Ralf Gerkmann

Mathematisches Institut der
Ludwig-Maximilians-Universität München

Sommersemester 2025

Analysis mehrerer Variablen

Inhaltsverzeichnis

§ 1. Konvergenz in metrischen Räumen *3*
§ 2. Stetige Abbildungen zwischen metrischen Räumen *16*
§ 3. Topologische Räume *28*
§ 4. Kompaktheit und Zusammenhang *39*
§ 5. Partielle Ableitungen und Richtungsableitungen *50*
§ 6. Totale Differenzierbarkeit und mehrdim. Ableitungsregeln *58*
§ 7. Lokale Umkehrbarkeit und implizit definierte Funktionen *69*

Literaturverzeichnis *80*

§ 1. *Konvergenz in metrischen Räumen*

Inhaltsübersicht

In Kapitel § 14 haben wir zwei wichtige neue Grundbegriffe eingeführt: die Normen auf einem $\mathbb{R}$-Vektorraum und die Metriken auf einer Menge. Allerdings kennen wir bisher abgesehen von der Norm $\|v\| = \sqrt{\langle v, v\rangle}$, die durch das euklidische Standard-Skalarprodukt induziert wird, keine weiteren konkreten Beispiele. Wir beginnen das aktuelle Kapitel mit der Einführung einer Vielzahl weiterer Normen auf dem $\mathbb{R}^n$, den sog. *p-Normen*. Außerdem behandeln wir das Konzept der *Äquivalenz* von Normen, das im Hinblick auf die mehrdimensionale Analysis eine wichtige Rolle spielt, weil viele analytische Eigenschaften von Funktionen (wie Stetigkeit oder Differenzierbarkeit) unverändert bleiben, wenn man zu einer äquivalenten Norm übergeht.

Außerdem 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. Auch das Konzept der *Cauchyfolgen*, dass die Feststellung der Konvergenz ohne die Angabe eines Grenzwerts ermöglicht, lässt sich auf die metrischen Räume übertragen. Zum Schluss beweisen wir den sog. *Banachschen Fixpunktsatz*, ein wichtiges Hilfsmittel sowohl für praktische Anwendungen, vor allem numerische Verfahren, als auch für theoretische Herleitungen. Beispielsweise werden wir diesen Satz später beim Kriterium für lokale Umkehrarkeit und implizit definierte Funktionen verwenden, und noch später werden wir ihm in der Theorie der Gewöhnlichen Differenzialgleichungen wieder begegnen.

Wichtige Begriffe und Sätze

- Konvergenz und Grenzwerte in metrischen Räumen
- Cauchyfolgen und Vollständigkeit metrischer Räume
- Definition der p-Norm auf $\mathbb{R}^n$ für $p \in [1, +\infty]$.
- Definition der Äquivalenz von Normen
- Je zwei Normen auf dem $\mathbb{R}^n$ sind äquivalent.
- Banachscher Fixpunktsatz

(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

$$\|\lambda x\|_p \quad = \quad \sqrt[p]{\textstyle\sum_{i=1}^n |\lambda x_i|^p} \quad = \quad |\lambda| \sqrt[p]{\textstyle\sum_{i=1}^n |x_i|^p} \quad = \quad |\lambda| \|x\|_p. \tag{1.2}$$

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

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

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.

(1.3) 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 $\xi^{1/p} \leq \frac{1}{p}(\xi - 1) + 1$ zuvbeweisen. 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 der Linearen Algebra lässt sich nun verallgemeinern zu

(1.4) 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 Lemmma (1.3), angewendet auf die nicht-negativen Zahlen $\frac{|x_i|^p}{\|x\|_p^p}$ und $\frac{|y_i|^q}{\|y\|_q^q}$ gilt

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

Daraus folgt

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

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

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.

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

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. Die Summe in der Klammer kann nach unten abgeschätzt werden durch $\frac{|x_k|^p}{|x_k|^p} = 1$, und nach oben wegen $\frac{|x_i|^p}{|x_k|^p} \leq 1$ für $1 \leq i \leq n$ durch den Wert n. Für jedes p gilt also jeweils $|x_k| \leq \|x\|_p \leq |x_k| \cdot \sqrt[p]{n}$. Wegen

$$\lim_{p \to +\infty} |x_k| \cdot \sqrt[p]{n} \quad = \quad |x_k| \cdot \lim_{p \to +\infty} \sqrt[p]{n} \quad = \quad |x_k| \cdot \lim_{p \to +\infty} e^{\frac{\ln(n)}{p}} \quad = \quad |x_k| \cdot e^0 \quad = \quad |x_k|$$

folgt nun aus dem Sandwich-Lemma, dass auch $\|x\|_p$ für $p \to +\infty$ gegen $|x_k| = \|x\|_\infty$ konvergiert. □

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 $\|x\|_\infty \leq \|x\|_1$.

(1.6) 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

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

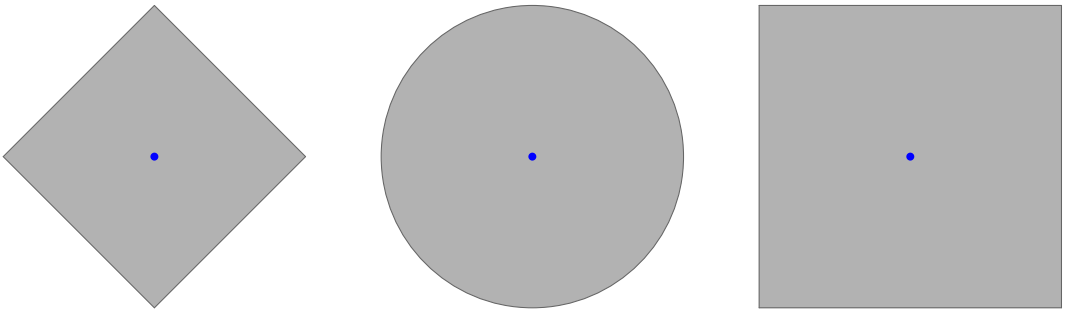
Offenbar gleichwertig mit dieser Bedingung ist die Existenz von reellen Konstanten $\delta_1, \delta_2 > 0$ mit $\|x\| \leq \delta_1\|x\|'$ und $\|x\|' \leq \delta_2\|x\|$ für alle $x \in V$.

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.



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

(1.7) 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.

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

Beweis: 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 $a = (a_1, ..., a_d) \in \mathbb{R}^d$, 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 $\phi : V \to \mathbb{R}^d$ von $\mathbb{R}$-Vektorräumen. Aus der Linearen Algebra ist bekannt, dass sich eine Norm durch einen Vektorraum-Isomorphismus von einem $\mathbb{R}$-Vektorraum auf einen anderen übertragen lässt. Demnach sind durch $\|v\|_* = \|\phi^{-1}(v)\|$ und $\|v\|'_* = \|\phi^{-1}(v)\|'$ 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 $v \in V$ gilt dann

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

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

Genau wie bei den normierten Vektorräumen definieren wir

(1.9) 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.

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.

(1.10) 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 $\delta_X(x,y) = 1$.

Dass es sich bei δ_X tatsächlich um eine Metrik handelt, kann leicht überprüft werden. 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 $B_r(x) = \bar{B}_r(x) = X$.

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 $1 \leq k \leq m$. In dieser Schreibweise gilt dann $x^{(n)} = (x_1^{(n)}, \ldots, x_m^{(n)})$.

(1.11) 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

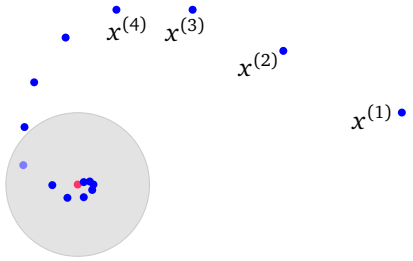
$$\lim_{n\to\infty} x^{(n)} \quad = \quad a \quad ,$$

wenn für jedes $\varepsilon \in \mathbb{R}^+$ ein $N \in \mathbb{N}$ existiert, so dass $x^{(n)} \in B_\varepsilon(a)$ für alle $n \geq N$ 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 $d(a, x) < \varepsilon$. Die Konvergenz der Folge $(x^{(n)})_{n\in\mathbb{N}}$ ist also äquivalent dazu, dass

$$\lim_{n\to\infty} d(a, x^{(n)}) \quad = \quad 0 \qquad \text{gilt.} \tag{1.12}$$

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 (1.12) einfach die Form $\lim_n |x^{(n)} - a| = 0$. Wie in der Analysis einer Variablen gilt auch hier

(1.13) 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.

(1.14) Proposition Sei X eine beliebige Menge. Eine Folge $(x^{(n)})_{n\in\mathbb{N}}$ im metrischen Raum (X, δ_X) ausgestattet mit der diskreten Metrik 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 $x^{(n)} = a$ für alle $n \geq N$. $\square$

(1.15) 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. $\square$

Nach Satz (1.8) 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 Proposition (1.15) ist diese Definition von der Wahl der Norm $\|\cdot\|$ unabhängig.

(1.16) 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 $1 \le k \le m$ 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 $n \ge N_k$ erfüllt ist. Setzen wir $N = \max\{N_1,\ldots,N_m\}$, dann gilt für alle $n \ge N$ 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 $1 \le k \le m$. Sei dazu $\varepsilon \in \mathbb{R}^+$ vorgegeben und $N \in \mathbb{N}$ so gewählt, dass $d_\infty(x^{(n)},a) < \varepsilon$ für alle $n \ge N$ gilt. Dann folgt

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

für alle $n \ge N$ und $1 \le k \le m$. 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

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

(1.17) 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 $m,n \ge N$ gilt.

Der Beweis der folgenden Aussage ist dem Beweis von Proposition (1.15) über die Konvergenz sehr ähnlich.

(1.18) 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.

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 Metrik $d(a,b) = |a-b|$. 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.

(1.19) 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.

(1.20) 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 $k \in \{1,...,d\}$ 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

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

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 Satz (1.16) 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 Proposition (1.15) und Proposition (1.18) 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, und durch $\|v\| = \|\phi^{-1}(v)\|_\infty$ ist 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.

(1.21) 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

(1.22) 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. □

(1.23) 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 betrachten wir die Funktion ϕ auf dem Intervall $X = [1, \frac{3}{2}]$. 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 $|\phi(x) - \phi(y)| = |\phi'(x_0)||x - y| \leq \frac{1}{2}|x - y|$.

Wählen wir nun $\varepsilon \in \mathbb{R}^+$ mit $\varepsilon \leq \frac{3}{2} - \sqrt{2}$, und setzen wir $X = [\sqrt{2} - \varepsilon, \sqrt{2} + \varepsilon]$, dann bildet ϕ die Menge X in sich ab und definiert auf X eine Kontraktion, mit Kontraktionskonstante $\gamma = \frac{1}{2}$. Denn wie wir oben gezeigt haben, gilt $|\phi(x) - \phi(y)| \leq \frac{1}{2}|x - y|$ für alle $x, y \in [1, \frac{3}{2}]$, also erst recht für alle $x, y \in X$. Ist nun $x \in X$ vorgegeben, dann gilt $|x - \sqrt{2}| \leq \varepsilon$ und somit $|\phi(x) - \sqrt{2}| = |\phi(x) - \phi(\sqrt{2})| \leq \frac{1}{2}|x - \sqrt{2}| \leq \frac{1}{2}\varepsilon$, also $\sqrt{2} - \frac{1}{2}\varepsilon \leq \phi(x) \leq \sqrt{2} + \frac{1}{2}\varepsilon$ und damit auch $\phi(x) \in X$. Wie wir später sehen werden, sind abgeschlossene Teilmengen vollständiger metrischer Räume selbst vollständig, und daraus wird sich ergeben, dass X ein vollständiger metrischer Raum ist. Der Banachsche Fixpunktsatz ist in dieser Situation also anwendbar.

