi(u_1))) \quad \Leftrightarrow \quad (u,v) = (u_1, \psi^{-1}(w) - (\psi^{-1} \circ \phi)(u_1)).\end{gathered}$$

Dies zeigt, dass $\Psi(u_1,w) = (u_1, \psi^{-1}(w) - (\psi^{-1} \circ \phi)(u_1))$ tatsächlich die Umkehrabbildung von Φ ist. □

(10.8) Satz. *(Satz über implizite Funktionen)*

Sei $X = \mathbb{R}^m$, $Y = \mathbb{R}^n$ und $U \subseteq X \times Y$ offen. Sei außerdem $f : U \to Y$ eine stetig differenzierbare Abbildung und $(a,b) \in U$ eine Nullstelle von f mit der Eigenschaft, dass $\partial_Y f(a,b)$ invertierbar ist. Dann gibt es Umgebungen $U' \subseteq X$ von a und $U'' \subseteq Y$ von b mit $U' \times U'' \subseteq U$ und eine stetig differenzierbare Abbildung $g : U' \to U''$, so dass die Äquivalenz

$$f(x,y) = 0_Y \quad \Leftrightarrow \quad y = g(x) \qquad \text{für alle } (x,y) \in U' \times U'' \quad \text{erfüllt ist.}$$

Beweis: Die wesentliche Idee besteht darin, die Auflösung der Gleichung $f(x,y) = 0_Y$ nach y so umzuformulieren, dass sie auf die Bestimmung einer lokalen Umkehrabbildung hinausläuft. Die Existenz der Umkehrabbildung wird dann mit dem Satz (10.3) über die lokale Umkehrbarkeit bewiesen. Sei $\Phi : U \to X \times Y$ die $\mathscr{C}^1$-Abbildung definiert durch $\Phi(x,y) = (x, f(x,y))$, und nehmen wir an, $\tilde{U} \subseteq U$ ist eine offene Teilmenge derart, dass $\Phi|_{\tilde{U}}$ einen $\mathscr{C}^1$-Diffeomorphismus zwischen $\tilde{U}$ und $\tilde{V} = \Phi(\tilde{U})$ definiert. Dessen Umkehrabbildung $\tilde{V} \to \tilde{U}$ bezeichnen wir mit Ψ. Seien $(w,z) \in \tilde{V}$ mit $w \in X$ und $z \in Y$, und sei $(x,y) = \Psi(w,z)$. Dann folgt $(x, f(x,y)) = \Phi(x,y) = (w,z)$, also insbesondere $x = w$. Dies zeigt, dass eine Abbildung $h : \tilde{V} \to Y$ mit $\Psi(w,z) = (w, h(w,z))$ für alle $(w,z) \in \tilde{V}$ existiert. Mit Ψ ist auch h eine $\mathscr{C}^1$-Abbildung.

Nehmen wir nun weiter an, dass $\tilde{U}$ die Form $U' \times U''$ hat, wobei $U' \subseteq X$ und $U'' \subseteq Y$ offene Teilmengen bezeichnen. Dann gilt für alle $x \in U'$, $y \in U''$ und $z \in Y$ die Äquivalenz

$$\begin{gathered}f(x,y) = z \quad \Leftrightarrow \quad (x, f(x,y)) = (x,z) \quad \Leftrightarrow \quad \Phi(x,y) = (x,z) \quad \Leftrightarrow \quad (x,y) = \Psi(x,z) \\ \Leftrightarrow \quad (x,y) = (x, h(x,z)) \quad \Leftrightarrow \quad y = h(x,z).\end{gathered}$$

Die Teilmenge aller $x \in U'$ mit $(x, 0_Y) \in \tilde{V}$ ist offen auf Grund der Offenheit von $\tilde{V}$ und der Stetigkeit der Abbildung $x \mapsto (x, 0_Y)$. Wir ersetzen U' durch diese Teilmenge und definieren $g : U' \to U''$ durch $g(x) = h(x, 0_Y)$. Dann

gilt die Äquivalenz $f(x,y) = 0_Y \iff y = g(x)$ für alle $x \in U'$ und $y \in U''$. Mit h ist auch die Abbildung g stetig differenzierbar.

Es bleibt zu zeigen, dass die im bisherigen Teil formulierten Annahmen unter den gegebenen Voraussetzungen erfüllt werden können. Sei also $(a,b) \in U$ eine Nullstelle von f mit der Eigenschaft, dass die partielle Ableitung $\partial_Y f(a,b)$ invertierbar ist. Wir zeigen, dass die oben definierte Abbildung $\Phi : U \to X \times Y$ bei (a,b) lokal umkehrbar ist. Dazu zerlegen wir Φ in die Komponenten

$$\Phi_X : U \longrightarrow X, \quad (x,y) \mapsto x \qquad \text{und} \qquad \Phi_Y : U \longrightarrow Y, \quad (x,y) \mapsto f(x,y).$$

Da Φ_X eine lineare Abbildung ist, gilt $\Phi'_X(x,y) = \Phi_X$ für alle (x,y) in U. Für die Ableitung von Φ' an der Stelle (a,b) und $(u,v) \in X \times Y$ erhalten wir

$$\Phi'(a,b)(u,v) \quad = \quad (\Phi'_X(a,b)(u,v), \Phi'_Y(a,b)(u,v)) \quad = \quad (\Phi_X(u,v), f'(a,b)(u,v)) \quad =$$

$$(u, \partial_X f(a,b)(u) + \partial_Y f(a,b)(v)).$$

Weil $\partial_Y f(a,b)$ invertierbar ist, erhalten wir mit Lemma (10.7) die Invertierbarkeit von $\Phi'(a,b)$. Der Satz über die lokale Umkehrbarkeit ist damit anwendbar und liefert eine offene Umgebung $\tilde{U}$ von (a,b), so dass $\Phi|_{\tilde{U}}$ zu einem $\mathscr{C}^1$-Diffeomorphismus auf die offene Bildmenge $\tilde{V} = \Phi(\tilde{U})$ wird. Nach Verkleinerung von $\tilde{U}$ und $\tilde{V}$ können wir annehmen, dass $\tilde{U} = U' \times U''$ mit offenen Umgebungen U' von a und U'' von b gilt. Wegen $(a,b) \in \tilde{U}$ und $\Phi(a,b) = (a, f(a,b)) = (a, 0_Y)$ ist $(a,0_Y)$ in $\tilde{V}$ enthalten. Wie oben ersetzen wir U' durch die Teilmenge aller $x \in U'$ mit $(x,0_Y) \in \tilde{V}$. Dann ist auch U' eine offene Umgebung von a, und die oben definierte Funktion $g : U' \to U''$ hat alle gewünschten Eigenschaften. □

(10.9) Folgerung. Die Funktion $g : U' \to U''$ aus dem Satz hat an der Stelle a die Ableitung

$$g'(a) \quad = \quad -\partial_Y f(a,b)^{-1} \circ \partial_X f(a,b).$$

Beweis: Hierzu verwenden wir die Kettenregel. Sei $\Phi : U' \to Z$ gegeben durch $x \mapsto f(x, g(x))$. Nach Definition gilt $\Phi = f \circ g_1$ mit den Funktionen $f : U \to Z$ und $g_1 : U' \to U$, $x \mapsto (x, g(x))$. Die Ableitung von g_1 kann komponentenweise gebildet werden: Es ist $g'_1(x) = (\mathrm{id}_X, g'(x))$ für alle $x \in U'$. Die Kettenregeln liefert nun

$$\Phi'(x) \quad = \quad f'(g_1(x)) \circ g'_1(x) \quad = \quad f'(x, g(x)) \circ (\mathrm{id}_X, g'(x)).$$

Wir stellen nun Φ' mit Hilfe der mehrdimensionalen partiellen Ableitungen von f dar. Für einen beliebigen Vektor $v \in X$ gilt

$$\Phi'(x)(v) \quad = \quad f'(x, g(x))((\mathrm{id}_X, g'(x))(v)) \quad = \quad f'(x, g(x))(v, g'(x)(v)) \quad =$$

$$\partial_X f(x, g(x))(v) + \partial_Y f(x, g(x))(g'(x)(v)) \quad = \quad \partial_X f(x, g(x))(v) + (\partial_Y f(x, g(x)) \circ g'(x))(v)$$

also $\Phi'(x) = \partial_X f(x, g(x)) + \partial_Y f(x, g(x)) \circ g'(x)$. Nun ist die Funktion g gerade so definiert, dass $\Phi(x) = f(x, g(x)) = 0$ für alle $x \in U'$ gilt. Folglich ist $\Phi'(x) = 0$ für alle $x \in U'$. Wegen $b = g(a)$ und auf Grund der Invertierbarkeit von

$\partial_Y f$ im Punkt (a,b) erhalten wir

$$\begin{aligned}\partial_X f(a,b) + \partial_Y f(a,b) \circ g'(a) = 0 \quad &\Leftrightarrow \quad \partial_Y f(a,b) \circ g'(a) = -\partial_X f(a,b) \\ &\Leftrightarrow \quad g'(a) = -\partial_Y f(a,b)^{-1} \circ \partial_X f(a,b).\end{aligned}$$

□

Beispiel. Wir untersuchen die Auflösbarkeit des Gleichungssystems

$$\begin{aligned} x^3 + y^3 + z^3 &= 7 \\ xy + yz + xz &= -2 \end{aligned}$$

nach (y,z) in einer Umgebung des Punktes $(2,-1,0)$. Es sollen also gezeigt werden, dass offene Umgebungen $U' \subseteq \mathbb{R}$ von 2 und $U'' \subseteq \mathbb{R}^2$ von $(1,0)$ und stetig differenzierbare Funktionen $g,h : I \to \mathbb{R}$ existieren, so dass für alle $x \in I$ und $(y,z) \in U'$ die Äquivalenz

$$\begin{aligned} x^3 + y^3 + z^3 &= 7 \\ xy + yz + xz &= -2 \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} y &= g(x) \\ z &= h(x) \end{aligned} \qquad \text{erfülllt ist.}$$

Außerdem berechnen wir die Ableitungen $g'(2)$ und $h'(2)$.

Um den Satz über implizite Funktionen anwenden zu können, setzen wir $X = \mathbb{R}$, $Y = Z = \mathbb{R}^2$ und definieren die Abbildung $f : X \times Y \to \mathbb{R}^2$ durch

$$f(x,y,z) \quad = \quad (x^3 + y^3 + z^3 - 7, xy + yz + xz + 2).$$

Die Ableitung von f an einem beliebigen Punkt $(x,y,z) \in \mathbb{R}^3$ ist gegeben durch

$$f'(x,y,z) \quad = \quad \begin{pmatrix} 3x^2 & 3y^2 & 3z^2 \\ y+z & x+z & x+y \end{pmatrix}$$

also insbesondere

$$f'(2,-1,0) = \begin{pmatrix} 12 & 3 & 0 \\ -1 & 2 & 1 \end{pmatrix} \quad , \quad \partial_X f(2,-1,0) = \begin{pmatrix} 12 \\ -1 \end{pmatrix} \quad , \quad \partial_Y f(2,-1,0) = \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix}.$$

Wegen $\det \partial_Y f(2,-1,0) = 3 \neq 0$ ist $\partial_Y f(2,-1,0)$ invertierbar, der Satz über implizite Funktionen kann also angewendet werden. Es liefert uns offene Umgebungen U' von 2 und U'' von $(1,0)$ und eine Funktion $k : U' \to U''$, so dass $f(x,y,z) = 0 \Leftrightarrow (y,z) = k(x)$ für alle $x \in U'$ und $(y,z) \in U''$ erfüllt ist. Die Gleichung $f(x,y,z) = 0$ entspricht dem Gleichungsystem auf der linken Seite der gewünschten Äquivalenz. Sind die Funktion $g,h : U' \to \mathbb{R}$ gegeben durch $k = (g,h)$, dann die entspricht $(y,z) = k(x)$ dem umgeformten Gleichungssystem oben rechts. Die zu $\partial_Y f(2,-1,0)$ inverse Matrix ist gegeben durch

$$\partial_Y f(2,-1,0)^{-1} \quad = \quad \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix}^{-1} \quad = \quad \tfrac{1}{3} \begin{pmatrix} 1 & 0 \\ -2 & 3 \end{pmatrix} \quad ,$$

und die Ableitung $k'(2)$ erhalten wir nach Folgerung (10.9) durch

$$\begin{aligned} k'(2) \quad &= \quad -\partial_Y f(2,-1,0)^{-1} \circ \partial_X f(2,-1,0) \quad = \quad -\tfrac{1}{3} \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 12 \\ -1 \end{pmatrix} \\ &= \quad -\tfrac{1}{3} \begin{pmatrix} 1 & 0 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 12 \\ -1 \end{pmatrix} \quad = \quad \begin{pmatrix} -4 \\ 9 \end{pmatrix}. \end{aligned}$$

Es gilt also $g'(2) = -4$ und $h'(2) = 9$.

§ 11. Untermannigfaltigkeiten

Zusammenfassung. Bisher haben wir bei der Bestimmung von Extrema nur Funktionen betrachtet, die auf offenen Teilmengen des $\mathbb{R}^n$ definiert sind. Häufig, zum Beispiel in vielen physikalischen Anwendungen, interessiert man sich aber für Extrema von Funktionen, deren Definitionsbereich auf Teilmengen eines allgemeinere Typs eingeschränkt ist. Oft handelt es sich dabei um eine glatte Kurve oder Fläche, allgemeiner um eine sog. *Untermannigfaltigkeit des* $\mathbb{R}^n$. Grob gesprochen ist dies eine Teilmengen des $\mathbb{R}^n$, die als Nullstellenmenge eines Systems stetig differenzierbarer Funktionen dargestellt werden kann. Dieses System muss eine zusätzliche technische Bedingung erfüllen, die Knoten, Knicke oder sonstige Unregelmäßigkeiten in der geometrischen Struktur ausschließt.

Ist nun $U \subseteq \mathbb{R}^n$ nun eine offene Teilmenge, $f : U \to \mathbb{R}$ eine Funktion und $M \subseteq \mathbb{R}^n$ eine solche Untermannigfaltigkeit, dann bezeichnet man die lokalen Extrema von $f|_M$ als lokale Extrema von f unter einer ***Nebenbedingung*** (nämlich der, dass das Extremum auf M liegt). Das Ziel dieses Abschnitts besteht darin, ein einfach zu überprüfendes hinreichendes Kriterium für solche Extrema aufzustellen und somit das Kriterium aus § 10 zu verallgemeinern.