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 Proposition (1.23) 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 $\sqrt{2} \approx 1{,}414213562373095048801$, 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 $0{,}5 \cdot 10^{-17}$.

§ 2. Stetige Abbildungen zwischen metrischen Räumen

Inhaltsübersicht

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

- stetige Abbildungen und Homöomorphismen zwischen metrischen Räumen
- Funktionsgrenzwert einer Abbildung in einem Berührpunkt des Definitionsbereichs
- Polar-, Zylinder- und Kugelkoordinaten
- Operatornorm einer stetigen linearen Abbildung

(2.1) 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.

(2.2) 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.

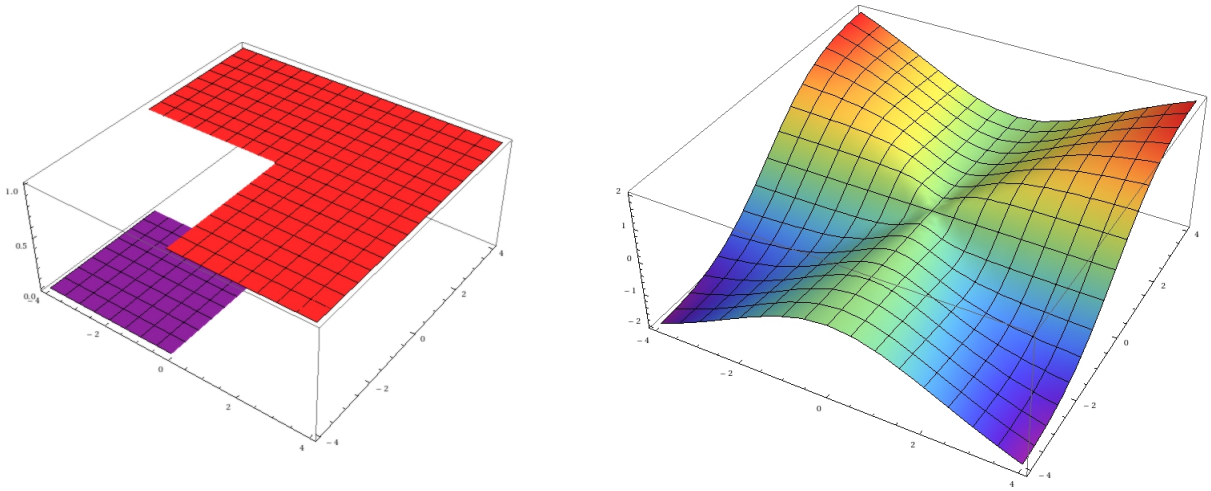
Beweis: zu (i) Sei $c \in \mathbb{R}$ und $f : X \to \mathbb{R}$ gegeben durch $f(x) = c$ für alle $x \in X$. Sei außerdem $a \in X$ ein beliebig gewählter Punkt und $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge mit $\lim_n x^{(n)} = a$. Dann gilt

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

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. Sei (X, d_X) ein metrischer Raum und $D \subseteq X$. Wir bezeichnen einen Punkt $a \in X$ als ***Berührpunkt*** von D, wenn $a \notin D$ gilt und eine Folge $(x^{(n)})_{n\in\mathbb{N}}$ in D mit $\lim_n x^{(n)} = a$ existiert.

(2.3) 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 Berührpunkt von D. Wir bezeichnen $b \in Y$ als ***Grenzwert*** von f für $x \to a$, wenn für jede Folge $(x^{(n)})$ in D mit $\lim_n x^{(n)} = a$ 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 Proposition (1.15) 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.

(2.4) 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. Existiert ein solcher Homöomorphismus, dann bezeichnet man (Y, d_Y) als ***homöomorph*** zu (X, d_X).

Wie man anhand der Definition leicht überprüft, ist durch die Eigenschaft „homöomorph“ eine Äquivalenzrelation auf der Klasse der metrischen Räume definiert: Ist (Y, d_Y) homöomorph zu (X, d_X), dann ist auch (X, d_X) homöomorph zu (Y, d_Y). Ist (Z, d_Z) ein weiterer metrischer Raum, und ist (Y, d_Y) homöomorph zu (X, d_X), und (Z, d_Z) homöomorph zu (Y, d_Y), dann ist (Z, d_Z) homöomorph zu (X, d_X). Man muss hier von der „Klasse“ der metrischen Räume sprechen, weil diese zu zahlreich sind, um eine Menge zu bilden. Dementsprechend ist die Eigenschaft „homöomorph“ keine Relation auf einer Menge, sondern eine sogenannte Klassenrelation.

Die Eigenschaft „homömorph“ spielt in der Analysis eine wichtige Rolle, weil zwischen den reellwertigen Funktionen zwischen homöomorphen metrischen Räumen eine natürliche 1-zu-1-Korrespondenz existiert: Ist $\phi : X \to Y$ ein Homöomorphismus zwischen metrischen Räumen (X, d_X) und (Y, d_Y) und $f : X \to \mathbb{R}$ eine stetige Funktion, dann ist $f \circ \phi^{-1}$ eine stetige Funktion auf Y. Ist umgekehrt g eine stetige reellwertige Funktion auf Y, dann ist $g \circ \phi$ eine stetige reellwertige Funktion auf X. Wenn man allgemeine Aussagen über stetige reellwertige auf diesen Räumen untersuchen möchte, dann genügt es, sich auf einen der beiden Räume zu konzentrieren.

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

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 Satz (1.16) gilt dann insbesondere $\lim\limits_{n\to\infty} x_i^{(n)} = a_i$, also insgesamt $\lim_n \pi_i(x^{(n)}) = \lim_n x_i^{(n)} = a_i = \pi_i(a)$. □

(2.6) 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 $n \geq N$ erfüllt ist. Es folgt dann $d_Y(f(a), f(x_n)) < \varepsilon$ für alle $n \geq N$.

„⇒“ 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 $d_Y(f(a), f(x)) \geq \varepsilon$ existiert. Insbesondere gibt es für $n \in \mathbb{N}$ ein $x^{(n)} \in X$ mit $d_X(a, x^{(n)}) < \frac{1}{n}$ und $d_Y(f(a), f(x^{(n)})) \geq \varepsilon$. 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. □

(2.7) 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 Satz (1.16) folgt daraus $\lim_n f_k(x^{(n)}) = f_k(a)$ für $1 \leq k \leq d$. 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 $1 \leq k \leq n$. Nach Satz (1.16) folgt daraus $\lim_n f(x^{(n)}) = a$. Also ist f im Punkt a stetig. □

(2.8) 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 Satz (1.16) 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. □

(2.9) 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 Proposition (2.7) 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 (2.2) (ii) stetig ist, sind also $f + g$, fg und f/g stetige Funktionen. □

Als Anwendungsbeispiel zeigen wir

(2.10) 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 (2.5) gezeigt haben, sind die Abbildungen $(x,y) \mapsto x$ und $(x,y) \mapsto y$ stetig. Ebenso ist nach Proposition (2.2) (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 Folgerung (2.9) 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 Folgerung (2.9) erhält man schließlich die Stetigkeit von f. □

(2.11) 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 $\|w\| = \|(w-v)+v\| \leq \|w-v\| + \|v\|$ und somit $\|w\| - \|v\| \leq \|v-w\|$. Da der Betrag $|\|v\| - \|w\||$ immer mit $\|v\| - \|w\|$ oder $\|w\| - \|v\|$ übereinstimmt, erhalten wir insgesamt $|\|v\| - \|w\|| \leq \|v - w\|$. □

(2.12) 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 Lemma (2.11) folgt $|\|x^{(n)}\| - \|v\|| \leq \|x^{(n)} - v\| < \varepsilon$ für alle $n \geq N$. □

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 $[a,b]$.

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

Beweis: zu (i) Die Abbildung ϕ ist nach Proposition (2.7) 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 Proposition (2.5) stetig sind, gilt dasselbe nach Folgerung (2.9) auch für $(r, \varphi) \mapsto r\cos(\varphi)$ und $(r, \varphi) \mapsto r\sin(\varphi)$. Eine erneute Anwendung von Proposition (2.7) 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

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

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 auszurechnen.

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

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

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

(2.14) 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

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

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 (2.14):

Der Beweis der Stetigkeit kann genau wie in Satz (2.13) 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 (2.13) 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 (2.13) 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.

(2.15) 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 $\|\phi(v)\| \leq \gamma\|v\|$ 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

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

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 $0 \leq \|\phi(v^{(n)}) - \phi(a)\| \leq \gamma\|v^{(n)} - a\|$ 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.

(2.16) 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 $\|\phi_1(f)\|_\infty \leq \gamma\|f\|_\infty$ 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 $V \to W$.

(2.17) 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. □

Aus der Definition der Operatornorm folgt unmittelbar $\|\phi(v)\| \leq \|\phi\|\|v\|$ für alle $\phi \in \mathscr{L}(V,W)$ und $v \in V$.

(2.18) 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

$$\|\phi_A\| \quad = \quad \max\left\{ \sum_{j=1}^{n} |a_{ij}| \;\middle|\; 1 \leq i \leq m \right\} \tag{2.19}$$

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 (2.19) 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 $\sum_{j=1}^n a_{i_0 j}v_j = \sum_{j=1}^n |a_{i_0 j}| = \|\phi_A\|$. □

(2.20) 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.

(2.21) Satz Sei X ein metrischer Raum, und seien V, W zwei endlich-dimensionale, normierte $\mathbb{R}$-Vektorräume. Eine Abbildung $\varphi : X \to \mathscr{L}(V, W)$ 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.

„⇐“ 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 $n \geq N$. □

§ 3. Topologische Räume

Inhaltsübersicht

Eine Teilmenge $U \subseteq X$ eines metrischen Raums (X, d_X) wird *offen* genannt, wenn um jeden Punkt von U ein (eventuell sehr kleiner) ε-Ball gelegt werden kann, der immer noch vollständig vollständig in U liegt. Komplemente offener Teilmengen werden als *abgeschlossen* bezeichnet. Die offenen Teilmengen eines metrischen Raums bilden ein Mengensystem, das bestimmten Gesetzmäßigkeiten unterliegt. Dies führt uns zum Begriff des *topologischen Raums*, einer Verallgemeinerung der metrischen Räume.

Die Konvergenz von Folgen und die Stetigkeit von Abbildungen lassen sich im allgemeinen Kontext topologischer Räume formulieren, und zumindest in metrischen Räumen lässt sich die Abgeschlossenheit einer Teilmenge anhand eines Folgenkriteriums überprüfen. Im Allgemeinen ist es möglich, dass eine Folge in einem topologischen Raum mehrere Grenzwerte besitzt; in den *hausdorffschen* topologischen Räumen ist der Grenzwert allerdings (im Falle der Existenz) weiterhin eindeutig.

Jeder Teilmenge A eines topologischen Rings lässt sich eine offene Menge A° und eine abgeschlossene Menge $\bar{A}$ mit $A^\circ \subseteq A \subseteq \bar{A}$ zuordnen, genannt das *Innere* und der *Abschluss* von A. Die Menge $\partial A = \bar{A} \setminus A^\circ$ wird der *Rand* von A genannt. Der Menge A kann selbst auf natürliche Weise die Struktur eines topologischen Raums gegeben werden; man bezeichnet diese als *induzierte Topologie*.