Wichtige Begriffe und Sätze in diesem Kapitel

- Untermannigfaltigkeiten des $\mathbb{R}^n$
- Definition von Extrema unter Nebenbedingungen
- notwendiges Kriterium für Extrema unter Nebenbedingungen

(11.1) Definition. Seien $d, n \in \mathbb{N}$ mit $d < n$. Eine Teilmenge $M \subseteq \mathbb{R}^n$ wird d-dimensionale ***Untermannigfaltigkeit*** des $\mathbb{R}^n$ genannt, wenn es für jeden Punkt $a \in M$ eine offene Umgebung U und $n-d$ stetig differenzierbare Funktionen $\varphi_1, ..., \varphi_{n-d} : U \to \mathbb{R}$ gibt, so dass gilt

(i) $U \cap M = \{x \in U \mid \varphi_1(x) = ... = \varphi_{n-d}(x) = 0\}$

(ii) $\dim \operatorname{lin}\{\varphi_1'(x), ..., \varphi_{n-d}'(x)\} = n-d$ für alle $x \in U \cap M$
(Mit anderen Worten, die Ableitungen von $\varphi_1, ..., \varphi_{n-d}$ im Punkt x sind als Elemente des $\mathbb{R}$-Vektorraums $\mathscr{L}(\mathbb{R}^n, \mathbb{R})$ linear unabhängig.)

Schauen wir uns eine Reihe von Beispielen für Untermannigfaltigkeiten an.

(i) Jeder d-dimensionale Untervektorraum $M \subseteq \mathbb{R}^n$ kann als Lösungsmenge eines Systems von $n-d$ linear unabhängigen Gleichungen

$$f_i(x) \quad = \quad a_{i1}x_1 + ... + a_{in}x_n \quad = \quad 0 \quad , \quad 1 \le i \le n-d$$

beschrieben werden. Weil die Funktionen f_i linear sind, stimmen sie in jedem Punkt mit ihrer eigenen Ableitung überein. Deshalb ist M eine d-dimensionale Untermannigfaltigkeit im $\mathbb{R}^n$.

(ii) Der Kreis $S^1 \subseteq \mathbb{R}^2$ mit Mittelpunkt 0 und Radius 1 wird beschrieben durch die Gleichung $f(x,y) = x^2+y^2-1 = 0$. Es gilt $f'(x,y) = (2x\ 2y)$ und somit $f'(x,y) \neq 0$ für alle $(x,y) \in S^1$. Also ist S^1 eine 1-dimensionale Untermannigfaltigkeit im $\mathbb{R}^2$. Genauso zeigt man, dass die ***Sphäre*** $S^2 = \{(x,y,z) \in \mathbb{R}^3 \mid x^2+y^2+z^2-1 = 0\}$ eine zweidimensionale Untermannigfaltigkeit im $\mathbb{R}^3$ und allgemeiner die sogenannte ***n-Sphäre***

$$S^n \quad = \quad \left\{ (x_0, ..., x_n) \in \mathbb{R}^{n+1} \;\middle|\; \sum_{i=0}^{n} x_i^2 - 1 = 0 \right\}$$

eine n-dimensionale Untermannigfaltigkeit im $\mathbb{R}^{n+1}$ ist.

(iii) Die ***Hyperbel*** gegeben durch $H = \{(x,y) \in \mathbb{R}^2 \mid xy - 1\}$ ist eine 1-dim. Untermannigfaltigkeit im $\mathbb{R}^2$.

(iv) Das ***Rotationsparaboloid*** $P = \{(x,y,z) \in \mathbb{R}^3 \mid z - x^2 - y^2 = 0\}$ ist eine 2-dim. Untermannigfaltigkeit im $\mathbb{R}^3$.

(v) Der ***Kegel*** $C = \{(x,y,z) \in \mathbb{R}^3 \mid z^2 - x^2 - y^2 = 0\}$ ist *keine* 2-dimensionale Untermannigfaltigkeit im $\mathbb{R}^3$, weil die Ableitung $(x,y) \mapsto (-2x\ -2y\ 2z)$ der definierenden Funktion im Punkt $(0,0,0)$ verschwindet. Entfernt man diesen Punkt, dann erhält man durch $\tilde{C} = C \setminus \{(0,0,0)\}$ eine 2-dimensionale Untermannigfaltigkeit.

(vi) Auch die sogenannte ***Neillsche Parabel*** $N = \{(x,y) \in \mathbb{R}^2 \mid y^2 - x^3 = 0\}$ und die Kurve gegeben durch

$$C \quad = \quad \{(x,y) \in \mathbb{R}^2 \mid y^2 - x^3 - x^2 = 0\}$$

werden erst dann zu 1-dimensionalen Untermannigfaltigkeiten im $\mathbb{R}^2$, wenn man den Punkt $(0,0)$ entfernt.

Wir formulieren nun ein notwendiges Kriterium für Extrema auf Mannigfaltigkeiten, dass wir allerdings aus Zeitgründen nicht beweisen. Für einen überschaubaren und transparenten Beweis des Kriteriums sind andere Zugänge zum Begriff der Untermannigkeit besser geeignet, die allerdings auch einen höheren technischen Aufwand erfordern.

(11.2) Satz. *(Satz über Extrema mit Nebenbedingungen)*

Sei M eine d-dimensionale Untermannigfaltigkeit im $\mathbb{R}^n$, $U \subseteq \mathbb{R}^n$ eine offene Teilmenge, und seien $\varphi_1, ..., \varphi_{n-d}$ stetig differenzierbare Funktionen $\varphi_i : U \to \mathbb{R}$ mit der Eigenschaft, dass die Bedingungen (i) und (ii) aus (11.1) erfüllt sind. Sei $f : U \to \mathbb{R}$ eine weitere stetig differenzierbare Funktion. Ist dann $a \in U \cap M$ ein lokales Extremum von f auf M, dann gibt es reelle Zahlen $\lambda_1, ..., \lambda_{n-d} \in \mathbb{R}$ (die sogenannten ***Lagrange-Multiplikatoren***), so dass in $\mathscr{L}(\mathbb{R}^n, \mathbb{R})$ die Gleichung

$$f'(a) \quad = \quad \sum_{j=1}^{n-d} \lambda_j \varphi_j'(a) \qquad \text{erfüllt ist.}$$

Wir betrachten ein Anwendungsbeispiel für den soeben formulierten Satz. Sei $U = \mathbb{R}^2$, die Funktion $f : U \to \mathbb{R}$ gegeben durch $f(x,y) = x + y$ und $K = S^1 = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$. Dann ist K die Nullstellenmenge der Funktion $\varphi(x,y) = x^2 + y^2 - 1$, und die Ableitungen von φ_1 und f sind in jedem Punkt (x,y) gegeben durch

$$\varphi_1'(x,y) = \begin{pmatrix} 2x & 2y \end{pmatrix} \qquad \text{und} \qquad f'(x,y) = \begin{pmatrix} 1 & 1 \end{pmatrix}.$$

Wenn f auf K im Punkt (x,y) ein lokales Extremum besitzt, dann muss nach (11.2) der Vektor $f'(x,y) = (1\ 1)$ ein skalares Vielfaches von $\varphi_1'(x,y) = (2x\ 2y)$ sein. Es muss also $x = y$ und $2x^2 = x^2 + x^2 = 1$ gelten. Dies ist nur für die

Punkte $\pm(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ der Fall. Durch eine direkte Rechnung kann man in der Tat leicht sehen, dass die eingeschränkte Funktion $f|_K$ im Punkt $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ ihr Minimum und im Punkt $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ ihr Maximum annimmt. Als Funktion auf $\mathbb{R}^2$ besitzt f natürlich keine lokalen Extremstellen, geschweige denn globale.

Als weiteres Beispiel bestimmen wir den Punkt p auf der Ebene $E = \{(x,y,z) \in \mathbb{R}^3 \mid z = x + y\}$, der vom Punkt $q = (1,0,0)$ den kleinsten Abstand hat. Dabei setzen wir als bekannt voraus, dass ein eindeutiger Punkt p mit dieser Eigenschaft existiert. Zu diesem Zweck bestimmen wir ein lokales Minimum der quadrierten Abstandsfunktion

$$f : \mathbb{R}^3 \to \mathbb{R}\ ,\ (x,y,z) \mapsto (x-1)^2 + y^2 + z^2 \qquad \text{auf} \qquad E = \{(x,y,z) \mid g(x,y,z) = 0\} \quad ,$$

wobei $g : \mathbb{R}^3 \to \mathbb{R}$ durch $(x,y,z) \mapsto x + y - z$ gegeben ist. Die Ableitungen von f und g sind in jedem Punkt gegeben durch

$$f'(x,y,z) = \begin{pmatrix} 2x-2 & 2y & 2z \end{pmatrix} \qquad \text{und} \qquad g'(x,y,z) = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}.$$

Ist $p = (x,y,z) \in E$ ein lokales Extremum von f, dann muss nach (11.2) die Gleichung $(2x-2\ \ 2y\ \ 2z) = \lambda(1\ \ 1\ \ -1)$ für ein $\lambda \in \mathbb{R}$ erfüllt sein. Dies liefert das Gleichungssystem $2x-2 = \lambda$, $2y = \lambda$, $2z = -\lambda$, was zu $x = \frac{1}{2}\lambda + 1$, $y = \frac{1}{2}\lambda$, $z = -\frac{1}{2}\lambda$ umgeformt werden kann. Wegen $p \in E$ gilt außerdem

$$g(x,y,z) = 0 \quad \Leftrightarrow \quad x + y = z \quad \Leftrightarrow \quad (\tfrac{1}{2}\lambda + 1) + \tfrac{1}{2}\lambda = -\tfrac{1}{2}\lambda \quad \Leftrightarrow \quad \lambda + 1 = -\tfrac{1}{2}\lambda$$
$$\Leftrightarrow \quad \tfrac{3}{2}\lambda = -1 \quad \Leftrightarrow \quad \lambda = -\tfrac{2}{3}.$$

Somit ist $p = (x,y,z) = (\frac{2}{3}, -\frac{1}{3}, \frac{1}{3})$ der gesuchte Punkt.

§ 12. Definition des mehrdimensionalen Riemann-Integrals

Zusammenfassung. In diesem Kapitel soll die aus dem ersten Semester bekannte Definition des Riemann-Integrals auf den $\mathbb{R}^n$ verallgemeinert werden. Im Eindimensionalen hatten wir die Riemann-Integrierbarkeit einer beschränkten Funktion $f : [a, b] \to \mathbb{R}$ folgendermaßen definiert: Zunächst hatten wir jeder Zerlegung $\mathscr{Z}$ des Intervalls $[a, b]$ eine Untersumme $\mathscr{S}_f^-(\mathscr{Z})$ und eine Obersumme $\mathscr{S}_f^+(\mathscr{Z})$ zugeordnet. Das *Unterintegral* war dann nach Definition das Supremum aller Untersummen, das *Oberintegral* das Infimum aller Obersummen. Die Funktion wurde als *Riemann-integrierbar* bezeichnet, wenn Unter- und Oberintegral übereinstimmten, und diesen gemeinsamen Wert wurde dann das *Riemann-Integral* von f genannt.

Im $\mathbb{R}^n$ betrachten wir als Definitionsbereich unserer Funktionen $f : Q \to \mathbb{R}$ anstelle eines Intervalls $[a, b]$ achsenparallele kompakte Quader der Form $Q = [a_1, b_1] \times ... \times [a_n, b_n]$. Jede Zerlegung der Intervalle $[a_j, b_j]$ liefert eine Zerlegung der Quaders in Teilquader. Mit Hilfe dieser Teilquader lassen sich wie im Eindimensionalen Unter- und Obersummen und damit auch Unter- und Oberintegrale definieren. Die Definitionen der Riemann-Integrierbarkeit und des Riemann-Integrals lassen sich dann analog zum eindimensionalen Fall formulieren. Neben diesen Definitionen ist das wichtigste Ergebnis dieses Kapitels die Riemann-Integrierbarkeit jeder stetigen Funktion $f : Q \to \mathbb{R}$ auf einem Bereich Q der angegebenen Form.

Wichtige Begriffe und Sätze in diesem Kapitel

- (achsenparalleler, kompakter) Quader
- Zerlegung mehrdimensionaler Quader, Ober- und Untersumme einer beschränkten Funktion
- Definition der Riemann-integrierbaren Funktionen und des Riemann-Integrals
- Integrierbarkeitskritierien
- Beispiel für eine nicht Riemann-integrierbare Funktion

Bevor wir die mehrdimensionalen Riemann-Integrale definieren, wiederholen wir zunächst die Definition im eindimensionalen Fall aus dem ersten Semester. Sei $[a, b] \subseteq \mathbb{R}$ ein endliches, abgeschlossenes Intervall und $f : [a, b] \to \mathbb{R}$ eine beschränkte Funktion. Eine ***Zerlegung*** von $[a, b]$ ist eine ist eine endliche Teilmenge $\mathscr{Z} = \{x_1, ..., x_{n-1}\}$ von $]a, b[$. Für jede solche Zerlegung hatten wir die ***Unter-*** und ***Obersumme*** von f definiert durch

$$\mathscr{S}_f^-(\mathscr{Z}) = \sum_{k=0}^{n-1} c_k(x_{k+1} - x_k) \qquad \text{und} \qquad \mathscr{S}_f^+(\mathscr{Z}) = \sum_{k=0}^{n-1} d_k(x_{k+1} - x_k)$$

wobei $c_k, d_k \in \mathbb{R}$ für $0 \leq k < n$ durch $c_k = \inf\{f(x) \mid x \in [x_k, x_{k+1}]\}$ und $d_k = \sup\{f(x) \mid x \in [x_k, x_{k+1}]\}$ gegeben sind und $x_0 = a$, $x_n = b$ gesetzt wird. Das ***Unterintegral*** ist dann das Supremum aller Untersummen und das ***Oberintegral*** das Infimum aller Obersummen, gebildet jeweils über sämtliche Zerlegungen $\mathscr{Z}$ von $[a, b]$. Stimmen beide überein, dann wird die Funktion f als ***Riemann-integrierbar*** bezeichnet, und der gemeinsamen Wert der Integrale ist das ***Riemann-Integral*** von f.

Diese Definitionen lassen sich problemlos auf höhere Dimension übertragen, indem man lediglich Intervalle durch ***achsenparallele kompakte Quader*** ersetzt, also Teilmengen von $Q \subseteq \mathbb{R}^n$ der Form $I_1 \times ... \times I_n$ mit $I_k = [a_k, b_k]$ und $a_k, b_k \in \mathbb{R}$, $a_k < b_k$ für $1 \leq k \leq n$. Der Einfachheit haber sprechen wir im Folgenden meist nur von ***Quadern***. Die Zahl

$$\nu(Q) \quad = \quad \prod_{k=1}^{n}(b_k - a_k)$$

bezeichnen wir als das (n-dimensionale) ***Volumen*** von Q. Der Begriff der Zerlegung muss ebenfalls geeignet angepasst werden.

(12.1) Definition. Sei $Q = I_1 \times ... \times I_n$ ein Quader im $\mathbb{R}^n$.

(i) Eine ***Zerlegung*** von Q ist ein Tupel $\mathscr{Z} = (\mathscr{Z}_1, ..., \mathscr{Z}_n)$, wobei $\mathscr{Z}_k$ für $1 \le k \le n$ jeweils eine Zerlegung des Intervalls I_k bezeichnet.

(ii) Ist $\mathscr{Z}$ eine solche Zerlegung, $I_k = [a_k, b_k]$ und $\mathscr{Z}_k = \{x_{k,1}, ..., x_{k,m_k-1}\}$ für $1 \le k \le n$ mit jeweils $a_k = x_{k,0} < x_{k,1} < ... < x_{k,m_k-1} < x_{k,m_k} = b_k$, dann bezeichnen wir eine Teilmenge $K \subseteq Q$ der Form

$$K \quad = \quad [x_{1,\ell_1-1}, x_{1,\ell_1}] \times ... \times [x_{n,\ell_n-1}, x_{n,\ell_n}]$$

mit $1 \le \ell_k \le m_k$ für $1 \le k \le n$ als ***Teilquader*** von Q bezüglich der Zerlegung $\mathscr{Z}$. Die Menge all dieser Teilquader bezeichnen wir mit $\mathscr{Q}(\mathscr{Z})$.

(iii) Ist $\mathscr{Z}' = (\mathscr{Z}'_1, ..., \mathscr{Z}'_n)$ eine weitere Zerlegung von Q, so bezeichnen wir $\mathscr{Z}'$ als ***Verfeinerung*** von $\mathscr{Z}$, wenn $\mathscr{Z}'_k \supseteq \mathscr{Z}_k$ für $1 \le k \le n$ erfüllt ist.

Je zwei Zerlegungen $\mathscr{Z}$ und $\mathscr{Z}'$ kann eine gemeinsame Verfeinerung zugeordnet werden, nämlich die Verfeinerung mit den Komponenten $\mathscr{Z}_k \cup \mathscr{Z}'_k$, die wir der Einfachheit halber mit $\mathscr{Z} \cup \mathscr{Z}'$ bezeichnen.

(12.2) Lemma. Sei Q ein Quader und $\mathscr{Z}$ eine Zerlegung von Q wie in Def. (12.1) angegeben. Dann addieren sich Volumina der Teilquader von Q bezüglich $\mathscr{Z}$ zum Volumen des Quaders Q, es gilt also

$$\sum_{K \in \mathscr{Q}(\mathscr{Z})} v(K) \quad = \quad v(Q).$$

Beweis: Dies ergibt sich aus der Rechnung

$$\sum_{K \in \mathscr{Q}(\mathscr{Z})} v(K) \quad = \quad \sum_{\ell_1=1}^{m_1} \cdots \sum_{\ell_n=1}^{m_n} v\left([x_{1,\ell_1-1}, x_{1,\ell_1}] \times ... \times [x_{n,\ell_n-1}, x_{n,\ell_n}]\right) \quad = \quad \sum_{\ell_1=1}^{m_1} \cdots \sum_{\ell_n=1}^{m_n} \prod_{k=1}^{n} (x_{k,\ell_k} - x_{k,\ell_k-1}) \quad =$$

$$\prod_{k=1}^{n} \left(\sum_{\ell_k=1}^{m_k} (x_{k,\ell_k} - x_{k,\ell_k-1}) \right) \quad = \quad \prod_{k=1}^{n} (x_{k,m_k} - x_{k,0}) \quad = \quad \prod_{k=1}^{n} (b_k - a_k) \quad = \quad v(Q). \qquad \square$$

Von der Anschauung her unmittelbar einleuchtend ist auch die folgende Aussage.

(12.3) Lemma. Seien $\mathscr{Z}, \mathscr{Z}'$ zwei Zerlegungen eines Quaders Q, wobei $\mathscr{Z}'$ eine Verfeinerung von $\mathscr{Z}$ sei. Dann definiert $\mathscr{Z}'$ für jeden Teilquader $K \in \mathscr{Q}(\mathscr{Z})$ auf natürliche Weise eine Zerlegung $\mathscr{Z}_K$ von K. Bezeichnen wir die Menge der Teilquader von K bezüglich $\mathscr{Z}_K$ mit $\mathscr{Q}(\mathscr{Z})_K$, dann gilt

$$\mathscr{Q}(\mathscr{Z}') = \bigcup_{K \in \mathscr{Q}(\mathscr{Z})} \mathscr{Q}(\mathscr{Z})_K \qquad \text{und} \qquad v(K) = \sum_{K' \in \mathscr{Q}(\mathscr{Z})_K} v(K') \quad ,$$

wobei es sich links um eine disjunkte Vereinigung handelt.

Beweis: Die zweite Gleichung ergibt sich unmittelbar aus Lemma (12.2), da die Elemente von $\mathscr{Q}(\mathscr{Z})_K$ jeweils die Teilquader von K bezüglich der Zerlegung $\mathscr{Z}_K$ sind. Wir müssen also lediglich die Definition von $\mathscr{Z}_K$ angeben und

die erste Gleichung sowie die Disjunktheit der Vereinigung überprüfen. Ist allgemein $S \subseteq \mathbb{R}$ eine endliche Teilmenge, so bezeichnen wir zwei Punkte $x, y \in S$ als *benachbart*, wenn kein $z \in S$ mit $x < z < y$ existiert. Ist $K = [c_1, d_1] \times \ldots \times [c_n, d_n]$, dann definieren wir die Zerlegung $\mathscr{Z}_K = (\mathscr{Z}_{K,1}, \ldots, \mathscr{Z}_{K,n})$ von K durch $\mathscr{Z}_{K,k} = \mathscr{Z}'_k \cap]c_k, d_k[$ für $1 \leq k \leq n$. Jeder Quader $K' \in \mathscr{Q}(\mathscr{Z}_K)$ ist in $\mathscr{Q}(\mathscr{Z}')$ enthalten. Denn ein solcher Quader hat die Form $K' = [c'_1, d'_1] \times \ldots \times [c'_n, d'_n]$, wobei c'_k, d'_k für $1 \leq k \leq n$ jeweils benachbarte Punkte von $\mathscr{Z}_{K,k} \cup \{c_k, d_k\}$ sind. Wegen $c_k, d_k \in \mathscr{Z}_k \cup \{a_k, b_k\}$ gilt auch $\mathscr{Z}_{K,k} \cup \{c_k, d_k\} = (\mathscr{Z}'_k \cup \{a_k, b_k\}) \cap [c_k, d_k]$, und dies zeigt, dass c'_k, d'_k zugleich auch benachbarte Punkte der Menge $\mathscr{Z}'_k \cup \{a_k, b_k\}$ sind. Damit ist $K' \in \mathscr{Q}(\mathscr{Z}')$ nachgewiesen.

Es ist auch klar, dass K der einzige Quader in $\mathscr{Q}(\mathscr{Z})$ mit $K' \in \mathscr{Q}(\mathscr{Z})_K$ ist. Sind nämlich K und K' wie oben definiert, dann folgt aus $K' \in \mathscr{Q}(\mathscr{Z})_K$, dass $c'_k, d'_k \in \mathscr{Z}_{K,k} \cup \{c_k, d_k\}$ für $1 \leq k \leq n$ gilt, mit $\mathscr{Z}_{K,k} = \mathscr{Z}'_k \cap]c_k, d_k[$, also insbesondere $c'_k, d'_k \in [c_k, d_k]$. Da sich die verschiedenen abgeschlossenen Intervalle $[c, d]$ gegeben durch benachbarte Punkte $c, d \in \mathscr{Z}_k \cup \{a_k, b_k\}$ sich in höchstens einem Punkt schneiden, ist $\{c_k, d_k\}$ durch diese Bedingung jeweils eindeutig festgelegt, und somit auch der Quader K. Dies zeigt, dass die Vereinigung $\bigcup_{K \in \mathscr{Q}(\mathscr{Z})} \mathscr{Q}(\mathscr{Z})_K$ tatsächlich disjunkt ist.

Setzen wir nun umgekehrt $K' \in \mathscr{Q}(\mathscr{Z}')$ voraus, mit $K' = [c'_1, d'_1] \times \ldots \times [c'_n, d'_n]$. Dann sind c'_k, d'_k jeweils benachbarte Punkte von $\mathscr{Z}'_k \cup \{a_k, b_k\}$, für $1 \leq k \leq n$. Sei $c_k = \max\{x \in \mathscr{Z}_k \cup \{a_k\} \mid x \leq c'_k\}$ und $d_k = \min\{x \in \mathscr{Z}_k \cup \{b_k\} \mid x \geq d'_k\}$. Wegen $\mathscr{Z}_k \subseteq \mathscr{Z}'_k$, und weil c'_k, d'_k in $\mathscr{Z}'_k$ benachbart sind, gibt es keinen Punkt $x \in \mathscr{Z}_k$ mit $c'_k < x < d'_k$, d.h. c_k und d_k sind benachbarte Punkte in $\mathscr{Z}_k \cup \{a_k, b_k\}$. Definieren wir nun $K = [c_1, d_1] \times \ldots \times [c_n, d_n]$, dann gilt jeweils $\mathscr{Z}_{K,k} = \mathscr{Z}'_k \cap]c_k, d_k[$, also $\mathscr{Z}_{K,k} \cup \{c_k, d_k\} = (\mathscr{Z}'_k \cup \{a_k, b_k\}) \cap [c_k, d_k]$, und somit sind c'_k, d'_k benachbarte Punkte von $\mathscr{Z}_{K,k} \cup \{c_k, d_k\}$. Es gilt also $K' \in \mathscr{Q}(\mathscr{Z}_K)$. □

Sei $Q \subseteq \mathbb{R}^n$ ein Quader, $\mathscr{Z}$ eine Zerlegung von Q und $f : Q \to \mathbb{R}$ eine beschränkte Funktion. Für jeden Quader $K \in \mathscr{Q}(\mathscr{Z})$ definieren wir die beiden Werte $c^-_{K,f} = \inf f(K)$ und $c^+_{K,f} = \sup f(K)$.

(12.4) Definition. Für jede Zerlegung $\mathscr{Z}$ von Q wie angegeben bezeichnet man die Werte

$$\mathscr{S}^-_f(\mathscr{Z}) \quad = \quad \sum_{K \in \mathscr{Q}(\mathscr{Z})} c^-_{K,f}\, v(K) \qquad \text{bzw.} \qquad \mathscr{S}^+_f(\mathscr{Z}) \quad = \quad \sum_{K \in \mathscr{Q}(\mathscr{Z})} c^+_{K,f}\, v(K)$$

als ***Unter-*** bzw. ***Obersumme*** von f bezüglich der Zerlegung $\mathscr{Z}$.

Vor dem Beweis des folgenden Lemmas erinnern wir an eine elementare Aussage der Analysis einer Variablen zum Infimum und Supremum: Sind $A, B \subseteq \mathbb{R}$ nichtleere beschränkte Teilmengen von $\mathbb{R}$ mit $A \subseteq B$, dann gilt $\inf(A) \geq \inf(B)$ und $\sup(A) \leq \sup(B)$. Denn wegen $A \subseteq B$ ist jede untere Schranke von B zugleich eine untere Schranke von A, die *größte* untere Schranke von A also mindestens so groß wie die größte untere Schranke von A. Der Beweis der zweiten Ungleich läuft analog.

(12.5) Lemma. Sind $\mathscr{Z}, \mathscr{Z}'$ Zerlegungen von Q, und ist $\mathscr{Z}'$ eine Verfeinerung von $\mathscr{Z}$, dann gilt

$$\mathscr{S}^-_f(\mathscr{Z}) \quad \leq \quad \mathscr{S}^-_f(\mathscr{Z}') \qquad \text{und} \qquad \mathscr{S}^+_f(\mathscr{Z}') \quad \leq \quad \mathscr{S}^+_f(\mathscr{Z}).$$

Beweis: Wir beschränken uns auf den Nachweis der ersten Ungleichung. Ist $K' \in \mathcal{Q}(Z')$ ein Teilquader von $K \in \mathcal{Q}(Z)$, dann gilt jeweils $c^-_{K,f} = \inf f(K) \le \inf f(K') = c^-_{K',f}$. Zusammen mit Lemma (12.3) erhalten wir

$$\begin{aligned} \mathcal{S}^-_f(\mathcal{Z}) &= \sum_{K\in\mathcal{Q}(\mathcal{Z})} c^-_{K,f}\, v(K) = \sum_{K\in\mathcal{Q}(\mathcal{Z})} \sum_{K'\in\mathcal{Q}(\mathcal{Z})_K} c^-_{K,f}\, v(K') = \sum_{K'\in\mathcal{Q}(\mathcal{Z}')} c^-_{K,f}\, v(K') \\ &\le \sum_{K'\in\mathcal{Q}(\mathcal{Z}')} c^-_{K',f}\, v(K') = \mathcal{S}^-_f(\mathcal{Z}'). \end{aligned}$$

□

(12.6) Folgerung. Für zwei beliebige Zerlegungen $\mathcal{Z}, \mathcal{Z}'$ von Q gilt $\mathcal{S}^-_f(\mathcal{Z}) \le \mathcal{S}^+_f(\mathcal{Z}')$.