Wichtige Begriffe und Sätze

- (hausdorffscher) topologischer Raum
- offene und abgeschlossene Teilmengen in metrischen und topologischen Räumen
- Umgebung eines Punktes
- Konvergenz von Folgen in topologischen Räumen
- Stetigkeit einer Abbildung zwischen topologischen Räumen
- Inneres A° und Abschluss $\bar{A}$ einer Teilmenge A eines topologischen Raums, Rand von A
- induzierte Topologie auf einer Teilmenge eines topologischen Raums

(3.1) Definition Sei (X, d) ein metrischer Raum.

(i) Eine Teilmenge $U \subseteq X$ wird ***offen*** genannt, wenn für jedes $x \in U$ ein $\varepsilon \in \mathbb{R}^+$ existiert, so dass $B_\varepsilon(x) \subseteq U$ gilt.

(ii) Man bezeichnet eine Teilmenge $A \subseteq X$ als ***abgeschlossen***, wenn das Komplement $X \setminus A$ von A in X offen ist.

Wir betrachten eine Reihe von Beispielen. Dabei legen wir auf der Menge $\mathbb{R}$ stets die Standardmetrik d_1 gegeben durch $d_1(x,y) = |x-y|$ zu Grunde.

(i) Offene Intervalle der Form $]a,b[$ mit $-\infty \leq a < b \leq +\infty$ sind offene Teilmengen im metrischen Raums $(\mathbb{R}, d_1)$. Sei nämlich $x \in]a,b[$ und $\varepsilon \in \mathbb{R}^+$ so gewählt, dass im Fall $a \in \mathbb{R}$ die Ungleichung $\varepsilon \leq x-a$ und im Fall $b \in \mathbb{R}$ auch die Ungleichung $\varepsilon \leq b-x$ erfüllt ist. Wir überprüfen, dass $B_\varepsilon(x) \subseteq]a,b[$ gilt. Sei dazu $y \in B_\varepsilon(x)$ vorgegeben; dann gilt $|y-x| < \varepsilon$, also $x-\varepsilon < y < x+\varepsilon$. Ist $a \in \mathbb{R}$, dann gilt $y > x-\varepsilon \geq x-(x-a) = a$. Ist $b \in \mathbb{R}$, dann gilt $y < x+\varepsilon \leq x+(b-x) = b$. Insgesamt ist damit $y \in]a,b[$ nachgewiesen. (Man beachte, dass im Fall $a = -\infty$ keine untere und im Fall $b = +\infty$ keine obere Abschätzung für y nachgewiesen werden muss.)

(ii) Abgeschlossene Intervalle der Form $[a,b]$ mit $-\infty < a < b < +\infty$ sind abgeschlossene Teilmengen in $(\mathbb{R}, d_1)$. Dazu müssen wir zeigen, dass $U = X \setminus [a,b]$ in $(\mathbb{R}, d_1)$ offen ist. Sei $x \in U$ vorgegeben; wir überprüfen, dass ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq U$ existiert. Wegen $x \in U$ gilt $x < a$ oder $x > b$. Ist $x < a$, dann setzen wir $\varepsilon = a-x$. Es gilt dann $B_\varepsilon(x) \subseteq U$. Ist nämlich $y \in B_\varepsilon(x)$, dann gilt $|y-x| < \varepsilon$ und insbesondere $y < x+\varepsilon = x+(a-x) = a$. Es folgt $y \notin [a,b]$, also $y \in U$. Betrachten wir nun den Fall $x > b$. Dann ist $B_\varepsilon(x) \subseteq U$ für $\varepsilon = x-b$ erfüllt. Denn ist $y \in B_\varepsilon(x)$, dann folgt aus $|y-x| < \varepsilon$ insbesondere $y > x-\varepsilon = x-(x-b) = b$, also $y \notin [a,b]$ und damit $y \in U$.

(iii) Halboffene Intervalle der Form $[a,b[$ mit $-\infty < a < b < +\infty$ sind weder offen noch abgeschlossen in $(\mathbb{R}, d_1)$. Wäre die Menge offen, dann müsste es ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(a) \subseteq [a,b[$ geben. Ist aber y ein beliebiger Punkt in $]a-\varepsilon, a[$, ist dieser wegen $|y-a| < \varepsilon$ zwar in $B_\varepsilon(a)$ enthalten, wegen $y < a$ aber nicht in $[a,b[$. Nehmen wir nun an, die Menge wäre abgeschlossen in $(\mathbb{R}, d_1)$. Dann wäre $V = \mathbb{R} \setminus [a,b[$ in $(\mathbb{R}, d_1)$ offen. Insbesondere gäbe es dann wegen $b \in V$ ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(b) \subseteq V$. Ist aber $y \in \mathbb{R}$ mit $\max\{a, b-\varepsilon\} < y < b$, dann gilt zwar $y \in B_\varepsilon(b)$, aber wegen $y \in [a,b[$ andererseits $y \notin V$. Ebenso ist jedes halboffene Intervall der Form $]a,b]$ mit $-\infty < a < b < +\infty$ weder offen noch abgeschlossen.

(iv) Sei (X,d) ein metrischer Raum, $x \in X$ und $r \in \mathbb{R}^+$. Dann ist der offene Ball $B_r(x)$ eine offene und der abgeschlossene Ball $\bar{B}_r(x)$ eine abgeschlossenen Teilmenge in (X,d). Zum Nachweis der Offenheit von $B_r(x)$ sei $y \in B_r(x)$ und $\varepsilon = r - d(x,y)$. Wegen $d(x,y) < r$ ist ε eine positive Zahl. Es gilt dann $B_\varepsilon(y) \subseteq B_r(x)$. Denn für jedes $z \in B_\varepsilon(y)$ gilt $d(y,z) < \varepsilon$ und damit $d(x,z) \leq d(x,y) + d(y,z) < d(x,y) + r - d(x,y) = r$, also $z \in B_r(x)$.

Um zu zeigen, dass $\bar{B}_r(x)$ abgeschlossen ist, müssen wir zeigen, dass es sich bei $U = X \setminus \bar{B}_r(x)$ um eine offene Teilmenge in (X,d) handelt. Ist $y \in U$ vorgegeben, dann gilt $d(x,y) > r$, und somit ist $\varepsilon = d(x,y) - r \in \mathbb{R}^+$. Wir zeigen, dass $B_\varepsilon(y) \subseteq U$ gilt. Nehmen wir an, dass dies nicht der Fall ist, und somit ein Punkt $z \in B_\varepsilon(y)$ mit $z \notin U$ existiert. Dann folgt $z \in \bar{B}_r(x)$. Es gilt dann sowohl $d(y,z) < \varepsilon$ als auch $d(x,z) \leq r$. Es folgt dann $d(x,y) \leq d(x,z) + d(z,y) = d(x,z) + d(y,z) < r + \varepsilon = r + (d(x,y) - r) = d(x,y)$. Der Widerspruch $d(x,y) < d(x,y)$ zeigt, dass $B_\varepsilon(y) \subseteq U$ gelten muss.

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

(3.2) 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. Weil nach Satz (1.8) alle Normen auf V äquivalent sind, spielt die Wahl der Norm wegen Proposition (3.2) in diesem Fall keine Rolle.

Sei X eine Menge und $\mathfrak{P}(X)$ seine Potenzmenge, also die Menge aller Teilmengen von X. Im weitere Verlauf 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 Abbildungs- 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 $]7, 8[$.

(3.3) Definition Sei X eine Menge und $\mathcal{T} \subseteq \mathfrak{P}(X)$. Man bezeichnet $\mathcal{T}$ als ***Topologie*** auf X, wenn folgende Bedingungen erfüllt sind.

(i) Es gilt $\emptyset \in \mathcal{T}$ und $X \in \mathcal{T}$.

(ii) Für alle $U, V \in \mathcal{T}$ gilt $U \cap V \in \mathcal{T}$.

(iii) Ist $(U_i)_{i\in I}$ eine Familie von Teilmengen von X, wobei $U_i \in \mathcal{T}$ für jedes $i \in I$ gilt, dann ist auch die Vereinigung $\bigcup_{i\in I} U_i$ ein Element von $\mathcal{T}$.

Ein Paar $(X, \mathcal{T})$ bestehend aus einer Menge X und einer Topologie $\mathcal{T}$ auf X bezeichnet man als ***topologischen Raum***. Die Elemente von $\mathcal{T}$ werden auch als ***offene Teilmengen*** des topologischen Raums bezeichnet.

(3.4) Satz Sei (X, d) ein metrischer Raum und $\mathcal{T}$ die Menge der in (X, d) offenen Teilmengen. Dann ist $\mathcal{T}$ eine Topologie auf X.

Beweis: Wir überprüfen, dass das Mengensystem $\mathscr{T}$ die Bedingungen (i), (ii) und (iii) in der Definition einer Topologie auf X erfüllt.

zu (i) Die Menge $Y = \emptyset$ 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) Seien $U, V \in \mathscr{T}$. Zu zeigen ist, dass $U \cap V$ in $\mathscr{T}$ liegt, also offen im metrischen Raum (X, d) ist. 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 $(U_i)_{i \in I}$ eine Familie von Mengen $U_i \in \mathscr{T}$, also eine Familie von in (X, d) offenen Teilmengen. Zu zeigen ist, dass $U = \bigcup_{i \in I} U_i$ in (X, d) offen ist. Sei dazu $x \in U$ 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 $B_\varepsilon(x) \subseteq U$. □

Eine Teilmenge $A \subseteq X$ in einem topologischen Raum $(X, \mathscr{T})$ wird ***abgeschlossen*** genannt, wenn die Menge $X \setminus A$ offen ist, also $X \setminus A \in \mathscr{T}$ gilt. Die Gleichungen $\emptyset = X \setminus X$ und $X = X \setminus \emptyset$ zeigen, dass $\emptyset$ und X nicht nur offene, sondern auch abgeschlossene Teilmengen in jedem topologischen Raum sind. Die Vereinigung zweier abgeschlossener Mengen A, B ist abgeschlossen, denn es gilt $X \setminus (A \cup B) = (X \setminus A) \cap (X \setminus B)$, und der Durchschnitt der beiden offenen Mengen $X \setminus A$, $X \setminus B$ ist offen. Der Durchschnitt $\bigcap_{i \in I} A_i$ einer Familie $(A_i)_{i \in I}$ abgeschlossener Teilmengen $A_i \subseteq X$ ist abgeschlossen, denn es gilt $X \setminus \left(\bigcap_{i \in I} A_i\right) = \bigcup_{i \in I} (X \setminus A_i)$, und die Vereinigung auf der rechten Seite der Gleichung ist offen.

Allerdings sind unendliche Vereinigungen abgeschlossener Teilmengen eines topologischen Raums im Allgemeinen nicht abgeschlossen. Beispielsweise ist für jedes $n \in \mathbb{N}$ die Teilmenge $[\frac{1}{n}, 1]$ in $(\mathbb{R}, d_1)$ abgeschlossen, die Vereinigung $\bigcup_{n \in \mathbb{N}} [\frac{1}{n}, 1] =]0, 1]$ aber nicht. Ebenso sind unendliche Durchschnitte offener Teilmengen im Allgemeinen nicht offen. Zum Beispiel ist $]-\frac{1}{n}, \frac{1}{n}[$ offen in $(\mathbb{R}, d_1)$ für jedes $n \in \mathbb{N}$, aber $\bigcap_{n \in \mathbb{N}}]-\frac{1}{n}, \frac{1}{n}[= \{0\}$ ist keine offene Teilmenge.