Beweis: Es ist $\mathcal{Z}'' = \mathcal{Z} \cup \mathcal{Z}'$ eine gemeinsame Verfeinerung von $\mathcal{Z}$ und $\mathcal{Z}'$. Mit Lemma (12.5) erhalten wir somit die Abschätzungen $\mathcal{S}^-_f(\mathcal{Z}) \le \mathcal{S}^-_f(\mathcal{Z}'') \le \mathcal{S}^+_f(\mathcal{Z}'') \le \mathcal{S}^+_f(\mathcal{Z}')$. □

Wie im eindimensionalen Fall definieren wir nun

(12.7) Definition. Sei $Q \subseteq \mathbb{R}^n$ ein Quader und $f : Q \to \mathbb{R}$ eine beschränkte Funktion. Dann nennt man

$$\int_{Q\star} f(x)\,dx = \sup_{\mathcal{Z}} \mathcal{S}^-_f(\mathcal{Z}) \quad \text{bzw.} \quad \int^{\star}_{Q} f(x)\,dx = \inf_{\mathcal{Z}} \mathcal{S}^+_f(\mathcal{Z})$$

das ***Unter-*** bzw. ***Oberintegral*** von f, wobei $\mathcal{Z}$ bei der Bildung von Supremum und Infimum jeweils alle Zerlegungen des Quaders Q durchläuft. Man bezeichnet f als ***Riemann-integrierbar***, wenn Unter- und Oberintegral übereinstimmen. In diesem Fall nennt man

$$\int_Q f(x)\,dx = \int_{Q\star} f(x)\,dx \quad \textbf{\textit{Riemann-Integral}} \quad \text{von } f.$$

Mit Folgerung (12.6) erhält man für jede beschränkte Funktion f die Ungleichung

$$\int_{Q\star} f(x)\,dx = \sup_{\mathcal{Z}} \mathcal{S}^-_f(\mathcal{Z}) \le \inf_{\mathcal{Z}} \mathcal{S}^+_f(\mathcal{Z}) = \int^{\star}_{Q} f(x)\,dx.$$

Jede konstante Funktion auf einem Quader Q ist Riemann-integrierbar. Sei nämlich $f : Q \to \mathbb{R}$ eine Funktion mit $f(x) = c$ für alle $x \in Q$ und $\mathcal{Z}$ eine beliebige Zerlegung. Dann gilt $c^-_{K,f} = c$ für alle $K \in \mathcal{Q}(\mathcal{Z})$. Es folgt

$$\mathcal{S}^-_f(\mathcal{Z}) = \sum_{K\in\mathcal{Q}(\mathcal{Z})} c^-_{K,f}\, v(K) = \sum_{K\in\mathcal{Q}(\mathcal{Z})} c\,v(K) = c \sum_{K\in\mathcal{Q}(\mathcal{Z})} v(K) = c\,v(Q)$$

und somit

$$\int_{\star Q} f(x)\,dx = \sup_{\mathcal{Z}} \mathcal{S}^-_f(\mathcal{Z}) = c\,v(Q).$$

Genauso beweist man, dass auch das Oberintegral den Wert $c\,v(Q)$ hat.

Als ein etwas weniger triviales Beispiel betrachten wir die Funktion $f : [0,1] \times [0,1] \to \mathbb{R}$, $(x,y) \mapsto xy$. Für jedes $n \in \mathbb{N}$ betrachten wir die Zerlegung $\mathscr{Z}_1^{(n)} = \mathscr{Z}_2^{(n)} = \{\frac{k}{n} \mid 1 \le k \le n-1\}$ des Intervalls $[0,1]$ und die Zerlegung $\mathscr{Z}^{(n)} = (\mathscr{Z}_1^{(n)}, \mathscr{Z}_2^{(n)})$ des Quaders $Q = [0,1] \times [0,1]$. Die Quader dieser Zerlegung sind offenbar gegeben durch $\mathscr{Q}(\mathscr{Z}^{(n)}) = \{ Q_{k\ell} \mid 1 \le k,\ell \le n\}$ mit $Q_{k\ell} = [\frac{k-1}{n}, \frac{k}{n}] \times [\frac{\ell-1}{n}, \frac{\ell}{n}]$, und es gilt $v(Q_{k\ell}) = \frac{1}{n^2}$ für alle k,ℓ. Nach Definition der Funktion f gilt

$$c_{k\ell}^- \;=\; c_{Q_{k\ell}}^- \;=\; \frac{(k-1)(\ell-1)}{n^2} \qquad \text{und} \qquad c_{k\ell}^+ \;=\; c_{Q_{k\ell}}^+ \;=\; \frac{k\ell}{n^2}$$

für $1 \le k,\ell \le n$. Als Untersumme erhalten wir

$$\mathscr{S}_f^-(\mathscr{Z}^{(n)}) \;=\; \sum_{K \in \mathscr{Q}(\mathscr{Z}^{(n)})} c_{K,f}^- v(K) \;=\; \sum_{k=1}^n \sum_{\ell=1}^n \frac{c_{k\ell}^-}{n^2} \;=\; \sum_{k=1}^n \sum_{\ell=1}^n \frac{(k-1)(\ell-1)}{n^4} \;=$$

$$\frac{1}{n^4}\left(\sum_{k=1}^n (k-1)\right)\left(\sum_{\ell=1}^n (\ell-1)\right) \;=\; \frac{1}{n^4} \cdot \tfrac{1}{2}(n-1)n \cdot \tfrac{1}{2}(n-1)n \;=\; \tfrac{1}{4}\left(1 - \frac{1}{n}\right)^2$$

und für die Obersumme ebenso

$$\mathscr{S}_f^+(\mathscr{Z}^{(n)}) \;=\; \sum_{K \in \mathscr{Q}(\mathscr{Z}^{(n)})} c_{K,f}^+ v(K) \;=\; \sum_{k=1}^n \sum_{\ell=1}^n \frac{c_{k\ell}^+}{n^2} \;=\; \sum_{k=1}^n \sum_{\ell=1}^n \frac{k\ell}{n^4} \;=$$

$$\frac{1}{n^4}\left(\sum_{k=1}^n k\right)\left(\sum_{\ell=1}^n \ell\text{‘}\right) \;=\; \frac{1}{n^4} \cdot \tfrac{1}{2}n(n+1) \cdot \tfrac{1}{2}n(n+1) \;=\; \tfrac{1}{4}\left(1 + \frac{1}{n}\right)^2.$$

Wir erhalten somit für jedes $n \in \mathbb{N}$ die Ungleichungskette

$$\tfrac{1}{4}\left(1 - \tfrac{1}{n}\right)^2 \;=\; \mathscr{S}_f^-(\mathscr{Z}^{(n)}) \;\le\; \int_{Q\bigstar} f(x)\,dx \;\le\; \int_Q^\bigstar f(x)\,dx \;\le\; \mathscr{S}_f^+(\mathscr{Z}^{(n)}) \;=\; \tfrac{1}{4}\left(1 + \tfrac{1}{n}\right)^2.$$

Durch Grenzübergang $n \to \infty$ folgt daraus schließlich $\displaystyle\int_Q f(x)\,dx \;=\; \tfrac{1}{4}$.

(12.8) Proposition. Sei $Q \subseteq \mathbb{R}^n$ ein Quader. Eine beschränkte Funktion $f : Q \to \mathbb{R}$ ist genau dann Riemann-integrierbar, wenn es für jedes $\varepsilon \in \mathbb{R}^+$ eine Zerlegung $\mathscr{Z}$ von Q gibt mit

$$\mathscr{S}_f^+(\mathscr{Z}) - \mathscr{S}_f^-(\mathscr{Z}) \;<\; \varepsilon.$$

Beweis: „⇒“ Sei $\varepsilon \in \mathbb{R}^+$ vorgegeben und s der Wert des Riemann-Integrals. Nach Definition des Unter- und Oberintegrals als Supremum und Infimum gibt es Zerlegungen $\mathscr{Z}$ und $\mathscr{Z}'$ mit

$$\mathscr{S}_f^-(\mathscr{Z}) \;>\; \int_{Q\bigstar} f(x)\,dx - \tfrac{1}{2}\varepsilon \;=\; s - \tfrac{1}{2}\varepsilon \qquad \text{und} \qquad \mathscr{S}_f^+(\mathscr{Z}') \;<\; \int_Q^\bigstar f(x)\,dx + \tfrac{1}{2}\varepsilon \;=\; s + \tfrac{1}{2}\varepsilon.$$

Ist $\mathscr{Z}''$ eine gemeinsame Verfeinerung von $\mathscr{Z}$ und $\mathscr{Z}'$, dann erhalten wir die Abschätzung $\mathscr{S}_f^+(\mathscr{Z}'') - \mathscr{S}_f^-(\mathscr{Z}'') \le \mathscr{S}_f^+(\mathscr{Z}') - \mathscr{S}_f^-(\mathscr{Z}) < (s + \frac{1}{2}\varepsilon) - (s - \frac{1}{2}\varepsilon) = \varepsilon$.

„⇐“ Für beliebig vorgegebenes $\varepsilon \in \mathbb{R}^+$ sei $\mathscr{Z}$ eine Zerlegung mit der angegebenen Eigenschaft. Dann gilt

$$\mathscr{S}_f^-(\mathscr{Z}) \;\le\; \int_{Q\bigstar} f(x)\,dx \;\le\; \int_Q^\bigstar f(x)\,dx \;\le\; \mathscr{S}_f^+(\mathscr{Z}).$$

Nach Voraussetzung ist die Differenz von Ober- und Untersumme kleiner als ε, also gilt diese Abschätzung erst recht für die Differenz von $\int_Q^\star f(x)\,dx$ und $\int_{Q\star} f(x)\,dx$. Da $\varepsilon \in \mathbb{R}^+$ beliebig klein gewählt werden kann, folgt $\int_Q^\star f(x)\,dx = \int_{Q\star} f(x)\,dx$, und damit ist f Riemann-integrierbar. □

(12.9) Proposition. Sei $Q \subseteq \mathbb{R}^n$ ein Quader, seien $f, g : Q \to \mathbb{R}$ Riemann-integrierbare Funktionen und $\lambda \in \mathbb{R}$. Dann sind auch f und g Riemann-integrierbar, und es gilt

$$\int_Q (f+g)(x)\,dx \quad = \quad \int_Q f(x)\,dx + \int_Q g(x)\,dx \quad \text{und} \quad \int_Q (\lambda f)(x)\,dx \quad = \quad \lambda \int_Q f(x)\,dx.$$

Aus $f \le g$ folgt $\displaystyle\int_Q f(x)\,dx \le \int_Q g(x)\,dx$.

Beweis: Wir beschränken uns auf den Beweis der ersten Gleichung. Sei $\varepsilon \in \mathbb{R}^+$ vorgegeben. Auf Grund der Integrierbarkeit von f und g gibt es nach Prop. (12.8) Zerlegungen $\mathscr{Z}'$, $\mathscr{Z}''$ von Q mit $\mathscr{S}_f^+(\mathscr{Z}') - \mathscr{S}_f^-(\mathscr{Z}') < \varepsilon$ und ebenso $\mathscr{S}_g^+(\mathscr{Z}'') - \mathscr{S}_g^-(\mathscr{Z}'') < \varepsilon$. Ist $\mathscr{Z}$ eine gemeinsame Verfeinerung von $\mathscr{Z}'$ und $\mathscr{Z}''$, dann sind die beiden Ungleichungen nach Lemma (12.5) auch für $\mathscr{Z}$ an Stelle von $\mathscr{Z}'$ bzw. $\mathscr{Z}''$ gültig. Für alle $K \in \mathscr{Q}(\mathscr{Z})$ gilt

$$c^-_{K,f+g} \quad = \quad \inf(f+g)(K) \quad \ge \quad \inf f(K) + \inf g(K) \quad = \quad c^-_{K,f} + c^-_{K,g} \quad ,$$

denn wegen $\inf f(K) + \inf g(K) \le f(x) + g(x) = (f+g)(x)$ für alle $x \in K$ ist $\inf f(K) + \inf g(K)$ eine untere Schranke von $(f+g)(K)$, und nach Definition ist $\inf(f+g)(K)$ die größte untere Schranke. Genauso beweist man $c^+_{K,f+g} \le c^+_{K,f} + c^+_{K,g}$ für alle $K \in \mathscr{Q}(\mathscr{Z})$. Durch Summation über alle Teilquader $K \in \mathscr{Q}(\mathscr{Z})$ erhält man

$$\mathscr{S}^-_{f+g}(\mathscr{Z}) \ge \mathscr{S}^-_f(\mathscr{Z}) + \mathscr{S}^-_g(\mathscr{Z}) \qquad \text{und} \qquad \mathscr{S}^+_{f+g}(\mathscr{Z}) \le \mathscr{S}^+_f(\mathscr{Z}) + \mathscr{S}^-_g(\mathscr{Z}).$$

Es folgt

$$\mathscr{S}^-_f(\mathscr{Z}) + \mathscr{S}^-_g(\mathscr{Z}) \quad \le \quad \mathscr{S}^-_{f+g}(\mathscr{Z}) \quad \le \quad \mathscr{S}^+_{f+g}(\mathscr{Z}) \quad \le \quad \mathscr{S}^+_f(\mathscr{Z}) + \mathscr{S}^+_g(\mathscr{Z})$$

Da die Differenz von rechter und linker Seite kleiner als 2ε ist, gilt dasselbe erst recht für die Differenz $\mathscr{S}^+_{f+g}(\mathscr{Z}) - \mathscr{S}^-_{f+g}(\mathscr{Z})$. Nach Prop. (12.8) zeigt dies, dass $f+g$ Riemann-integrierbar ist. Was den Wert des Integrals angeht, gilt sowohl

$$\mathscr{S}^-_f(\mathscr{Z}) + \mathscr{S}^-_g(\mathscr{Z}) \quad \le \quad \mathscr{S}^-_{f+g}(\mathscr{Z}) \quad \le \quad \int_Q (f+g)(x)\,dx \quad \le \quad \mathscr{S}^+_{f+g}(\mathscr{Z}) \quad \le \quad \mathscr{S}^+_f(\mathscr{Z}) + \mathscr{S}^+_g(\mathscr{Z})$$

als auch

$$\mathscr{S}^-_f(\mathscr{Z}) + \mathscr{S}^-_g(\mathscr{Z}) \quad \le \quad \int_Q f(x)\,dx + \int_Q g(x)\,dx \quad \le \quad \mathscr{S}^+_f(\mathscr{Z}) + \mathscr{S}^+_g(\mathscr{Z}).$$

Da rechte und linke Seite der beiden Ungleichungsketten eine Differenz $< 2\varepsilon$ haben, erhalten wir für die Differenz der Integrale der Mitte den Abstand $|\int_Q f(x)\,dx + \int_Q g(x)\,dx - \int_Q (f+g)(x)\,dx| < 2\varepsilon$. Weil $\varepsilon \in \mathbb{R}^+$ beliebig klein gewählt werden kann, erhalten wir $\int_Q f(x)\,dx + \int_Q g(x)\,dx = \int_Q (f+g)(x)\,dx$.

Für den Beweis der „≤“-Aussage genügt es zu bemerken, dass für jede Zerlegung $\mathscr{Z}$ von Q und jedes $K \in \mathscr{Q}(\mathscr{Z})$ aus $f \le g$ auch $c^-_{K,f} \le c^-_{K,g}$ folgt. Dies wiederum bedeutet $\mathscr{S}^-_f(\mathscr{Z}) \le \mathscr{S}^-_g(\mathscr{Z})$ für jede Zerlegung $\mathscr{Z}$, und aus der Definition des Riemann-Integrals folgt

$$\int_Q f(x)\,dx \quad = \quad \int_{Q\star} f(x)\,dx \quad = \quad \sup_{\mathscr{Z}} \mathscr{S}^-_f(\mathscr{Z}) \quad \le \quad \sup_{\mathscr{Z}} \mathscr{S}^-_g(\mathscr{Z}) \quad = \quad \int_{Q\star} g(x)\,dx \quad = \quad \int_Q g(x)\,dx. \quad \square$$

(12.10) Proposition. Jede stetige Funktion $f : Q \to \mathbb{R}$ auf einem Quader Q ist Riemann-integrierbar.

Beweis: Sei $\varepsilon \in \mathbb{R}^+$ beliebig vorgegeben und $\|\cdot\|_\infty$ die Maximums-Norm auf $\mathbb{R}^n$. Nach Satz (5.19) ist jede stetige Funktion auf einer kompakten Menge gleichmäßig stetig. Es gibt also ein $\delta \in \mathbb{R}^+$, so dass für alle $x, y \in Q$ aus $\|x-y\|_\infty < \delta$ jeweils $|f(x)-f(y)| < \varepsilon$ folgt. Sei nun $\mathscr{Z}$ eine so feine Zerlegung von Q, dass der Durchmesser von jedem Quader $K \in \mathscr{Z}$ jeweils kleiner als δ ist. Dann gilt für jeden Quader $K \in \mathscr{Z}$ und alle $x, y \in K$ jeweils $|f(x)-f(y)| < \varepsilon$.

Daraus folgt jeweils $c^+_{K,f} - c^-_{K,f} \le \varepsilon$. Denn wegen $f(x) < \varepsilon + f(y)$ für alle $x, y \in K$ ist $\varepsilon + f(y)$ für jedes $y \in K$ eine obere Schranke von $f(K)$, es gilt also $\sup f(K) \le \varepsilon + f(y)$ für alle $y \in K$. Aus $f(y) \ge \sup f(K) - \varepsilon$ für alle $y \in K$ wiederum folgt, dass $\sup f(K) - \varepsilon$ eine untere Schranke von $f(K)$ ist. Daraus folgt $\sup f(K) - \varepsilon \le \inf f(K)$, was zu $c^+_{K,f} - c^-_{K,f} = \sup f(K) - \inf f(K) \le \varepsilon$ umgeformt werden kann. Diese Abschätzung wiederum liefert

$$\begin{aligned} \mathscr{S}^+_f(\mathscr{Z}) - \mathscr{S}^-_f(\mathscr{Z}) &= \sum_{K\in\mathscr{Q}(\mathscr{Z})} c^+_{K,f} v(K) - \sum_{K\in\mathscr{Q}(\mathscr{Z})} c^-_{K,f} v(K) = \sum_{K\in\mathscr{Q}(\mathscr{Z})} (c^+_{K,f} - c^-_{K,f}) v(K) \\ &\le \varepsilon \sum_{K\in\mathscr{Q}(\mathscr{Z})} v(K) \le \varepsilon v(Q). \end{aligned}$$

Weil ε beliebig klein gewählt werden kann, folgt mit Prop. (12.8) daraus die Behauptung. □

Andererseits gibt es wie im eindimensionalen auch in jeder höheren Dimension Funktionen, die nicht Riemann-integrierbar sind. Sei zum Beispiel $Q = [0,1]^n$ und die Funktion $f : Q \to \mathbb{R}$ definiert durch

$$f(x_1, \ldots, x_n) = \begin{cases} 1 & \text{falls } x_1, \ldots, x_n \in \mathbb{Q} \\ 0 & \text{sonst} \end{cases}$$

Sei $\mathscr{Z}$ eine beliebige Zerlegung von Q und $K \in \mathscr{Q}(\mathscr{Z})$. Wir schreiben K in der Form $[a_1, b_1] \times \ldots \times [a_n, b_n]$ mit $a_k < b_k$ für $1 \le k \le n$. Weil die rationalen Zahlen dicht in $\mathbb{R}$ liegen, gibt es für jedes k ein $x_k \in [a_k, b_k] \cap \mathbb{Q}$. Es folgt $f(x) = 1$ für den Punkt $x = (x_1, \ldots, x_n)$ in K und somit $c^+_{K,f} \ge 1$. Weil auch die irrationalen Zahlen dicht in $\mathbb{R}$ liegen, finden wir ebenso ein $x \in K$ mit $f(x) = 0$. Daraus folgt $c^-_{K,f} \le 0$. Insgesamt erhalten wir für jede Zerlegung $\mathscr{Z}$ also

$$\mathscr{S}^-_f(\mathscr{Z}) = \sum_{K\in\mathscr{Q}(\mathscr{Z})} c^-_{K,f} v(K) \le 0 \qquad \text{und} \qquad \mathscr{S}^+_f(\mathscr{Z}) = \sum_{K\in\mathscr{Q}(\mathscr{Z})} c^+_{K,f} v(K) \ge \sum_{K\in\mathscr{Q}(\mathscr{Z})} 1 \cdot v(K) = v(Q) = 1.$$

Daraus folgt $\int_{Q\bigstar} f(x)\, dx \le 0$ und $\int^{\bigstar}_Q f(x)\, dx \ge 1$. Die Funktion ist also nicht Riemann-integrierbar.

§ 13. Der Satz von Fubini

Zusammenfassung. Als der aus Anwendungssicht wichtigste Satz der mehrdimensionalen Integralrechnung kann der *Satz von Fubini* angesehen werden. Durch ihn ist es möglich, die Berechnung eines mehrdimensionalen Integrals auf die Berechnung mehrerer eindimensionaler Integrale zurückzuführen. Als Anwendung berechnen wir mit diesem Satz die Volumina der Kugel und des dreidimensionalen Simplex.

Wichtige Begriffe und Sätze in diesem Kapitel

– Satz von Fubini

Für die Berechnung mehrdimensionaler Integrale kommt folgender Satz zur Anwendung, der die mehrdimensionale Integration auf den eindimensionalen Fall zurückführt.

(13.1) Satz. *(Satz von Fubini)*

Seien $P \subseteq \mathbb{R}^m$ und $Q \subseteq \mathbb{R}^n$ Quader, und sei $f : P \times Q \to \mathbb{R}$ eine Riemann-integrierbare Funktion auf dem Quader $P \times Q \subseteq \mathbb{R}^m \times \mathbb{R}^n$. Für jedes $x \in P$ sei die Funktion $f_x : Q \to \mathbb{R}$ definiert durch $f_x(y) = f(x,y)$. Dann sind die Funktionen

$$f_\bigstar : P \to \mathbb{R} \quad , \quad x \mapsto \int_{Q\bigstar} f_x(y)\,dy \qquad \text{und} \qquad f^\bigstar : P \to \mathbb{R} \quad , \quad x \mapsto \int_Q^\bigstar f_x(y)\,dy$$

beide Riemann-integrierbar, und es gilt

$$\int_{P\times Q} f(x,y)\,d(x,y) \quad = \quad \int_P \left(\int_{Q\bigstar} f_x(y)\,dy \right) dx \quad = \quad \int_P \left(\int_Q^\bigstar f_x(y)\,dy \right) dx.$$

Beweis: Sei $\mathscr{Z}_P = (\mathscr{Z}_1, \ldots, \mathscr{Z}_m)$ eine Zerlegung von P und $\mathscr{Z}_Q = (\mathscr{Z}_{m+1}, \ldots, \mathscr{Z}_{m+n})$ eine Zerlegung von Q. Dann ist durch $\mathscr{Z} = (\mathscr{Z}_1, \ldots, \mathscr{Z}_{m+n})$ eine Zerlegung von $P \times Q$ gegeben, und jede Zerlegung von $P \times Q$ kann auf diese Weise aus Zerlegungen von P und Q zusammengesetzt werden. Außerdem gilt

$$\mathscr{Q}(\mathscr{Z}) \quad = \quad \left\{ K_P \times K_Q \;\middle|\; K_P \in \mathscr{Q}(\mathscr{Z}_P) \text{ und } K_Q \in \mathscr{Q}(\mathscr{Z}_Q) \right\}.$$

Unser Ansatz zum Beweis der Riemann-Integrierbarkeit von $f_\bigstar$ und $f^\bigstar$ ist folgender: Sei $\varepsilon \in \mathbb{R}^+$ vorgegeben. Weil f Riemann-integrierbar ist, können wir durch Wahl einer geeigneten Zerlegung $\mathscr{Z}$ erreichen, dass die Differenz $\mathscr{S}_f^+(\mathscr{Z}) - \mathscr{S}_f^-(\mathscr{Z})$ kleiner als ε wird. Wenn wir nun zeigen könen, dass $\mathscr{S}_{f_\bigstar}^-(\mathscr{Z}_P) \geq \mathscr{S}_f^-(\mathbb{Z})$ und $\mathscr{S}_{f_\bigstar}^+(\mathscr{Z}_P) \leq \mathscr{S}_f^+(\mathscr{Z})$ gilt, dann erhalten wir auch $\mathscr{S}_{f_\bigstar}^+(\mathscr{Z}_P) - \mathscr{S}_{f_\bigstar}^-(\mathscr{Z}_P) < \varepsilon$, und daraus ergibt sich die Riemann-Integrierbarkeit von $f_\bigstar$. Durch eine vollkommen analoge Überlegung (die wir hier nicht ausführen), ergibt sich auch die Riemann-Integrierbarkeit der Funktion $f^\bigstar$.

Um die Abschätzung $\mathscr{S}^-_{f_\star}(\mathscr{Z}_P) \geq \mathscr{S}^-_f(\mathscr{Z})$ herzuleiten, bemerken wir zunächst, dass

$$\mathscr{S}^-_f(\mathscr{Z}) \;=\; \sum_{K_P\in\mathscr{Q}(\mathscr{Z}_P)}\sum_{K_Q\in\mathscr{Q}(\mathscr{Z}_Q)} c^-_{K_P\times K_Q,f}\,\nu(K_P\times K_Q) \;=\; \sum_{K_P\in\mathscr{Q}(\mathscr{Z}_P)}\left(\sum_{K_Q\in\mathscr{Q}(\mathscr{Z}_Q)} c^-_{K_P\times K_Q,f}\,\nu(K_Q)\right)\nu(K_P)$$

und $\mathscr{S}^-_{f_\star}(\mathscr{Z}_P) = \sum_{K_P\in\mathscr{Q}(\mathscr{Z}_P)} c^-_{K_P,f_\star}\,\nu(K_P)$ gilt. Wir müssen also für jedes $K_P\in\mathscr{Q}(\mathscr{Z}_P)$ die Ungleichung

$$\sum_{K_Q\in\mathscr{Q}(\mathscr{Z}_Q)} c^-_{K_P\times K_Q,f}\,\nu(K_Q) \quad\leq\quad c^-_{K_P,f_\star} \tag{5}$$

beweisen. Nach Definition ist $c^-_{K_P,f_\star} = \inf f_\star(K_P)$, und für jedes $x_0\in K_P$ ist $f_\star(x_0) = \int_{Q\star} f_{x_0}(y)\,dy$. Das Unterintegral wiederum kann nach unten abgeschätzt werden durch die Untersumme $\mathscr{S}^-_{f_{x_0}}(\mathscr{Z}_Q) = \sum_{K_Q\in\mathscr{Q}(\mathscr{Z}_Q)} c^-_{K_Q,f_{x_0}}\,\nu(K_Q)$. Weiter gilt

$$c^-_{K_P\times K_Q} \quad=\quad \inf\{f(x,y)\mid x\in K_P, y\in K_Q\} \quad\leq\quad \inf\{f(x_0,y)\mid y\in K_Q\} \quad=\quad c^-_{K_Q,f_{x_0}}.$$

Insgesamt erhalten wir damit für jedes $x_0\in K_P$ die Ungleichung

$$\sum_{K_Q\in\mathscr{Q}(\mathscr{Z}_Q)} c^-_{K_P\times K_Q,f}\,\nu(K_Q) \quad\leq\quad \sum_{K_Q\in\mathscr{Q}(\mathscr{Z}_Q)} c^-_{K_Q,f_{x_0}}\,\nu(K_Q) \quad\leq\quad \int_{Q\star} f_{x_0}(y)\,dy \quad=\quad f_\star(x_0)$$

Die Summe $\sum_{K_Q\in\mathscr{Q}(\mathscr{Z}_Q)} c^-_{K_P\times K_Q,f}\,\nu(K_Q)$ ist also eine untere Schranke für $f_\star(K_P)$, und wegen $c^-_{K_P,f_\star} = \inf f_\star(K_P)$ erhalten wir (5), wie gewünscht. Wie zu Beginn erläutert, ist die Riemann-Integrierbarkeit von $f_\star$ damit bewiesen.