(3.5) Definition Sei $(X, \mathscr{T})$ ein topologischer Raum und $x \in X$. Eine Teilmenge $U \subseteq X$ wird ***Umgebung*** von x genannt, wenn eine offene Teilmenge V von X mit $x \in V$ und $V \subseteq U$ existiert.

Ist die Topologie $\mathscr{T}$ auf X durch eine Metrik gegeben, so ist $U \subseteq X$ genau dann eine Umgebung des Punktes $x \in X$, wenn ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq U$ existiert. Die Implikation „⇐“ ist unmittelbar klar, denn $B_\varepsilon(x)$ ist eine offene Teilmenge von X mit $x \in B_\varepsilon(x)$. Zum Nachweis der Richtung „⇒“ setzen wir voraus, dass V eine offene Teilmenge von X mit $x \in V$ und $V \subseteq U$ ist. Auf Grund der Offenheit von V existiert ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(x) \subseteq V$. Dann gilt insbesondere $B_\varepsilon(x) \subseteq U$.

(3.6) Definition Ein topologischer Raum $(X, \mathscr{T})$ wird ***hausdorffsch*** genannt, wenn für zwei beliebige verschiedene Punkte $x, y \in X$ jeweils Umgebungen U von x und V von y mit $U \cap V = \emptyset$ existieren.

Ist der topologische Raum $(X,\mathscr{T})$ durch eine Metrik d definiert, dann ist er hausdorffsch. Seien nämlich x und y zwei verschiedene Punkte von X. Dann ist $\varepsilon = \frac{1}{2}d(x,y) \in \mathbb{R}^+$, und $U = B_\varepsilon(x)$ und $V = B_\varepsilon(y)$ sind Umgebungen von x bzw. y mit $U \cap V = \varnothing$. Wäre nämlich $z \in U \cap V$, dann müsste $d(x,z) < \varepsilon$ und $d(y,z) < \varepsilon$ gelten, und aus der Dreiecksungleichung würde sich der Widerspruch $d(x,y) \leq d(x,z) + d(z,y) < \varepsilon + \varepsilon = 2\varepsilon = d(x,y)$ ergeben. Es gibt aber topologische Räume, die nicht hausdorffsch sind, was zeigt, dass nicht jede Topologie durch eine Metrik definiert ist.

(i) Auf jeder Menge X existiert die ***triviale Topologie*** (auch indiskrete Topologie, chaotische Topologie oder Klumpentopologie genannt), gegeben durch $\mathscr{T} = \{\varnothing, X\}$. Sobald X mindestens zwei verschiedene Punkte x, y enthält, ist $(X,\mathscr{T})$ ein nicht-hausdorffscher topologischer Raum. Denn die einzige Umgebung von x ist die Menge X, und dasselbe gilt für y. Aber der Durchschnitt $X \cap X = X$ ist nicht die leere Menge.

(ii) Als weiteres konkretes Beispiel für einen nicht-hausdorffschen topologischen Raum betrachten wir $(X,\mathscr{T})$ gegeben durch $X = [1,2] \cup [3,4]$ und $\mathscr{T} = \{\varnothing, [1,2], [3,4], X\}$. Man kann leicht kontrollieren, dass durch $\mathscr{T}$ tatsächlich eine Topologie auf X definiert ist. Aber die Topologie ist nicht hausdorffsch. Denn die einzigen Umgebungen der Punkte 1 und 2 sind $[1,2]$ und X, und für alle $U, V \in \{[1,2], X\}$ ist der Durchschnitt $U \cap V$ jeweils nicht leer.

Mit Hilfe des Umgebungsbegriffs lässt sich die Konvergenz von Folgen definieren.

(3.7) Definition Sei $(X,\mathscr{T})$ ein topologischer Raum, $a \in X$ und $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge in X. Wir sagen, die Folge $(x^{(n)})_{n\in\mathbb{N}}$ ***konvergiert*** gegen den Punkt a und bezeichnen a als ***Grenzwert*** der Folge, wenn für jede Umgebung U von x in $(X,\mathscr{T})$ ein $N \in \mathbb{N}$ existiert, so dass $x^{(n)} \in U$ für alle $n \geq N$ erfüllt ist.

Ist a der einzige Grenzwert von $(x^{(n)})_{n\in\mathbb{N}}$, dann verwendet man auch in diesem Kontext die Notation $\lim_n x^{(n)} = a$. Allerdings kann es in einem topologischen Raum $(X,\mathscr{T})$ vorkommen, dass eine Folge mehrere Grenzwerte besitzt. Als Beispiel betrachten wir die konstante Folge gegeben durch $x^{(n)} = 1$ im topologischen Raum $(X,\mathscr{T})$ aus Beispiel (ii) von oben. Jeder Punkt $a \in [1,2]$ ist ein Grenzwert dieser Folge. Denn jede Umgebung U von a enthält das Intervall $[1,2]$, und folglich gilt $x^{(n)} \in U$ für alle $n \in \mathbb{N}$. Es gilt aber

(3.8) Proposition Ist $(X,\mathscr{T})$ ein hausdorffscher topologischer Raum, dann besitzt jede Folge in X höchstens einen Grenzwert.

Beweis: Nehmen wir an, dass $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge in X mit zwei verschiedenen Grenzwerten $a, b \in X$ ist. Da $(X,\mathscr{T})$ hausdorffsch ist, gibt es Umgebungen U von a und V von b mit $U \cap V = \varnothing$. Da a ein Grenzwert der Folge ist, existiert ein $N_1 \in \mathbb{N}$ mit $x^{(n)} \in U$ für alle $n \geq N_1$. Ebenso existiert ein $N_2 \in \mathbb{N}$ mit $x^{(n)} \in V$ für alle $n \geq N_2$. Für $n = \max\{N_1, N_2\}$ muss dann $x^{(n)} \in U \cap V$ gelten. Aber das ist wegen $U \cap V = \varnothing$ unmöglich. □

Mit Hilfe von Folgengrenzwerten lässt sich ein notwendiges Kriterium für die Abgeschlossenheit einer Teilmenge formulieren.

(3.9) Satz Sei $(X, \mathcal{T})$ ein topologischer Raum und $A \subseteq X$.

(i) Ist A abgeschlossen und $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge in A, die in X einen Grenzwert x besitzt, dann gilt $x \in A$.

(ii) Ist die Topologie $\mathcal{T}$ durch eine Metrik d definiert, dann gilt auch die Umkehrung: Liegt der Grenzwert jeder in A liegenden, in X konvergenten Folge in A, dann ist A eine abgeschlossene Teilmenge von X.

Beweis: zu (i) Nehmen wir an, dass der Grenzwert x der Folge $(x^{(n)})_{n\in\mathbb{N}}$ nicht in A liegt. Wegen der Abgeschlossenheit von A ist die Teilmenge $U = X \setminus A$ offen und wegen $x \in U$ eine Umgebung von x. Auf Grund der Konvergenz existiert ein $N \in \mathbb{N}$ mit $x^{(n)} \in U$ für alle $n \geq N$. Aber dies widerspricht unserer Voraussetzung, dass $x^{(n)} \in A$ für alle $n \in \mathbb{N}$ gilt.

„⇐“ Nehmen wir an, dass für jede in A liegende auch der Grenzwert in A liegt, dass aber A nicht abgeschlossen ist. Dann ist die Menge $U = X \setminus A$ nicht offen; existiert also ein Punkt $x \in U$ mit der Eigenschaft, dass $B_\varepsilon(x) \subseteq U$ für kein $\varepsilon \in \mathbb{R}^+$ erfüllt ist. Dies wiederum bedeutet, dass für jedes $\varepsilon \in \mathbb{R}^+$ der Durchschnitt $B_\varepsilon(x) \cap A$ nicht leer 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 $x \notin A$. 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.

(3.10) Definition Sei $(X, \mathcal{T})$ ein topologischer Raum und $A \subseteq X$ eine Teilmenge.

(i) Man nennt x einen ***inneren Punkt*** von A, wenn x eine Umgebung U mit $U \subseteq A$ besitzt. 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*** $\bar{A}$ von A, wenn für jede Umgebung U von x der Durchschnitt $A \cap U$ 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 d_1.

(3.11) Proposition Sei $(X, \mathcal{T})$ ein topologischer 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 jede Umgebung U von x 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 U eine Umgebung von x mit $U \subseteq A$, dann gilt $x \in U$ und somit $x \in A$. Die Menge A° ist auch offen. Für jeden Punkt $x \in A^\circ$ gibt es nämlich eine Umgebung U_x von x mit $U_x \subseteq A$ und (nach Definition des Umgebungsbegriffs) eine offene Menge V_x mit $x \in V_x$ und $V_x \subseteq U_x \subseteq A$. Die Menge V_x ist nicht nur Umgebung von x, sondern auch Umgebung jedes ihrer Punkte. Es gilt also $V_x \subseteq A^\circ$ und somit auch $\bigcup_{x \in A^\circ} V_x \subseteq A^\circ$. Aus $x \in V_x$ für alle $x \in A^\circ$ folgt andererseits $A^\circ \subseteq \bigcup_{x \in A^\circ} V_x$. Die Gleichung $A^\circ = \bigcup_{x \in A^\circ} V_x$ zeigt, dass A° (als Vereinigung offener Teilmengen) selbst offen ist.

Sei nun U eine beliebige offene Teilmenge von X mit $U \subseteq A$, und sei $x \in U$. Weil U eine Umgebung von x für jedes $x \in U$ ist, besteht U vollständig aus inneren Punkten, es gilt also $U \subseteq A^\circ$. Dies zeigt, dass A° die größte offene Teilmenge von X ist, die in A enthalten ist.

zu (ii) Ist $x \in A$, dann gilt $x \in U \cap A$ für jede Umgebung U von x und somit $U \cap A \neq \emptyset$. Dies zeigt, dass x in $\bar{A}$ liegt. Damit ist $A \subseteq \bar{A}$ nachgewiesen. Nun zeigen wir, dass $\bar{A}$ abgeschlossen ist, indem wir nachweisen, dass $U = X \setminus \bar{A}$ eine offene Teilmenge ist. Ist $x \in U$ vorgegeben, dann existiert (wegen $x \notin \bar{A}$) eine Umgebung V_x von x mit $V_x \cap A = \emptyset$, und somit auch eine offene Menge W_x mit $x \in W_x$ und $W_x \cap A = \emptyset$, also $W_x \subseteq U$. Es gilt also $\bigcup_{x \in U} W_x \subseteq U$, andererseits aber auch $U \subseteq \bigcup_{x \in U} W_x$ wegen $x \in W_x$ für alle $x \in U$. Die Gleichung $U = \bigcup_{x \in U} W_x$ zeigt, dass U tatsächlich offen ist.

Nun zeigen wir noch, dass $\bar{A}$ die kleinste abgeschlossene Teilmenge von X ist, die A enthält. Sei $Z \subseteq X$ abgeschlossen mit $Z \supseteq A$, und nehmen wir an, dass $\bar{A}$ keine Teilmenge von Z ist. Dann gibt es einen Punkt $x \in \bar{A} \setminus Z$. Setzen wir $U = X \setminus Z$, dann ist U offen, und es gilt $x \in U$. Außerdem folgt aus $A \subseteq Z$, dass $U \cap A = \emptyset$ gilt. Somit ist U eine Umgebung von x, die mit A leeren Durchschnitt besitzt; aber dies steht zur Annahme $x \in \bar{A}$ im Widerspruch.

zu (iii) „⇒“ Sei $x \in \partial A$. Dann gilt insbesondere $x \in \bar{A}$, somit enthält jede Umgebung U von x einen Punkt aus A. Nehmen wir nun an, es gibt eine Umgebung U, die keinen Punkt des Komplements $X \setminus A$ enthält. Dann folgt daraus $U \subseteq A$. Aber dies würde bedeuten, dass x ein innerer Punkt von A ist, was der Annahme $x \in \partial A$ widerspricht.

„⇐“ Weil nach Voraussetzung jede Umgebung U von x einen Punkt aus A enthält, gilt $x \in \bar{A}$. Nehmen wir nun $x \in A^\circ$. Dann gibt es eine Umgebung U von x mit $U \subseteq A$. Aber dies bedeutet, dass U mit $X \setminus A$ leeren Durchschnitt hat, im Widerspruch zur Voraussetzung. Insgesamt ist x also in $\partial A = \bar{A} \setminus A^\circ$ enthalten. □

Kommen wir nun zum Stetigkeitsbegriff im Zusammenhang mit topologischen Räumen.

(3.12) Definition Seien $(X, \mathscr{T})$ und $(Y, \mathscr{U})$ topologische Räume, $f : X \to Y$ eine Abbildung, und sei $a \in X$. Wir bezeichnen f als ***stetig im Punkt*** a, wenn für jede Umgebung V von $f(a)$ in $(Y, \mathscr{U})$ eine Umgebung U von a in $(X, \mathscr{T})$ mit $f(U) \subseteq V$ existiert. Die Funktion f wird ingesamt als ***stetig*** bezeichnet, wenn sie in jedem Punkt von X stetig ist.

Diese Definition ist das natürliche Analogon zum ε-δ-Kriterium in $\mathbb{R}$ oder (allgemeiner) in metrischen Räumen. Es sei daran erinnert, dass die intuitive und ungenauer Formulierung dieses Kriteriums lautet: „Wenn ein Punkt x nahe bei a liegt, dann liegt $f(x)$ nahe bei $f(a)$." Die „Nähe" kommt bei der neuen Formulierung durch den Umgebungsbegriff zum Ausdruck: Dass x nahe bei a liegen soll, wird durch die Forderung $x \in U$ konkretisiert, und dass sich $f(x)$ nahe bei $f(a)$ befindet, wird durch $f(x) \in V$ ausgedrückt. Die Gültigkeit der Implikation $x \in U \Rightarrow f(x) \in V$ wiederum ist äquivalent zu $f(U) \subseteq V$.

Sind die Topologien $\mathscr{T}$ auf X und $\mathscr{Y}$ auf Y durch Metriken d_X und d_Y gegeben, dann ist die neue Formulierung tatsächlich auch äquivalent zum Stetigkeitsbegriff aus §9: Setzen wir zunächst voraus, dass $f : X \to Y$ stetig im Punkt a als Abbildung zwischen den metrischen Räumen (X, d_X) und (Y, d_Y) ist. Dies ist gleichbedeutend mit der Gültigkeit des ε-δ-Kriteriums an der Stelle a. Sei nun V eine Umgebung von $f(a)$ in $(Y, \mathscr{U})$. Dann existiert eine offene Teilmenge V' von Y mit $f(a) \in V'$ und $V' \subseteq V$. Weil die Topologie $\mathscr{U}$ durch die Metrik d_Y definiert ist, bedeutet dies wiederum, dass ein $\varepsilon \in \mathbb{R}^+$ existiert, so dass der offene Ball $B_{Y,\varepsilon}(f(a))$ von Radius ε um $f(a)$ bezüglich d_Y in V', und somit auch in V, enthalten ist. Durch Anwendung des ε-δ-Kriteriums erhalten wir nun ein $\delta \in \mathbb{R}^+$ mit der Eigenschaft, dass für alle $x \in X$ die Implikation $d_X(a, x) < \delta \Rightarrow d_Y(f(a), f(x))$ erfüllt ist. Setzen wir $U = B_{X,\delta}(a)$, dann ist U eine Umgebung von a in $(X, \mathscr{T})$, denn nach Definition der Topologie $\mathscr{T}$ ist U eine offene Teilmenge von X, die a enthält. Die Implikation ist gleichbedeutend mit $f(B_{X,\delta}(a)) \subseteq B_{Y,\varepsilon}(f(a))$, und wir erhalten insgesamt $f(U) \subseteq B_{Y,\varepsilon}(f(a)) \subseteq V' \subseteq V$. Dies zeigt, dass f stetig als Abbildung zwischen den topologischen Räumen $(X, \mathscr{T})$ und $(Y, \mathscr{U})$ im Sinne von Definition (3.12) ist.

Setzen wir nun umgekehrt voraus, dass f in diesem Sinne stetig im Punkt a ist. Wir zeigen, dass f als Abbildung zwischen den metrischen Räumen (X, d_X) und (Y, d_Y) bezüglich a das ε-δ-Kriterium erfüllt und somit als Abbildung zwischen den metrischen Räumen stetig ist. Sei dazu $\varepsilon \in \mathbb{R}^+$ vorgegeben. Dann ist $V = B_{Y,\varepsilon}(f(a))$ eine (offene) Umgebung des Punktes $f(a)$ im topologischen Raum $(Y, \mathscr{U})$. Auf Grund der Stetigkeit existiert eine Umgebung U von a in $(X, \mathscr{T})$ mit $f(U) \subseteq V$. Da es sich bei U um eine Umgebung handelt, existiert eine in $(X, \mathscr{T})$ offene Teilmenge U' mit $a \in U'$ und $U' \subseteq U$. Dies wiederum bedeutet nach Definition der Topologie $\mathscr{T}$, dass ein $\delta \in \mathbb{R}^+$ mit $B_{X,\delta}(a) \subseteq U'$ existiert. Es gilt dann $f(B_{X,\delta}(a)) \subseteq f(U') \subseteq f(U) \subseteq V = B_{Y,\varepsilon}(f(a))$. Für alle $x \in X$ gilt also die Implikationskette

$$d(a, x) < \delta \quad \Rightarrow \quad x \in B_{X,\delta}(a) \quad \Rightarrow \quad f(x) \in B_{Y,\varepsilon}(f(a)) \quad \Rightarrow \quad d(f(a), f(x)) < \varepsilon \quad ,$$

mit anderen Worten, das ε-δ-Kriterium ist erfüllt.

(3.13) Proposition Seien $(X,\mathscr{T})$ und $(Y,\mathscr{U})$ topologische Räume. Für eine Abbildung $f : X \to Y$ sind die folgenden Aussagen äquivalent.

(i) Die Abbildung f ist stetig.

(ii) Für jede offene Teilmenge V in $(Y,\mathscr{U})$ ist $f^{-1}(V)$ offen.

(iii) Für jede abgeschlossene Teilmenge A in $(Y,\mathscr{U})$ ist $f^{-1}(A)$ abgeschlossen.

Beweis: Die Äquivalenz „(ii) $\Leftrightarrow$ (iii)“ folgt direkt aus der Tatsache, dass für jede Teilmenge $B \subseteq Y$ jeweils $f^{-1}(Y \setminus B) = X \setminus f^{-1}(B)$ gilt. Tatsächlich gilt für jedes $x \in X$ die Äquivalenz

$$x \in f^{-1}(Y \setminus B) \quad \Leftrightarrow \quad f(x) \in Y \setminus B \quad \Leftrightarrow \quad f(x) \notin B \quad \Leftrightarrow \quad x \notin f^{-1}(B) \quad \Leftrightarrow \quad x \in X \setminus f^{-1}(B).$$

Zum Beweis von „(ii) $\Rightarrow$ (iii)“ setzen wir nun voraus, dass $f^{-1}(V)$ für jede offene Teilmenge von Y offen ist. Ist A eine abgeschlossene Teilmenge von Y, dann ist $Y \setminus A$ offen. Auf Grund der Voraussetzung ist dann auch $f^{-1}(Y \setminus A) = X \setminus f^{-1}(A)$ offen, und dies wiederum ist gleichbedeutend mit der Abgeschlossenheit von $f^{-1}(A)$. Der Beweis der Implikation „(iii) $\Rightarrow$ (ii)“ läuft vollkommen analog.

Beweisen wir nun die Implikation „(i) $\Rightarrow$ (ii)“. Auf Grund der Voraussetzung ist $f : X \to Y$ in jedem Punkt $a \in X$ stetig. Sei nun $V \subseteq Y$ eine offene Teilmenge; wir müssen zeigen, dass $f^{-1}(V)$ in $(X,\mathscr{T})$ offen ist. Für jeden Punkt $a \in f^{-1}(V)$ ist V eine Umgebung von $f(a)$. Es existiert deshalb eine Umgebung U_a von a in $(X,\mathscr{T})$ mit $f(U_a) \subseteq V$. Nach Definition des Umgebungsbegriffs wiederum gibt es eine offene Teilmenge U'_a mit $U'_a \subseteq U_a$ und $a \in U'_a$. Aus $f(U'_a) \subseteq f(U_a) \subseteq V$ folgt $U'_a \subseteq f^{-1}(V)$. Insgesamt erhalten wir damit $\bigcup_{a \in f^{-1}(V)} U'_a \subseteq f^{-1}(V)$, und wegen $a \in f^{-1}(U'_a)$ für alle $a \in f^{-1}(V)$ gilt andererseits auch $f^{-1}(V) \subseteq \bigcup_{a \in f^{-1}(V)} U'_a$. Die Gleichung $f^{-1}(V) = \bigcup_{a \in f^{-1}(V)} U'_a$ zeigt, dass $f^{-1}(V)$ offen ist, als Vereinigung der offenen Teilmengen U'_a mit $a \in f^{-1}(V)$.

Nun beweisen wir noch die Implikation „(ii) $\Rightarrow$ (i)“. Unter Verwendung der Voraussetzung (ii) müssen wir zeigen, dass f in jedem Punkt von X stetig ist. Sei also $a \in X$ vorgegeben, und sei V eine Umgebung von $f(a)$ in $(Y,\mathscr{U})$. Dann gibt es eine offene Teilmenge V' von Y mit $f(a) \in V'$ und $V' \subseteq V$. auf Grund unserer Voraussetzung ist $f^{-1}(V')$ offen. Wegen $f(a) \in V'$ gilt außerdem $a \in f^{-1}(V')$, also ist $U = f^{-1}(V')$ eine Umgebung von a. Die Inklusion $f(U) \subseteq V' \subseteq V$ zeigt, dass f tatsächlich in a stetig ist. □

Ist $(X,\mathscr{T})$ ein topologischer Raum, dann existiert auf natürliche Weise eine Topologie auf jeder Teilmenge von X.

(3.14) Satz Sei $(X,\mathscr{T})$ ein topologischer Raum und $A \subseteq X$ eine Teilmenge. Es sei $\mathscr{T}_A$ die Menge aller Teilmengen der Form $U \cap A$ mit $U \in \mathscr{T}$. Dann ist $(A,\mathscr{T}_A)$ ein topologischer Raum. Man nennt $\mathscr{T}_A$ die auf A ***induzierte*** Topologie.

Beweis: Wir überprüfen die einzelnen Punkte aus Definition (3.3). Wegen $\varnothing \cap A = \varnothing$ und $X \cap A = A$ sind $\varnothing$ und A in $\mathscr{T}_A$ enthalten. Seien nun $U_A, V_A \in \mathscr{T}_A$ vorgegeben. Dann gibt es $U, V \in \mathscr{T}$ mit $U_A = U \cap A$ und $V_A = V \cap A$. Wegen $U_A \cap V_A = (U \cap A) \cap (V \cap A) = (U \cap V) \cap A$ und $U \cap V \in \mathscr{T}$ ist auch $U_A \cap V_A$ in $\mathscr{T}_A$ enthalten.