In einem letzten Schritt beweisen wir nun die im Satz angegebenen Integralgleichungen. Für jede Zerlegung $\mathscr{Z}$ von $P\times Q$ gilt einerseits

$$\mathscr{S}^-_f(\mathscr{Z}) \quad\leq\quad \mathscr{S}^-_{f_\star}(\mathscr{Z}_P) \quad\leq\quad \int_P f_\star(x)\,dx \quad\leq\quad \int_P f^\star(x)\,dx \quad\leq\quad \mathscr{S}^+_{f^\star}(\mathscr{Z}_P) \quad\leq\quad \mathscr{S}^+_f(\mathscr{Z}) \quad,$$

andererseits aber auch

$$\mathscr{S}^-_f(\mathscr{Z}) \quad\leq\quad \int_{P\times Q} f(x,y)\,d(x,y) \quad\leq\quad \mathscr{S}^+_f(\mathscr{Z}).$$

Ist nun $\varepsilon\in\mathbb{R}^+$ vorgegeben und $\mathscr{Z}$ so gewählt, dass die Differenz $\mathscr{S}^+_f(\mathscr{Z}) - \mathscr{S}^-_f(\mathscr{Z})$ kleiner als ε wird, so folgt sowohl $|\int_{P\times Q} f(x,y)\;d(x,y) - \int_P f_\star(x)\;dx| < \varepsilon$ als auch $|\int_{P\times Q} f(x,y)\;d(x,y) - \int_P f^\star(x)\;dx| < \varepsilon$. Da ε beliebig klein gewählt werden kann, stimmen alle drei Integrale überein. □

Häufig ist f in Anwendungsfällen eine stetige Funktion auf $P\times Q$. In diesem Fall ist auch f_x für alle $x\in P$ jeweils auf ganz Q stetig und somit nach Satz (12.10) Riemann-integrierbar. Dies bedeutet, dass die Funktionen $f_\star$ und $f^\star$ zusammenfallen und wir den Satz von Fubini einfach in der Form

$$\int_{P\times Q} f(x,y)\,d(x,y) \quad=\quad \int_P\left(\int_Q f_x(y)\,dy\right)dx \quad=\quad \int_P\left(\int_Q f(x,y)\,dy\right)dx \qquad \text{schreiben können.}$$

Beispiel 1:
Betrachten wir noch einmal die Funktion $f(x,y) = xy$ auf $P\times Q = [0,1]^2$ aus dem vorherigen Abschnitt. Offenbar ist f auf ihrem gesamten Definitionsbereich stetig. Deshalb ist der Satz von Fubini anwendbar, und wir erhalten

$$\int_{P\times Q} f(x,y)\,d(x,y) \quad=\quad \int_P\left(\int_Q xy\;dy\right)dx \quad=\quad \int_0^1\left(\int_0^1 xy\;dy\right)dx \quad=\quad \int_0^1 \left[\tfrac{1}{2}xy^2\right]_{y=0}^{y=1} dx$$

$$=\quad \int_0^1 \tfrac{1}{2}x\;dx \quad=\quad \left[\;\tfrac{1}{4}x^2\;\right]_0^1 \quad=\quad \tfrac{1}{4}.$$

□

Beispiel 2: *Volumen des Simplex*

Als ***Simplex*** bezeichnen wir die Menge gegeben durch

$$S \quad = \quad \{(x,y,z) \in [0,1]^3 \ \mid \ x+y+z \leq 1\}.$$

Definieren wir die Funktion $f : [0,1]^2 \to \mathbb{R}$ durch

$$f(x,y) \quad = \quad \begin{cases} 1-x-y & \text{falls } x+y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

dann ist das Volumen von S das Integral der Funktion f über $[0,1]^2$. Für jedes $x \in [0,1]$ ist die Funktion $f_x : [0,1] \to \mathbb{R}$ definiert durch

$$f_x(y) \quad = \quad \begin{cases} 1-x-y & \text{für } 0 \leq y \leq 1-x \\ 0 & \text{sonst.} \end{cases}$$

Das Integral dieser Funktion über $[0,1]$ ist gegeben durch

$$\int_0^1 f_x(y)\,dy \quad = \quad \int_0^{1-x} (1-x-y)\,dy \quad = \quad \left[y - xy - \tfrac{1}{2}y^2\right]_{y=0}^{y=1-x} \quad = \quad (1-x) - x(1-x) - \tfrac{1}{2}(1-x)^2$$

$$= \quad (1-x) + \left(-x+x^2\right) + \left((-\tfrac{1}{2}) + x - \tfrac{1}{2}x^2\right) \quad = \quad \tfrac{1}{2}x^2 - x + \tfrac{1}{2}.$$

Setzen wir die Stetigkeit von f als bekannt voraus, dann liefert der Satz von Fubini

$$\int_{[0,1]^2} f(x,y)\,d(x,y) \quad = \quad \int_0^1 \left(\int_0^1 f_x(y)\,dy \right) dx \quad = \quad \int_0^1 \left(x^2 - x + \tfrac{1}{2}\right) dx \quad =$$

$$\left[\tfrac{1}{6}x^3 - \tfrac{1}{2}x^2 + \tfrac{1}{2}x\right]_0^1 \quad = \quad \tfrac{1}{6} - \tfrac{1}{2} + \tfrac{1}{2} \quad = \quad \tfrac{1}{6}. \qquad \square$$

Beispiel 3: *Volumen der Einheitskugel*

Zur Vorbereitung berechnen wir das Integral $\int_{-r}^{r} \sqrt{r^2 - x^2}\,dx$ für beliebiges $r \in \mathbb{R}^+$. In der Analysis einer Variablen wurde die Gleichung für den Flächeninhalt des Halbkreises

$$\int_{-1}^{1} \sqrt{1-x^2}\,dx \quad = \quad \tfrac{1}{2}\pi$$

bewiesen. Mit Hilfe der Substitutionsregel folgt daraus für $r \in \mathbb{R}^+$ jeweils

$$\int_{-r}^{r} \sqrt{r^2-x^2}\,dx \quad = \quad r\int_{-r}^{r} \sqrt{1-(\tfrac{x}{r})^2}\,dx \quad = \quad r^2 \int_{-r}^{r} \tfrac{1}{r} \cdot \sqrt{1-(\tfrac{x}{r})^2}\,dx \quad = \quad r^2 \int_{-1}^{1} \sqrt{1-x^2}\,dx \quad = \quad \tfrac{1}{2}\pi r^2.$$

Das Volumen der *Halbkugel* ist nun offenbar das Integral der Funktion

$$f(x,y) \quad = \quad \begin{cases} \sqrt{1-x^2-y^2} & \text{falls } x^2+y^2 \leq 1 \\ 0 & \text{sonst} \end{cases}$$

über den Bereich $Q = [-1,1]^2$. In diesem Fall ist $f_x(y)$ für jedes $x \in [-1,1]$ gegeben durch $f_x(y) = \sqrt{1-x^2-y^2}$ für $-\sqrt{1-x^2} \le y \le \sqrt{1-x^2}$ und $f_x(y) = 0$ für alle $y \notin [-\sqrt{1-x^2}, \sqrt{1-x^2}]$. Mit Hilfe des soeben berechneten Integrals $\int_{-r}^{r} \sqrt{r^2-x^2}\,dx$ erhalten wir nun

$$\int_{-1}^{1} f_x(y)\,dy \quad = \quad \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \sqrt{1-x^2-y^2}\,dy \quad = \quad \tfrac{1}{2}\pi(1-x^2)$$

für alle $x \in [-1,1]$. Weil die Funktion f stetig ist, können wir den Satz von Fubini anwenden und erhalten

$$\int_{[-1,1]^2} f(x,y)\,d(x,y) \quad = \quad \int_{-1}^{1}\left(\int_{-1}^{1} f_x(y)\,dy\right)dx \quad = \quad \int_{-1}^{1} \tfrac{1}{2}\pi(1-x^2)\,dx \quad = \quad \tfrac{1}{2}\pi\int_{-1}^{1}\left(1-x^2\right)\,dx$$

$$= \quad \tfrac{1}{2}\pi\left[x-\tfrac{1}{3}x^3\right]_{-1}^{1} \quad = \quad \tfrac{1}{2}\pi\left(\tfrac{2}{3}-(-\tfrac{2}{3})\right) \quad = \quad \tfrac{1}{2}\pi\cdot\tfrac{4}{3} \quad = \quad \tfrac{2}{3}\pi.$$

Dies ist das Volumen der Halbkugel. Das Volumen der (Voll-)Kugel vom Radius 1 beträgt also $2\cdot\frac{2}{3}\pi = \frac{4}{3}\pi$.

§ 14. Nullmengen und Lebesguesches Integrabilitätskriterium

Zusammenfassung. Wie wir in § 12 gezeigt haben, sind stetige Funktionen auf einem Quader stets integrierbar. Andererseits wissen wir bereits aus dem ersten Semester, dass auch unstetige Funktionen integrierbar sein können; zum Beispiel ist auch die *stückweise* Stetigkeit hinreichend für die Integrierbarkeit. In diesem Kapitel werden wir ein Kriterium angeben, das für die Integrierbarkeit sowohl hinreichend als auch notwendig ist: dass die Unstetigkeitsstellen der Funktion eine sog. *Nullmenge* bilden. Es handelt sich dabei um Teilmengen des $\mathbb{R}^n$, denen man von der Anschauung her das n-dimensionale Volumen Null zuordnen würde, beispielsweise eine Strecke im $\mathbb{R}^2$, ein Flächenstück im $\mathbb{R}^3$ usw. Um das Kriterium beweisen zu können, benötigen wir den Begriff der *Oszillation* einer Funktion, die als Maß für die Unstetigkeit angesehen werden kann. Die hier entwickelten Konzepte werden wir auch in den nächsten Kapiteln benötigen, wenn wir das Jordan-Volumen einführen und die Integrationstheorie auf Funktionen erweitern, deren Definitionsbereiche keine Quader sind.

Wichtige Begriffe und Sätze in diesem Kapitel

- Nullmengen, Jordansche Nullmengen
- Oszillation einer Funktion in einem Punkt
- Lebesguesches Integrabilitätskriterium

Zur Erinnerung: Eine Menge I wird (höchstens) ***abzählbar*** genannt, wenn sie endlich oder gleichmächtig mit der Menge $\mathbb{N}$ der natürlichen Zahlen ist. Letzteres bedeutet, dass eine bijektive Abbildung $\varphi : I \to \mathbb{N}$ existiert. Eine Familie $(A_i)_{i\in I}$ von Mengen bezeichnen wir als abzählbar, wenn die Indexmenge I abzählbar ist.

In diesem Abschnitt werden wir neben den kompakten auch offene Quader im $\mathbb{R}^n$ betrachten. Dabei handelt es sich um kartesische Produkte der Form $]a_1, b_1[\times \ldots \times]a_n, b_n[$ mit $a_k, b_k \in \mathbb{R}$, $a_k < b_k$. Das Volumen eines solchen Quaders sei weiterhin das Produkt $\prod_{k=1}^n (b_k - a_k)$ der Intervalllängen $b_k - a_k$. Ist $Q = [a_1, b_1] \times \ldots \times [a_n, b_n]$ ein kompakter Quader, so bezeichnen wir mit $Q^\circ =]a_1, b_1[\times \ldots \times]a_n, b_n[$ den offenen Quadern, der mit den entsprechenden offenen Intervallen gebildet wird. Umgekehrt können wir jedem offenen Quader Q der oben angegebenen Form durch $\bar{Q} = [a_1, b_1] \times \ldots \times [a_n, b_n]$ einen kompakten Quader zuordnen. Bei Q° handelt es sich um das Innere und bei $\bar{Q}$ um den Abschluss der Menge Q, wie es in § 7 definiert wurde.

Für die spätere Verwendung bemerken wir noch, dass zu jedem $\varepsilon \in \mathbb{R}^+$ und jedem kompakten Quader Q ein offener Quader $K \supseteq Q$ mit $v(K) < v(Q) + \varepsilon$ gebildet werden kann. Ist Q nämlich ein kartesisches Produkt der Intervalle $[a_k, b_k]$, dann ist für jedes $\delta \in \mathbb{R}^+$ der Quader

$$K_\delta \quad = \quad]a_1 - \delta, b_1 + \delta[\times \ldots \times]a_n - \delta, b_n + \delta[$$

eine Obermenge von Q. Für $\delta \to 0$ gilt offenbar $v(K_\delta) \to v(Q)$, denn das Volumen eines Quaders ist eine stetige Funktion der Kantenlängen $b_k - a_k$. Für hinreichend kleines $\delta \in \mathbb{R}^+$ kann also $v(K_\delta) < v(Q) + \varepsilon$ erreicht werden.

(14.1) Definition. Eine Teilmenge $A \subseteq \mathbb{R}^n$ heißt ***Nullmenge***, wenn es für jedes $\varepsilon \in \mathbb{R}^+$ eine abzählbare Familie $(K_i)_{i\in I}$ von offenen Quadern mit $A \subseteq \bigcup_{i\in I} K_i$ und $\sum_{i\in I} v(K_i) < \varepsilon$ existiert. Findet man für jedes $\varepsilon \in \mathbb{R}^+$ sogar eine endliche Familie von offenen Quadern mit diesen Eigenschaften, dann sprechen wir von einer ***Jordanschen Nullmenge***.

Offenbar ist jede Jordansche Nullmenge eine Nullmenge, denn jede endliche Familie von Quadern ist insbesondere auch eine abzählbare Familie.

(14.2) Proposition.

(i) Gilt $A \subseteq B \subseteq \mathbb{R}^n$ und ist B eine Nullmenge, dann ist auch A eine Nullmenge. Ist B sogar eine Jordansche Nullmenge, dann ist auch A eine Jordansche Nullmenge.

(ii) Jede abzählbare Menge ist eine Nullmenge, und jede endliche Menge ist eine Jordansche Nullmenge.

(iii) Endliche Vereinigungen von Jordanschen Nullmengen sind Jordansche Nullmengen. Abzählbare Vereinigungen von Nullmengen sind Nullmengen.

(iv) Eine kompakte Teilmenge $A \subseteq \mathbb{R}^n$ ist genau dann eine Nullmenge, wenn sie eine Jordansche Nullmenge ist.