Sei nun $(U_{A,i})_{i\in I}$ eine Familie in $\mathscr{T}_A$. Dann gibt es für jedes $i \in I$ ein $U_i \in \mathscr{T}$ mit $U_{A,i} = U_i \cap A$, und $\bigcup_{i\in I} U_i$ ist in $\mathscr{T}$ enthalten. Die Gleichung $\bigcup_{i\in I} U_{A,i} = \bigcup_{i\in I}(U_i \cap A) = \left(\bigcup_{i\in I} U_i\right) \cap A$ zeigt nun, dass auch die Vereinigung $\bigcup_{i\in I} U_{A,i}$ in $\mathscr{T}_A$ enthalten ist. □

Ist die Topologie auf X durch eine Metrik definiert, dann lässt sich die auf $A \subseteq X$ induzierte Topologie ebenfalls durch eine Metrik beschreiben.

(3.15) Satz Sei (X, d) ein metrischer Raum und $A \subseteq X$. Sei $\mathscr{T}$ die durch d definierte Topologie, und sei $\mathscr{T}_A$ die auf A induzierte Topologie. Dann ist $d_A = d|_{A\times A}$ eine Metrik auf A, und $\mathscr{T}_A$ ist genau die durch d_A definierte Topologie.

Beweis: Weil d eine Metrik auf X ist, gelten für alle $x, y, z \in X$ die Aussagen $x = y \Leftrightarrow d(x, y) = 0$, $d(x, y) = d(y, x)$ und $d(x, z) \leq d(x, y) + d(y, z)$. Insbesondere gelten diese Aussagen für alle $x, y, z \in A$. Dies zeigt, dass $d_A = d|_{A\times A}$ eine Metrik auf A ist.

Sei nun $U \subseteq X$ eine beliebige Teilmenge. Wir müssen zeigen, dass genau dann $U \in \mathscr{T}_A$ gilt, wenn U im metrischen Raum (A, d_A) eine offene Teilmenge ist. Setzen wir $U \in \mathscr{T}_A$ voraus, und sei $a \in A$ beliebig vorgegeben. Wir müssen zeigen, dass ein $\varepsilon \in \mathbb{R}^+$ mit $B_{A,\varepsilon}(a) \subseteq U$ existiert, wobei $B_{A,\varepsilon}(a)$ den offenen Ball vom Radius ε um a im metrischen Raum (A, d_A) bezeichnet. Nach Definition von $\mathscr{T}_A$ existiert ein Element $U' \in \mathscr{T}$ mit $U = U' \cap A$. Weil U in (X, d) offen ist, existiert ein $\varepsilon \in \mathbb{R}^+$ mit $B_\varepsilon(a) \subseteq U'$, wobei $B_\varepsilon(a)$ den offenen Ball vom Radius ε um a in (X, d) bezeichnet. Für alle $x \in X$ gilt nun die Implikation

$$\begin{aligned} x \in B_{A,\varepsilon}(a) \quad &\Rightarrow \quad x \in A \wedge d_A(a, x) < \varepsilon \quad \Rightarrow \quad x \in A \wedge d(a, x) < \varepsilon \quad \Rightarrow \quad x \in A \wedge x \in B_\varepsilon(a) \\ &\Rightarrow \quad x \in A \wedge x \in U' \quad \Rightarrow \quad x \in U' \cap A \quad \Rightarrow \quad x \in U. \end{aligned}$$

Dies zeigt, dass tatsächlich $B_{A,\varepsilon}(a) \subseteq U$ erfüllt ist.

Setzen wir nun umgekehrt voraus, dass U offen in (A, d_A) ist. Für jeden Punkt $a \in U$ gibt es ein $\varepsilon_a \in \mathbb{R}^+$ mit $B_{A,\varepsilon_a}(a) \subseteq U$. Die Menge U stimmt dann mit der Vereinigung $\bigcup_{a\in U} B_{A,\varepsilon_a}(a)$ überein. Außerdem ist $B_{A,\varepsilon_a}(a) = B_{\varepsilon_a}(a) \cap A$ für jedes $a \in U$, denn für alle $x \in X$ gilt die Äquivalenz

$$\begin{aligned} x \in B_{\varepsilon_a}(a) \cap A \quad &\Leftrightarrow \quad x \in B_{\varepsilon_a}(a) \wedge x \in A \quad \Leftrightarrow \quad d(a, x) < \varepsilon_a \wedge x \in A \\ &\Leftrightarrow \quad d_A(a, x) < \varepsilon_a \wedge x \in A \quad \Leftrightarrow \quad x \in B_{A,\varepsilon_a}(a). \end{aligned}$$

Als Vereinigung offener Mengen ist $U' = \bigcup_{a\in U} B_{\varepsilon_a}(a)$ offen in X. Die Gleichung

$$U \quad = \quad \bigcup_{a\in U} B_{A,\varepsilon_a}(a) \quad = \quad \bigcup_{a\in U}(B_{\varepsilon_a}(a) \cap A) \quad = \quad \left(\bigcup_{a\in U}(B_{\varepsilon_a}(a)\right) \cap A \quad = \quad U' \cap A$$

zeigt nun, dass U somit in $\mathscr{T}_A$ enthalten ist. □

Als konkretes Beispiel betrachten wir $(\mathbb{R}^2, d_\infty)$ mit der durch die Norm $\|\cdot\|_\infty$ induzierten Metrik d_∞ und die Teilmenge $A = \mathbb{R} \times \{0\}$. Nach Satz (3.15) erhalten wir eine Metrik d_A auf A durch

$$\begin{aligned} d_A((x,0),(y,0)) &= d_\infty((x,0),(y,0)) = \|(x,0)-(y,0)\|_\infty = \|(x-y,0\|_\infty \\ &= \max\{|x-y|,0\} = |x-y|. \end{aligned}$$

Identifizieren wir A mit $\mathbb{R}$ über die Bijektion $\mathbb{R} \to A$, $x \mapsto (x,0)$, dann entspricht d_A der Standardmetrik d_1 auf $\mathbb{R}$. Genauso, wie wir am Anfang des Kapitels gezeigt haben, dass offene Intervalle in $\mathbb{R}$ offene Teilmengen des metrischen Raums $(\mathbb{R}, d_1)$ sind, kann man hier überprüfen, dass zum Beispiel $U_A =]0,1[\times \{0\}$ eine offene Teilmenge im metrischen Raum (A, d_A) ist. Auf Grund des Satzes kann dies auch alternativ daraus gefolgert werden, dass $U =]0,1[\times \mathbb{R}$ eine offene Teilmenge von $(\mathbb{R}^2, d_\infty)$ ist (was natürlich gezeigt werden muss) und $U_A = U \cap A$ gilt.

Andererseits ist U_A natürlich keine offene Teilmenge von $(\mathbb{R}^2, d_\infty)$. Wäre dies der Fall, dann müsste es zum Beispiel ein $\varepsilon \in \mathbb{R}^+$ geben, so dass der offene Ball $B_\varepsilon((\frac{1}{2},0))$ vom Radius ε um $(\frac{1}{2},0)$ bezüglich d_∞ in U_A enthalten ist. Aber wegen $d_\infty((\frac{1}{2},0),(\frac{1}{2},\frac{1}{2}\varepsilon)) = \max\{|\frac{1}{2}-\frac{1}{2}|, |0-\frac{1}{2}\varepsilon|\} = \max\{0, \frac{1}{2}\varepsilon\} = \frac{1}{2}\varepsilon < \varepsilon$ liegt der Punkt $(\frac{1}{2},\frac{1}{2}\varepsilon)$ in $B_\varepsilon((\frac{1}{2},0))$, ohne ein Element von U_A zu sein.

§ 4. Kompaktheit und Zusammenhang

Inhaltsübersicht

Die *Kompaktheit* eines abstrakten topologischen Raums wird durch eine Überdeckungseigenschaft definiert; im $\mathbb{R}^n$ sind die kompakten Teilmengen genau die beschränkten und abgeschlossenen Teilmengen. Die kompakten topologischen Räume teilen als Definitionsbereiche reellwertiger Funktionen viele Eigenschaften mit den endlichen abgeschlossenen Intervallen. So gilt für sie das Maximumsprinzip und für kompakte metrische Räume auch der Satz von Bolzano-Weierstraß.

Ein topologischer Raum A wird als *zusammenhängend* bezeichnet, wenn er nicht als Vereinigung echter relativ offener Teilmengen dargestellt werden kann (die man als „Komponenten“ eines nicht zusammenhängenden topologischen Raums betrachten kann). Ein stärkerer Begriff ist der *Wegzusammenhang* eines topologischen Raums A, der besagt, dass man je zwei Punkte durch eine ganz in A verlaufende Kurve verbinden kann. Die Bedeutung der zusammenhängenden Räume für die Analysis rührt daher, dass dies genau die Definitionsbereiche stetiger Funktionen sind, für die der Zwischenwertsatz seine Gültigkeit behält.

Wichtige Begriffe und Sätze

- offene Überdeckung einer Teilmenge eines topologischen Raums
- kompakte Teilmenge eines topologischen Raums, kompakter topologischer Raum
- Durchmesser und Beschränktheit einer Teilmenge eines metrischen Raums
- Schachtelungsprinzip für metrische Räume
- zusammenhängender / wegzusammenhängender topologischer Raum (Beispiele: Intervalle, konvexe Teilmengen von $\mathbb{R}$-Vektorräumen)
- Kompaktheit abgeschlossener Quader und Satz von Heine-Borel
- Satz von Bolzano-Weierstrass für metrische Räume
- gleichmäßige Stetigkeit stetiger Funktionen auf kompakten metrischen Räumen
- Zwischenwertsatz für zusammenhängende topologische Räume

(4.1) Definition Sei I eine Menge, $(X, \mathcal{T})$ ein topologischer 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 $\bigcup_{i\in I} U_i \supseteq A$.

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.

(4.2) Definition Sei $(X, \mathscr{T})$ ein topologischer 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.

Ist die Teilmenge $A = X$ von $(X, \mathscr{T})$ kompakt, so bezeichnet man $(X, \mathscr{T})$ als ***kompakten topologischen Raum***. Häufig beschreibt man die Kompaktheit einer Menge A mit der Formulierung „Jede offene Überdeckung enthält eine endliche Teilüberdeckung.“. Die bedeutet aber *nicht* einfach, dass A eine endliche offene Überdeckung besitzt. Letzteres trifft für jede Teilmenge A eines topologischen Raums $(X, \mathscr{T})$ zu: Die einelementige Familie $(U_i)_{i\in\{1\}}$ gegeben durch $U_1 = X$ ist offensichtlich eine offene 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.

(4.3) Proposition Sei $(X, \mathscr{T})$ ein topologischer 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. □

(4.4) Proposition Die Menge $\mathbb{R}$ im metrischen Raum $(\mathbb{R}, d_1)$ mit der Standardmetrik $d_1(x, y) = |x - y|$ 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 $\mathbb{R}$. □

(4.5) 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 Proposition (3.2) unmittelbar, dass auch diese Eigenschaft nicht von der Wahl der Norm abhängig ist.

(4.6) 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 $d(x,y) \leq d(a,x) + d(a,y) < 2\gamma$ für alle $x, y \in A$. Die Umkehrung ist offensichtlich.