Beweis: zu (i) Sei $\varepsilon \in \mathbb{R}^+$ vorgegeben. Weil B eine Nullmenge ist, gibt es eine abzählbare Familie $(K_i)_{i\in I}$ offener Quader mit $\sum_{i\in I} v(K_i) < \varepsilon$ und $\bigcup_{i\in I} K_i \supseteq B \supseteq A$. Daraus folgt, dass auch A eine Nullmenge ist. Für Jordansche Nullmengen läuft das Argument genauso, die abzählbare Familie muss lediglich durch eine endliche ersetzt werden.

zu (ii) Sei $X = \{x_m \mid m \in \mathbb{N}\} \subseteq \mathbb{R}^n$ abzählbar und $\varepsilon \in \mathbb{R}^+$ vorgegeben. Dann können wir für jedes $m \in \mathbb{N}$ einen Quader K_m mit $x_m \in K_m$ und $v(K_m) < 2^{-m}\varepsilon$ wählen. Es gilt dann $\bigcup_{m\in\mathbb{N}} K_m \supseteq X$ und

$$\sum_{m\in\mathbb{N}} v(K_m) \quad < \quad \varepsilon \sum_{m=1}^{\infty} 2^{-m} \quad = \quad \varepsilon.$$

Somit ist X eine Nullmenge. Sei nun X endlich, $X = \{x_1, \ldots, x_m\}$ mit $m \in \mathbb{N}$ und wieder $\varepsilon \in \mathbb{R}^+$ beliebig vorgegeben. Dann wählen wir für jedes $\ell \in \{1, \ldots, m\}$ einen Quader K_ℓ mit $x_\ell \in K_\ell$ und $v(K_\ell) < \frac{1}{m}\varepsilon$. Dann ist $\bigcup_{\ell=1}^{m} K_\ell \supseteq X$ und

$$\sum_{\ell=1}^{m} v(K_\ell) \quad < \quad \sum_{\ell=1}^{m} \tfrac{1}{m}\varepsilon \quad = \quad \varepsilon.$$

Dies zeigt, dass X eine Jordansche Nullmenge ist.

zu (iii) Wir beschränken uns auf den Beweis der Aussage für Nullmengen. Sei $(X_m)_{m\in\mathbb{N}}$ eine Familie von Nullmengen $X_m \subseteq \mathbb{R}^n$ und $X = \bigcup_{m\in\mathbb{N}} X_m$. Sei $\varepsilon \in \mathbb{R}^+$ beliebig vorgegeben. Dann gibt es für jedes $m \in \mathbb{N}$ eine abzählbare Familie $(K_{m\ell})_{\ell\in\mathbb{N}}$ von Quadern mit $\bigcup_{\ell=1}^{\infty} K_{m\ell} \supseteq X_m$ und $\sum_{\ell=1}^{\infty} v(K_{m\ell}) < 2^{-m}\varepsilon$. Da jede abzählbare Vereinigung von abzählbaren Mengen wieder abzählbar ist, handelt es sich bei $(K_{m\ell})_{m,\ell\in\mathbb{N}}$ um eine abzählbare Familie von Quadern. Außerdem gilt

$$\bigcup_{m=1}^{\infty}\left(\bigcup_{\ell=1}^{\infty} K_{m\ell}\right) \quad \supseteq \quad \bigcup_{m=1}^{\infty} X_m \quad = \quad X$$

und

$$\sum_{m=1}^{\infty}\left(\sum_{\ell=1}^{\infty} v(K_{m\ell})\right) \quad < \quad \sum_{m=1}^{\infty} 2^{-m}\varepsilon \quad = \quad \varepsilon.$$

zu (iv) Sei $A \subseteq \mathbb{R}^n$ kompakt. Die Implikation „⇒“ ist gültig, weil jede (und nicht nur jede kompakte) Jordansche Nullmenge auch eine Nullmenge ist. Zum Beweis von „⇐“ setzen wir nun voraus, dass A eine Nullmenge ist. Für vorgegebenes $\varepsilon \in \mathbb{R}^+$ finden wir eine abzählbare Familie $(K_m)_{m\in\mathbb{N}}$ offener Quader mit $\bigcup_{m\in\mathbb{N}} K_m \supseteq A$ und $\sum_{m=1}^{\infty} v(K_m) < \varepsilon$.

Weil A kompakt und die Quader K_m offen sind, können wir in $(K_m)_{m\in\mathbb{N}}$ eine endliche Teilüberdeckung $K_{\ell_1},\dots,K_{\ell_r}$ wählen. Es gilt dann $\bigcup_{i=1}^{r} K_{\ell_i} \supseteq A$ und $\sum_{i=1}^{r} \nu(K_{\ell_i}) \le \sum_{m=1}^{\infty} \nu(K_m) < \varepsilon$. Dies zeigt, dass A eine Jordansche Nullmenge ist. □

Wir bemerken noch, dass in der Definition der Nullmengen und der Jordanschen Nullmengen die offenen Quader durch kompakte Quader ersetzt werden können, ohne dass sich an der Definition etwas ändert. Ist nämlich $A \subseteq \mathbb{R}^n$, $\varepsilon \in \mathbb{R}^+$ und $(K_m)_{m\in\mathbb{N}}$ eine abzählbare Familie von offenen Quadern mit $\bigcup_{m\in\mathbb{N}} K_m \supseteq A$ und $\sum_{m=1}^{\infty} \nu(K_m) < \varepsilon$, dann ist durch $(\bar{K}_m)_{m\in\mathbb{N}}$ eine Familie abgeschlossener Quader mit $\bigcup_{m\in\mathbb{N}} \bar{K}_m \supseteq \bigcup_{m\in\mathbb{N}} K_m \supseteq A$ gegeben. Wegen $\nu(\bar{K}_m) = \nu(K_m)$ für alle $m \in \mathbb{N}$ gilt auch $\sum_{m=1}^{\infty} \nu(\bar{K}_m) < \varepsilon$.

Setzen wir nun umgekehrt voraus, dass $(K_m)_{m\in\mathbb{N}}$ eine Familie abgeschlossener Quader mit den beiden Eigenschaften $\bigcup_{m\in\mathbb{N}} K_m \supseteq A$ und $\sum_{m=1}^{\infty} \nu(K_m) < \varepsilon$ ist. Sei $\delta = \varepsilon - \sum_{m=1}^{\infty} \nu(K_m)$. Auf Grund der Bemerkung zu Beginn dieses Abschnitts finden wir für jedes $m \in \mathbb{N}$ einen offenen Quader $\tilde{K}_m \supseteq K_m$ mit $\nu(\tilde{K}_m) < \nu(K_m) + 2^{-m}\delta$. Es gilt dann $\bigcup_{m\in\mathbb{N}} \tilde{K}_m \supseteq A$ und

$$\sum_{m=1}^{\infty} \nu(\tilde{K}_m) \quad < \quad \sum_{m=1}^{\infty} \nu(K_m) + \sum_{m=1}^{\infty} 2^{-m}\delta \quad \le \quad \sum_{m=1}^{\infty} \nu(K_m) + \delta \quad = \quad \varepsilon.$$

Nicht jede Nullmenge ist eine Jordansche Nullmenge. Setzen wir beispielsweise $A = \mathbb{N} \subseteq \mathbb{R}$, dann gibt es keine endliche Familie $\{K_1,\dots,K_m\}$ von Quadern (in diesem Fall Intervalle) mit $\bigcup_{\ell=1}^{m} K_\ell \supseteq A$, erst recht keine Familie mit der Eigenschaft $\sum_{\ell=1}^{m} \nu(K_\ell) < \varepsilon$ für ein vorgegebenes $\varepsilon \in \mathbb{R}^+$. Denn jedes Intervall K_ℓ hat endliche Länge und enthält deshalb auch nur endlich viele Punkte der unendlichen Menge A. Also ist A keine Jordansche Nullmenge. Andererseits ist A abzählbar und damit nach (14.2) (ii) eine Nullmenge.

Unser Ziel in diesem Abschnitt besteht darin, mit Hilfe der Nullmengen ein notwendiges und hinreiches Kriterium für die Riemann-Integrierbarkeit einer Funktion anzugeben. Sei $f : D \to \mathbb{R}$ eine beschränkte Funktion auf einem beliebigen Definitionsbereich $D \subseteq \mathbb{R}^n$. Für jedes $\delta \in \mathbb{R}^+$ und jedes $a \in D$ definieren wir

$$f_\delta^-(a) = \inf\{f(x) \mid x \in D, \|x-a\| < \delta\} \qquad \text{und} \qquad f_\delta^+(a) = \sup\{f(x) \mid x \in D, \|x-a\| < \delta\} \quad ,$$

wobei $\|\cdot\| = \|\cdot\|_2$ die euklidische Norm auf dem $\mathbb{R}^n$ bezeichnet. Dann nennt man

$$\omega(f,a) \quad = \quad \inf\{f_\delta^+(a) - f_\delta^-(a) \mid \delta \in \mathbb{R}^+\}$$

die ***Oszillation*** von f im Punkt a. Sie ist ein Maß für die „Sprungweite“ der Funktion f in unmittelbarer Nähe des Punkte a. Wir bemerken noch, dass für $0 < \delta_1 < \delta_2$ jeweils $f_{\delta_1}^-(a) \ge f_{\delta_2}^-(a)$ und $f_{\delta_1}^+(a) \le f_{\delta_2}^+(a)$ gilt. Insbesondere ist also $f_{\delta_1}^+(a) - f_{\delta_1}^-(a) \le f_{\delta_2}^+(a) - f_{\delta_2}^-(a)$.

(14.3) Proposition. Sei $D \subseteq \mathbb{R}^n$ und $f : D \to \mathbb{R}$ eine beschränkte Funktion.

(i) In jedem Punkt $a \in D$ ist f genau dann stetig, wenn $\omega(f,a) = 0$ ist.

(ii) Für jedes $\varepsilon \in \mathbb{R}^+$ ist $D_\varepsilon = \{x \in D \mid \omega(f,x) \ge \varepsilon\}$ relativ abgeschlossen in D.

Beweis: zu (i) „⇒“ Sei $\varepsilon \in \mathbb{R}^+$ vorgegeben und nehmen wir an, dass f im Punkt a stetig ist. Dann gibt es nach dem ε-δ-Kriterum ein $\delta_1 \in \mathbb{R}^+$, so dass $f(a) - \varepsilon < f(x) < f(a) + \varepsilon$ für alle $x \in D$ mit $\|x-a\| < \delta_1$ erfüllt ist. Es folgt $f_{\delta_1}^+(a) - f_{\delta_1}^-(a) \le 2\varepsilon$ und damit erst recht $\omega(f,a) \le 2\varepsilon$. Weil ε beliebig vorgegeben war, erhalten wir $\omega(f,a) = 0$. „⇐“ Setzen wir $\omega(f,a) = 0$ voraus, dann gibt es für vorgegebenes $\varepsilon \in \mathbb{R}^+$ ein $\delta_1 \in \mathbb{R}^+$ mit $f_{\delta_1}^+(a) - f_{\delta_1}^-(a) < \varepsilon$. Für alle $x \in D$ mit $\|x-a\| < \delta_1$ gilt

$$f_{\delta_1}^-(a) \quad \le \quad f(x), f(a) \quad \le \quad f_{\delta_1}^+(a) \quad ,$$

und daraus folgt $|f(x) - f(a)| < \varepsilon$. Damit ist das ε-δ-Kriterium erfüllt, also ist f im Punkt a stetig.

zu (ii) Wir zeigen, dass $U_\varepsilon = D \setminus D_\varepsilon$ eine in D relativ offene Teilmenge ist. Sei $a \in U_\varepsilon$ vorgegeben. Dann gilt $\omega(f,a) < \varepsilon$ nach Definition von U_ε, insbesondere finden wir ein $\delta_1 \in \mathbb{R}^+$ mit $f^+_{\delta_1}(a) - f^-_{\delta_1}(a) < \varepsilon$. Sei nun z ein beliebiger Punkt in $B_{\delta_1}(a) \cap D$ und $\delta_2 \in \mathbb{R}^+$ so gewählt, dass $B_{\delta_2}(z) \subseteq B_{\delta_1}(a)$ für die offenen Bälle bezüglich $\|\cdot\|$ gilt. Wegen

$$\begin{aligned} \inf\{f(x) \mid x \in D, \|x-a\| < \delta_1\} \quad &\leq \quad \inf\{f(x) \mid x \in D, \|x-z\| < \delta_2\} \quad \leq \\ \sup\{f(x) \mid x \in D, \|x-z\| < \delta_2\} \quad &\leq \quad \sup\{f(x) \mid x \in D, \|x-a\| < \delta_1\} \end{aligned}$$

folgt dann $f^+_{\delta_2}(z) - f^-_{\delta_2}(z) \leq f^+_{\delta_1}(a) - f^-_{\delta_1}(a)$ und somit $\omega(f,z) \leq f^+_{\delta_1}(a) - f^-_{\delta_1}(a) < \varepsilon$. Dies zeigt, dass die Menge $B_{\delta_1}(a) \cap D$ in U_ε enthalten ist, und damit ist die relative Offenheit von U_ε in D bewiesen. □

(14.4) Lemma. Seien $r \in \mathbb{N}$ und $Q, K_1, \ldots, K_r \subseteq \mathbb{R}^n$ kompakte Quader, wobei $K_i \subseteq Q$ für $1 \leq i \leq r$ gilt. Dann gibt es eine Zerlegung $\mathscr{Z}$ von Q mit der Eigenschaft, dass $K_1, \ldots, K_r$ als Vereinigung von Quadern aus $\mathscr{Q}(\mathscr{Z})$ dargestellt werden können.

Beweis: Seien die Quader Q und $K_1, \ldots, K_r$ gegeben durch $Q = [a_1, b_1] \times \ldots \times [a_n, b_n]$ und $K_i = [a_{i1}, b_{i1}] \times \ldots \times [a_{in}, b_{in}]$ für $1 \leq i \leq r$. Wegen $K_i \subseteq Q$ gilt jeweils $a_k \leq a_{ik} < b_{ik} \leq b_k$, für $1 \leq k \leq n$. Wir können also durch $\mathscr{Z} = (\mathscr{Z}^{(1)}, \ldots, \mathscr{Z}^{(n)})$ mit

$$\mathscr{Z}^{(k)} \quad = \quad \{a_{1k}, b_{1k}, \ldots, a_{rk}, b_{rk}\} \setminus \{a_k, b_k\} \qquad \text{für} \quad 1 \leq k \leq n$$

eine Zerlegung von Q definieren. Für $1 \leq i \leq r$ und $1 \leq k \leq n$ gilt dann jeweils $\{a_{ik}, b_{ik}\} \subseteq \mathscr{Z}^{(k)} \cup \{a_k, b_k\}$. Die Intervallgrenzen a_{ik}, b_{ik} sind also jeweils in der Menge der Zerlegungspunkte $\mathscr{Z}^{(k)}$ unter Hinzunahme von a_k, b_k enthalten, so dass $[a_{ik}, b_{ik}]$ als Vereinigung der Teilintervalle dargestellt werden kann, die durch die Zerlegung $\mathscr{Z}^{(k)}$ gegeben sind. Somit ist K_i eine Vereinigung der in $\mathscr{Q}(\mathscr{Z})$ enthaltenen Teilquader von Q. □

(14.5) Lemma. Sei $Q \subseteq \mathbb{R}^n$ ein kompakter Quader, $\varepsilon \in \mathbb{R}^+$ und $f : Q \to \mathbb{R}$ eine Funktion mit $\omega(f,x) < \varepsilon$ für alle $x \in Q$. Dann gibt es eine Zerlegung $\mathscr{Z}$ von Q mit der Eigenschaft, dass $c^+_{K,f} - c^-_{K,f} \leq \varepsilon$ für alle $K \in \mathscr{Q}(\mathscr{Z})$ erfüllt ist.