(4.7) 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 $d(a,a') \leq d(A_n)$ 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 $n \geq N$ erfüllt ist. Sind nun $m, n \in \mathbb{N}$ mit $n \geq m \geq N$, dann folgt $d(x^{(n)}, x^{(m)}) \leq d(A_m) < \varepsilon$, 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 $n \geq m$. Weil A_m abgeschlossen ist, liegt nach Satz (3.9) 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. □

(4.8) 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 $d(Q) = \max\{b_k - a_k \mid 1 \leq k \leq n\}$.

(4.9) 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 Satz (4.7), dem Schachtelungsprinzip, 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 $2^{-m} d(Q) < \varepsilon$. Wir zeigen, dass Q_m in U_{i_0} liegt. Sei $x \in Q_m$ vorgegeben. Aus $a, x \in Q_m$ und $d(Q_m) < 2^{-m} d(Q) < \varepsilon$ 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. □

(4.10) Satz Sei A eine abgeschlossene Teilmenge eines kompakten topologischen Raums $(X, \mathcal{T})$. 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

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

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. □

(4.11) 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 Satz (4.9) eine kompakte Teilmenge von $\mathbb{R}^n$. Als abgeschlossene Teilmenge von $\bar{B}_r(0_{\mathbb{R}^n})$ ist A nach Satz (4.10) ebenfalls kompakt. □

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

(4.12) Definition Ein Punkt x in einem topologischen Raum $(X,\mathcal{T})$ 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 in $\mathbb{R}$ gilt nun

(4.13) 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. Nach Definition der Umgebungen gibt es jeweils eine offene Teilmenge V_x von X mit $V_x \subseteq U_x$ und $x \in V_x$. Die Familie $(V_x)_{x\in X}$ bildet eine offene Überdeckung von A. Auf Grund der Kompaktheit existiert eine endliche Teilmenge $J \subseteq X$, so dass A bereits von $(V_x)_{x\in J}$ überdeckt wird. Aber dies würde bedeuten, dass

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

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. □

(4.14) Folgerung Jede beschränkte Folge im $\mathbb{R}^n$ 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. □

(4.15) 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 Satz (3.9) 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 in einem metrischen Raum nach Proposition (1.13) nicht gegen zwei verschiedene Punkte konvergieren kann, erhalten wir auch hier einen Widerspruch. □

Zusammen mit Folgerung (4.11) erhalten wir

(4.16) 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

(4.17) Proposition Sei X eine unendliche Menge und δ_X die diskrete Metrik 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 § 10 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. □

(4.18) Satz Sei $f : X \to Y$ eine stetige Abbildung zwischen topologischen 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 nach Proposition (3.13) offen, 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 $y = f(x) \in V_i$. □

(4.19) Satz *(Maximumsprinzip)*

Jede stetige, reellwertige Funktion $f : X \to \mathbb{R}$ auf einem kompakten topologischen Raum $(X, \mathcal{T})$ mit $X \neq \emptyset$ ist beschränkt und nimmt auf X ihr Maximum und Minimum an.

Beweis: Nach Satz (4.18) und dem Satz von Heine-Borel, angewendet auf den $\mathbb{R}^1$, 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 Satz (3.9) auch der Grenzwert m_+ in $f(X)$ liegen. Genauso beweist man $m_- \in f(X)$. □

Den Begriff der gleichmäßigen Stetigkeit wurde im ersten Semester bereits für eindimensionale Funktionen definiert. Wir verallgemeinern ihn nun auf metrische Räume.

(4.20) 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.

Im ersten Semester wurde gezeigt, dass jede stetige Funktion auf einem kompakten Intervall auch gleichmäßig stetig ist (womit dann später die Riemann-Integrierbarkeit stetiger Funktionen nachgewiesen wurde). Das Analogon dieser Aussage für metrische Räume lautet

(4.21) 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 (Satz (2.6)) 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

$$d_Y(f(x), f(y)) \quad \leq \quad d_Y(f(a), f(x)) + d_Y(f(a), f(y)) \quad < \quad \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon \quad = \quad \varepsilon.$$ □

Kommen wir nun zum zweiten großen Thema dieses Kapitels, dem Zusammenhang. Ist $(X, \mathcal{T})$ ein topologischer Raum und $A \subseteq X$, so bezeichnen wir eine Teilmenge $V \subseteq A$ als in A ***relativ offen*** bzw. ***relativ abgeschlossen***, wenn V bezüglich der auf A induzierten Topologie offen bzw. abgeschlossen ist.

(4.22) Definition Sei $(X, \mathcal{T})$ ein topologischer 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 topologischen Raum $(X, \mathcal{T})$ selbst als *zusammenhängend*, wenn die Teilmenge $A = X$ zusammenhängend ist.

(4.23) Proposition Eine Teilmenge A eines topologischen Raums $(X, \mathcal{T})$ ist genau dann zusammenhängend, wenn $\emptyset$ 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 $U \neq \emptyset, A$, 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 $V \neq \emptyset$. 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 $U \neq A$, denn ansonsten wäre $V = \emptyset$. Also ist U eine in A relativ offene Menge ungleich $\emptyset$ und A. □

Die Teilmenge $A = \{(x, \frac{1}{x}) \mid 0 \neq x \in \mathbb{R}\} \subseteq \mathbb{R}^2\}$ des $\mathbb{R}^2$ 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.

(4.24) 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. □

(4.25) Proposition Seien $(X, \mathcal{T}_X)$ und $(Y, \mathcal{T}_Y)$ topologische 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 $U, V \subseteq f(A)$ mit $f(A) = U \cup V$, 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 Proposition (3.13). 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. □

(4.26) Satz *(Zwischenwertsatz)*

Sei $(X, \mathcal{T})$ ein topologischer 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 Proposition (4.25) 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 Satz (4.24) ein Intervall. Dies bedeutet, dass mit $f(a)$ und $f(b)$ auch jeder Wert dazwischen in $f(A)$ enthalten ist. □

(4.27) Definition Sei $(X, \mathcal{T})$ ein topologischer 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.

(4.28) 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.

(4.29) 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.

(4.30) Proposition Jeder wegzusammenhängende Teilmenge $A \subseteq X$ eines topologischen Raums $(X, \mathcal{T})$ 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 Proposition (3.13) offen in $[0,1]$. Weil aber $[0,1]$ als metrischer Raum nach Satz (4.24) 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

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

begegnet, als Beispiel für eine (im Punkt 0) unstetige Funktion, bei der sich die Unstetigkeit aber am Funktionsgraphen nicht direkt erkennen 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 Proposition (4.30) 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 Satz (3.9) 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 Proposition (2.7) 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

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

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.

§ 5. *Partielle Ableitungen und Richtungsableitungen*

Inhaltsübersicht

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

- Richtungsableitung $\partial_v f(p)$ einer Funktion f in Richtung v einem Punkt p in Richtung
- partielle Ableitung $\partial_j f$ einer Funktion
- Mittelwertsatz für Richtungsableitungen
- höhere partielle 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 $\phi(t) = a + tv$. Da es sich bei U um eine Umgebung von a handelt, und auf Grund der Stetigkeit von ϕ, ist $\phi^{-1}(U)$ eine Umgeung von 0 in $\mathbb{R}$. 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.

(5.1) Definition Sei $U \subseteq V$ offen und $f : U \to \mathbb{R}$ eine Funktion. Sei $a \in U$, $v \in V$, und sei $\phi : \mathbb{R} \to V$ gegeben durch $\phi(t) = a + tv$. 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. Ist $V = \mathbb{R}^n$ und $v = e_j$ für ein $j \in \{1,\dots,n\}$, dann spricht man auch von der j-ten ***partiellen Ableitung*** und verwendet die Bezeichung $\partial_j f(a)$.

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 Berechnung von Richtungsableitungen und partiellen Ableitungen wird durch folgende Rechenregel vereinfacht.

(5.2) Lemma Sei $U \subseteq V$ offen und $f : U \to \mathbb{R}$ eine Funktion. Seien $a, v \in V$, und sei $\phi : \mathbb{R} \to V$ gegeben durch $\phi(t) = a + tv$. 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. □

(5.3) Folgerung Sei $U \subseteq V$ offen, $f : U \to \mathbb{R}$ eine Funktion und $a \in U$. Dann gilt

$$\partial_j f(a) = \frac{\partial}{\partial t}\big(f(a_1, ..., a_{j-1}, t, a_{j+1}, ..., a_n)\big)\Big|_{t=a_j}.$$

Dies soll bedeuten, dass genau dann die j-te partielle Ableitung von f an der Stelle a definiert ist, wenn die Funktion $t \mapsto f(a_1, ..., a_{j-1}, t, a_{j+1}, ..., a_n)$ an der Stelle a_j differenzierbar ist, und dass in diesem Fall die Ableitung dieser Funktion an der Stelle a_j mit $\partial_j f(a)$ übereinstimmt.

Beweis: Dies folgt unmittelbar durch Anwendung von Lemma (5.2) auf $V = \mathbb{R}^n$, $v = e_j$ und die Stelle $t = a_j$. □

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

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

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 $\partial_2 f(x, y) = 3\sin(2x)e^{3y}$.

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

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

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

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

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

(5.4) 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

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

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

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

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.

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

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

Dann gilt $\partial_1 f(0,0) = 2$, $\partial_2 f(0,0) = 3$, aber für $v\notin \mathrm{lin}(e_1)\cup\mathrm{lin}(e_2)$ 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

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

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

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

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

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

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.

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

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

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 $]a,b[= [a,b] \setminus \{a,b\}$.

(5.5) 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 Lemma (5.2) 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.

(5.6) 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 Satz (5.5) 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 Satz (5.5), 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. □

(5.7) 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 (5.6) 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 $\partial_w \partial_v f(a) = \lim_n \partial_w \partial_v f(q_n) = \lim_n \partial_v \partial_w f(p_n) = \partial_v \partial_w f(a)$. □

§ 6. *Totale Differenzierbarkeit und mehrdim. Ableitungsregeln*

Inhaltsübersicht

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 $df(a)$ 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 $\partial_v f(a) = df(a)(v)$ für alle $v \in \mathbb{R}^n$. Wie in der eindimensionalen Analysis existieren auch für die mehrdimensionale totale Ableitung Summen-, Produkt- und Kettenregel.

Wichtige Begriffe und Sätze

- totale Differenzierbarkeit einer Funktion in einem Punkt
- totale Ableitung $df(a)$ einer Funktion f in einem Punkt a
- Zusammenhang $df(a) = \partial_v f(a)$ zwischen totaler Ableitung und Richtungsableitungen
- stetige partielle Differenzierbarkeit $\Rightarrow$ totale Differenzierbarkeit
- mehrdimensionale Summen- und Produktregel
- mehrdimensionale Kettenregel
- Jacobimatrix (synomym: Funktionalmatrix)
  in einem Punkt

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.

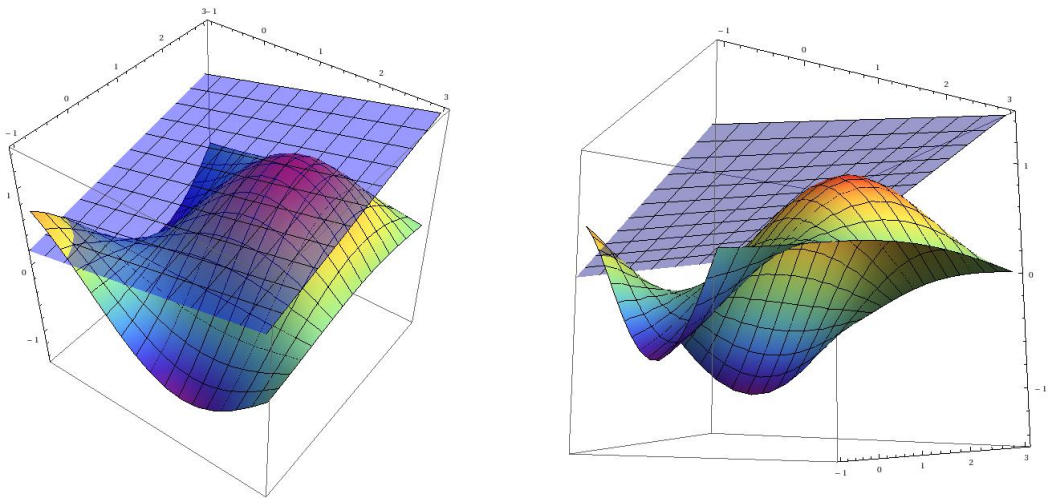
Im gesamten Abschnitt seien V, W jeweils endlich-dimensionale, normierte $\mathbb{R}$-Vektorräume.