Beweis: Auf Grund der Voraussetzung gibt es für jeden Punkt $a \in Q$ ein $\delta_a \in \mathbb{R}^+$ mit $f^+_{\delta_a}(a) - f^-_{\delta_a}(a) \leq \varepsilon$. Für jedes a sei K_a ein kompakter Quader mit $a \in K_a$ und $K_a \subseteq B_{\delta_a}(a)$, wobei $B_{\delta_a}(a)$ den offenen Ball vom Radius δ_a bezüglich der euklidischen Norm bezeichnet. Dann bilden die offenen Quader K°_a eine offene Überdeckung von Q, aus der wir auf Grund der Kompaktheit von Q eine endliche Teilüberdeckung $K^\circ_{a_1}, \ldots, K^\circ_{a_r}$ wählen können. Nach Konstruktion gilt $c^+_{K_{a_i},f} - c^-_{K_{a_i},f} \leq \varepsilon$ für $1 \leq i \leq r$. Wird nun die Zerlegung $\mathscr{Z}$ mit (14.4) so fein gewählt, dass jeder Quader K_{a_i} als Vereinigung von Quadern aus $\mathscr{Q}(\mathscr{Z})$ dargestellt werden kann, dann ist $c^+_{K,f} - c^-_{K,f} \leq \varepsilon$ für alle $K \in \mathscr{Q}(\mathscr{Z})$ erfüllt. □

(14.6) Satz. *(Lebesguesches Integrierbarkeitskriterium)*

Sei $Q \subseteq \mathbb{R}^n$ ein kompakter Quader, $f : Q \to \mathbb{R}$ eine beschränkte Funktion und $U \subseteq Q$ die Teilmenge der Punkte, in denen f unstetig ist. Dann sind äquivalent

(i) Die Funktion f ist Riemann-integrierbar.

(ii) Die Menge U ist eine Nullmenge.

Beweis: „⇒“ Nach Voraussetzung ist f Riemann-integrierbar, und wir müssen zeigen, dass U eine Nullmenge ist. Definieren wir für jedes $\varepsilon \in \mathbb{R}^+$ jeweils $U_\varepsilon = \{x \in Q \mid \omega(f,x) \geq \varepsilon\}$, dann gilt $U = \bigcup_{n\in\mathbb{N}} U_{1/n}$ nach (14.3) (i). Wegen (14.2) (iii) genügt es zu zeigen, dass $U_{1/n}$ für jedes $n \in \mathbb{N}$ eine Nullmenge ist. Seien also $n \in \mathbb{N}$ und $\varepsilon \in \mathbb{R}^+$ vorgegeben. Auf Grund der Riemann-Integrierbarkeit von f gibt es eine Zerlegung $\mathscr{Z}$ von Q mit $\mathscr{S}_f^+(\mathscr{Z}) - \mathscr{S}_f^-(\mathscr{Z}) < \frac{\varepsilon}{n}$. Dann ist durch

$$\mathscr{Q} \quad = \quad \{K \in \mathscr{Q}(\mathscr{Z}) \mid K^\circ \cap U_{1/n} \neq \emptyset\}$$

eine endliche Menge von Quadern mit $U_{1/n} \subseteq \bigcup_{K\in\mathscr{Q}} K$ definiert. Für jedes $K \in \mathscr{Q}$ gibt es wegen $K^\circ \cap U_{1/n} \neq \emptyset$ einen Punkt $a \in K^\circ$ mit $\omega(f,a) \geq \frac{1}{n}$ und ein zugehöriges $\delta \in \mathbb{R}^+$ mit $B_\delta(a) \subseteq K^\circ$. Nach Definition der Oszillation gilt

$$\sup\{f(x) \mid x \in B_\delta(a)\} - \inf\{f(x) \mid x \in B_\delta(a)\} \quad \geq \quad \tfrac{1}{n}$$

und damit insbesondere $c_{K,f}^+ - c_{K,f}^- \geq \frac{1}{n}$ für alle $K \in \mathscr{Q}$. Daraus folgt nun

$$\begin{aligned} \tfrac{1}{n}\sum_{K\in\mathscr{Q}} v(K) \quad \leq \quad \sum_{K\in\mathscr{Q}} (c_{K,f}^+ - c_{K,f}^-)v(K) \quad \leq \quad \sum_{K\in\mathscr{Q}(\mathscr{Z})} (c_{K,f}^+ - c_{K,f}^-)v(K) \\ = \quad \mathscr{S}_f^+(\mathscr{Z}) - \mathscr{S}_f^-(\mathscr{Z}) \quad < \quad \tfrac{\varepsilon}{n} \end{aligned}$$

und $\sum_{K\in\mathscr{Q}} v(K) < \varepsilon$. Damit ist gezeigt, dass es sich bei $U_{1/n}$ um eine Nullmenge handelt.

„⇐“ Nach Voraussetzung ist die Teilmenge $U \subseteq Q$ der Unstetigkeitsstellen von f eine Nullmenge. Sei $\varepsilon \in \mathbb{R}^+$ vorgegeben. Für den Nachweis der Riemann-Integrierbarkeit von f genügt es zu zeigen, dass eine Zerlegung $\mathscr{Z}$ mit $\mathscr{S}_f^+(\mathscr{Z}) - \mathscr{S}_f^-(\mathscr{Z}) < (2\|f\|_\infty + v(Q))\varepsilon$ existiert, wobei $\|f\|_\infty = \sup\{|f(x)| \mid x \in Q\}$ gesetzt wird. Wegen $U_\varepsilon \subseteq U$ ist auch U_ε eine Nullmenge (siehe (14.2) (i)). Die Menge U_ε ist nach (14.3) (ii) abgeschlossen, und als abgeschlossene Teilmenge U_ε der kompakten Menge Q ist auch U_ε kompakt. Aus (14.2) (iv) folgt, dass U_ε sogar eine Jordansche Nullmenge ist. Es gibt also offene Teilquader $K_1, \ldots, K_p$ von Q mit $U_\varepsilon \subseteq \bigcup_{i=1}^p K_i$ und $\sum_{i=1}^r v(K_i) < \varepsilon$.

Wir wählen nun mit (14.4) eine Zerlegung $\mathscr{Z}$ von Q so, dass jeder der kompakten Quader $\bar{K}_i$ als Vereingung von Quadern aus $\mathscr{Q}(\mathscr{Z})$ dargestellt werden kann. Für jedes $K \in \mathscr{Q}(\mathscr{Z})$ gilt dann entweder $K \subseteq \bar{K}_i$ für ein $i \in \{1, \ldots, p\}$, oder K und $\bar{K}_i$ haben für alle i nur Randpunkte gemeinsam, so dass $K \cap U_\varepsilon = \emptyset$ gilt. Die Quader in der ersten Kategorie fassen wir zur Menge $\mathscr{Q}_1$, die der zweiten zur Menge $\mathscr{Q}_2$ zusammen, so dass $\mathscr{Q}_1 \cup \mathscr{Q}_2 = \mathscr{Q}(\mathscr{Z})$ und $\mathscr{Q}_1 \cap \mathscr{Q}_2 = \emptyset$ gilt. Nach (14.5), angewendet auf die einzelnen Teilquader $K \in \mathscr{Q}_2$ und die Einschränkungen $f|_K$, können wir durch eine weitere Verfeinerung von $\mathscr{Z}$ errreichen, dass $c_{K,f}^+ - c_{K,f}^- \leq \varepsilon$ für alle $K \in \mathscr{Q}_2$ gilt. Weil zudem $|c_{K,f}^-|, |c_{K,f}^+| \leq \|f\|_\infty$ für alle $K \in \mathscr{Q}_1$ gilt, erhalten wir

$$\begin{aligned} \mathscr{S}_f^+(\mathscr{Z}) - \mathscr{S}_f^-(\mathscr{Z}) \quad = \quad \sum_{K\in\mathscr{Q}(\mathscr{Z})} (c_{K,f}^+ - c_{K,f}^-)v(K) \quad = \quad \sum_{K\in\mathscr{Q}_1} (c_{K,f}^+ - c_{K,f}^-)v(K) + \sum_{K\in\mathscr{Q}_2} (c_{K,f}^+ - c_{K,f}^-)v(K) \\ \leq \quad 2\|f\|_\infty \sum_{K\in\mathscr{Q}_1} v(K) + \varepsilon \sum_{K\in\mathscr{Q}_2} v(K) \quad \leq \quad 2\|f\|_\infty \varepsilon + \varepsilon v(Q) \quad = \quad (2\|f\|_\infty + v(Q))\,\varepsilon. \end{aligned}$$

□

Wir illustrieren den soeben bewiesenen Satz an einem Beispiel. Die Funktion auf $Q = [0,1]^2$ gegeben durch

$$f : Q \longrightarrow \mathbb{R} \quad , \quad (x,y) \mapsto \begin{cases} 3 & \text{falls } x+y < 1 \\ 5 & \text{falls } x+y \geq 1 \end{cases}$$

ist offenbar unstetig, aber die Menge der Unstetigkeitsstellen ist in der Menge $U = \{(x,y) \in Q \mid x+y = 1\}$ enthalten. Ist nämlich $(x,y) \in Q$ mit $x+y > 1$ vorgegeben und $((x_n,y_n))_{n\in\mathbb{N}}$ eine Folge mit $\lim_n (x_n,y_n) = (x,y)$, dann gibt es

für $\varepsilon = \frac{1}{2}(x+y-1)$ ein $N \in \mathbb{N}$ mit $\|(x_n,y_n)-(x,y)\|_\infty < \varepsilon$ für alle $n \geq N$. Wegen $|x_n - x| < \varepsilon$ und $|y_n - y| < \varepsilon$ gilt insbesondere $x_n > x-\varepsilon$, $y_n > y-\varepsilon$ und damit $x_n + y_n > x+y-2\varepsilon = x+y-(x+y-1) = 1$ für alle $n \geq N$. Daraus folgt $f(x_n,y_n) = 5 = f(x,y)$ für alle $n \geq N$ und somit $\lim_n f(x_n,y_n) = f(x,y)$. Ebenso beweist man die Stetigkeit von f in jedem Punkt $(x,y) \in Q$ mit $x+y<1$. Also kann f nur an den Punkten (x,y) mit $x+y=1$ unstetig sein.

Die Menge U (und damit die Menge der Unstetigkeitsstellen von f) ist eine Nullmenge. Definieren wir nämlich für beliebig vorgegebenes $n \in \mathbb{N}$ und $k \in \{1,\ldots,n\}$ die Quader $Q_k^{(n)} = [\frac{n-k}{n}, \frac{n+1-k}{n}] \times [\frac{k-1}{n}, \frac{k}{n}]$, dann gilt jeweils

$$U \quad \subseteq \quad \bigcup_{k=1}^{n} Q_k^{(n)}.$$

Sei nämlich $(x,y) \in U$ vorgegeben (mit $0 \leq x,y \leq 1$ und $x+y=1$) und $k = \lceil ny \rceil$. Nach Definition der oberen Gaußklammer gilt dann $k-1 < ny \leq k \Leftrightarrow \frac{k-1}{n} < y \leq \frac{k}{n}$ und

$$\tfrac{k-1}{n} < 1-x \leq \tfrac{k}{n} \quad \Leftrightarrow \quad \tfrac{k-1-n}{n} < -x \leq \tfrac{k-n}{n} \quad \Leftrightarrow \quad \tfrac{n-k}{n} \leq x < \tfrac{n+1-k}{n} \quad ,$$

also $(x,y) \in Q_k^{(n)}$. Das Volumen der Quader beträgt jeweils $\nu(Q_k^{(n)}) = \frac{1}{n^2}$, also ist

$$\sum_{k=1}^{n} \nu(Q_k^{(n)}) \quad = \quad \sum_{k=1}^{n} \tfrac{1}{n^2} \quad = \quad \tfrac{1}{n}.$$

Ist nun $\varepsilon \in \mathbb{R}^+$ vorgegeben und $n \in \mathbb{N}$ so gewählt, dass $\frac{1}{n} < \varepsilon$ erfüllt ist, dann gilt $U \subseteq \bigcup_{k=1}^n Q_k^{(n)}$ und $\sum_{k=1}^n \nu(Q_k^{(n)}) < \varepsilon$. Damit haben wir gezeigt, dass U sogar eine Jordansche Nullmenge ist.

Nach (14.6) ist die Funktion f also Riemann-integrierbar. Das Integral können wir wie zuvor mit dem Satz von Fubini ausrechen. Für jedes $x \in [0,1]$ gilt $f_x(y) = 3 \Leftrightarrow x+y<1 \Leftrightarrow y < 1-x$ und $f_x(y) = 5 \Leftrightarrow x+y \geq 1 \Leftrightarrow y \geq 1-x$, somit

$$\int_0^1 f_x(y)\,dy \quad = \quad \int_0^{1-x} 3\,dy + \int_{1-x}^1 5\,dy \quad = \quad 3(1-x)+5x \quad = \quad 2x+3$$

und

$$\int_Q f(x,y)\,d(x,y) \quad = \quad \int_0^1 \left(\int_0^1 f_x(y)\,dy \right) dx \quad = \quad \int_0^1 (2x+3)\,dx \quad = \quad \left[x^2+3x\right]_{x=0}^{x=1} \quad = \quad 4.$$

Literaturverzeichnis

[BF] M. Barner, F. Flor, *Analysis II*. de Gruyter Lehrbuch.

[Fo] O. Forster, *Analysis 2*. vieweg studium - Grundkurs Mathematik.

[He] H. Heuser, *Lehrbuch der Analysis, Teil 2*. Teubner-Verlag.

[Kö] K. Königsberger, *Analysis 2*. Springer-Verlag.