(6.1) 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 $df(a)$.

Im Gegensatz zum eindimensionalen Fall ist die Ableitung also keine reelle Zahl mehr, sondern eine lineare Abbildung, genauer ein Element des $\mathbb{R}$-Vektorraums $\mathrm{Hom}_{\mathbb{R}}(V, W)$ 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 $\tilde{f}(a+h) = f(p) + df(a)(h)$ dar, die der Ableitung der Funktion in einem Punkt $a \in \mathbb{R}^2$ entspricht.

(6.2) 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 $f(a+h) = f(a) + df(a)(h) + \psi(h)$ eine Darstellung von f wie in der Definition der Differenzierbarkeit angegeben. Als lineare Abbildung auf einem endlich-dimensionalen $\mathbb{R}$-Vektorraum ist $df(a)$ 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

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

(6.3) 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 $df(a)$ sind dann die Ableitungen $df_1(a), ..., df_m(a)$.

Beweis: Sei $\phi \in \mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R}^m)$ eine lineare Abbildung, und seien $\phi_i \in \mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R})$ 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 Satz (1.16) 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 $\lim\limits_{h\to 0_V} \psi(h)/\|h\| = 0_{\mathbb{R}^m}$ äquivalent zu $\lim\limits_{h\to 0_V} \psi_i(h)/\|h\| = 0$ für $1 \leq i \leq m$.

„⇒“ Sei f in a differenzierbar. Dann setzen wir $\phi = df(a)$ 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, $\phi_i = df_i(a)$ für $1 \leq i \leq m$, und sei $\phi \in \mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R}^m)$ die eindeutig bestimmte lineare Abbildung mit den Komponentenfunktionen $\phi_i \in \mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R})$. 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 $1 \leq i \leq m$. 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 $\phi = df(a)$. □

(6.4) 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 $\partial_v f(a) = df(a)(v)$.

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

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

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 $(f \circ \phi)'(0) = df(a)(v)$. Für alle $t \in \mathbb{R}$ mit $|t| < \varepsilon$ gilt

$$(f \circ \phi)(t) \quad = \quad f(a + tv) \quad = \quad f(a) + df(a)(tv) + \psi(tv) \quad = \quad f(a) + t\,df(a)(v) + \psi(tv).$$

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

$$h(t) \quad = \quad \frac{(f \circ \phi)(t) - (f \circ \phi)(0)}{t} \quad = \quad \frac{f(a+tv) - f(a)}{t} \quad = \quad df(a)(v) + \frac{\psi(tv)}{t}$$

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 $(f \circ \phi)'(0) = 0 = df(a)(0)$. 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 $\lim\limits_{h\to 0} \psi(h)/\|h\| = 0$ 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 $(f\circ\phi)'(0) = \lim\limits_{n\to\infty} h(t_n) = df(a)(v) + \lim\limits_{n\to\infty} \psi(t_n v)/t_n = df(a)(v)$. □

(6.5) 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 Proposition (6.4) ist sichergestellt, dass die drei angegebenen Richtungsableitungen existieren. Weil $df(a) : V \to \mathbb{R}$ eine lineare Abbildung ist, gilt außerdem

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

(6.6) 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

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

(6.7) Proposition Seien die Bezeichnungen wie in Definition (6.6) gewählt, und sei $J = \mathrm{Jac}(f)(a)$. Dann gilt $df(a)(v) = J \cdot v$ 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 $df(a) : \mathbb{R}^n \to \mathbb{R}^m$ bezüglich der Einheitsbasen.

Beweis: Für $1 \leq i \leq m$ seien $df(a)_i$ und $(\phi_J)_i$ jeweils die i-te Komponentenfunktion linearen Abbildungen $df(a)$ und $\phi_J : \mathbb{R}^n \to \mathbb{R}^m$, $v \mapsto J \cdot v$. Nach Proposition (6.3) gilt jeweils $df(a)_i = df_i(a)$. Es genügt also, $(\phi_J)_i = df_i(a)$ für $1 \leq i \leq m$ zu beweisen. Auf Grund der Linearität der beiden Abbildungen genügt es wiederum, die Gleichungen $df_i(a)(e_j) = (\phi_J)_i(e_j)$ für $1 \leq i \leq m$ und $1 \leq j \leq n$ 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 $1 \leq j \leq n$. Aus Proposition (6.4) folgt andererseits auch $df_i(a)(e_j) = \partial_{e_j} f_i(a) = \partial_j f_i(a)$. □

Um die Notation möglichst einfach zu halten, schreiben wir in Zukunft für die Jacobi-Matrix ebenfalls $df(a)$ 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 $df(a)(v) = df(a) \cdot v$ wobei auf der linken Seite der Gleichung $df(a)$ als lineare Abbildung $\mathbb{R}^n \to \mathbb{R}^m$ interpretiert wird, während auf der rechten Seite das Matrix-Vektor-Produkt von $df(a) \in \mathscr{M}_{m\times n,\mathbb{R}}$ 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 $df(x, y) = (2x \ 2y)$ gelten. Ist f insbesondere an der Stelle $(3, 4)$ total differenzierbar, dann gilt $df(3, 4) = (6 \ \ 8)$, 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 $\lim\limits_{h\to 0} \psi(h)/\|h\|_\infty = \lim\limits_{h\to 0} \|h\|_1 = 0$. 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.

(6.8) 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 $v \in B_\varepsilon(0_{\mathbb{R}^n})$, $v = (v_1, ..., v_n)$ ordnen wir durch

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

insgesamt $n+1$ Punkte $z^{(k)}(v) \in B_\varepsilon(a)$ zu, wobei $e_1, ..., e_n$ 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 $1 \le k \le n$ jeweils in $B_\varepsilon(a)$. Nach dem Mittelwertsatz für Richtungsableitungen, Satz (5.5), gibt es für $1 \le k \le n$ im Fall $v_k \neq 0$ jeweils ein $y^{(k)}(v) \in [z^{(k-1)}(v), z^{(k)}(v)]$ mit

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

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

$$f(a+v) \quad = \quad f(z^{(n)}(v)) \quad = \quad f(z^{(0)}(v)) + \sum_{k=1}^{n} \big(f(z^{(k)}(v)) - f(z^{(k-1)}(v)) \big) \quad =$$

$$f(a) + \sum_{k=1}^{n} v_k \partial_k f(y^{(k)}(v)) \quad = \quad f(a) + \sum_{k=1}^{n} v_k \partial_k f(a) + \psi(v) \quad = \quad f(a) + \phi(v) + \psi(v).$$

Betrachten wir nun den Grenzübergang $\lim\limits_{v \to 0} \psi(v)/\|v\|_\infty$. 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 $df(a) = \phi$. □

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 § 12 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 Folgerung (6.5) erfüllt.

(iii) Die Funktion f aus Beispiel 5 in § 12 ist im Punkt $(0, 0)$ nicht differenzierbar, weil sie dort nicht einmal stetig ist, vgl. Proposition (6.2).

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. Bevor wir diese beweisen, bemerken wir zur Vorbereitung, dass der Fehlerterm ψ in der Definition der Differenzierbarkeit stets in der Form $\psi(h) = \|h\|\psi_1(h)$ schreiben lässt, mit einer Funktion $\psi_1 : U_a \to W$, die $\psi(0_V) = 0_W$ erfüllt und stetig im Nullpunkt ist, nämlich indem man $\psi_1(0_V) = 0_W$ und $\psi_1(h) = \|h\|^{-1}\psi(h)$ für $h \neq 0_V$ definiert. Ist umgekehrt $\psi_1 : U_a \to W$ eine solche Funktion, dann erfüllt die Funktion $\psi : U \to W$ gegeben durch $\psi(h) = \|h\|\psi_1(h)$ offenbar die Bedingung $\lim_{h\to 0} \|h\|^{-1}\psi(h) = 0_W$. Wir können die Differenzierbarkeit einer Funktion f wie in Definition (6.1) im Punkt a also auch folgendermaßen formulieren: Es gibt ein $\phi \in \mathrm{Hom}_{\mathbb{R}}(V, W)$ und eine Funktion $\psi_1 : U_a \to V$, stetig im Nullpunkt mit $\psi_1(0_V) = 0_W$, so dass

$$f(a+h) \quad = \quad f(a) + \phi(h) + \|h\|\psi_1(h)$$

für alle $h \in U_a$ erfüllt ist. Mit dieser Formulierung lässt sich in den nachfolgenden Beweisen bequemer arbeiten.

(6.9) Satz *(Summenregel)*

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

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

Beweis: Nach Definition der Differenzierbarkeit im Punkt a bzw. der soeben eingeführten Umformulierung können wir beide Funktionen in der Form

$$\begin{aligned} f(a+h) &= f(a) + df(a)(h) + \|h\|\varphi_1(h) \\ g(a+h) &= g(a) + dg(a)(h) + \|h\|\psi_1(h) \end{aligned}$$

darstellen, wobei die Funktionen $\varphi_1, \psi_1 : U_a \to W$ definiert auf $U_a = \{h \in V \mid a + h \in U\}$ beide stetig in 0_V sind und $\varphi_1(0_V) = \psi_1(0_V) = 0_W$ erfüllen. Es folgt

$$(f+g)(a+h) \quad = \quad (f+g)(a) + (df(a) + dg(a))(h) + \|h\|(\varphi_1 + \psi_1)(h)$$

für alle $h \in U_a$. Dabei ist $df(a) + dg(a) \in \mathrm{Hom}_{\mathbb{R}}(V, W)$, und die Funktion $\varphi_1 + \psi_1$ ist stetig im Nullpunkt und erfüllt $(\varphi_1 + \psi_1)(0_V) = 0_W$. Also ist $f + g$ in a differenzierbar, und es gilt $d(f+g)(a) = df(a) + dg(a)$. Ebenso erhalten wir $(\lambda f)(a+h) = (\lambda f)(a) + (\lambda df(a))(h) + \|h\|(\lambda\varphi_1)(h)$, wobei auch $\lambda\varphi_1$ stetig im Nullpunkt ist und $(\lambda\varphi_1)(0_V) = 0_W$ erfüllt. Dies zeigt, dass auch λf in a differerenzierbar ist, mit der Ableitung $d(\lambda f)(a) = \lambda df(a)$ im Punkt a. □

(6.10) 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 fg in a differenzierbar, und es gilt

$$d(fg)(a) \quad = \quad f(a)dg(a) + g(a)df(a).$$

Beweis: Wieder verwenden wir die Differenzierbarkeit der Funktionen f und g im Punkt a, um die Funktionen in der oben angegebene Form darzustellen, also durch

$$\begin{aligned} f(a+h) &= f(a) + df(a)(h) + \|h\|\varphi_1(h) \\ g(a+h) &= g(a) + dg(a)(h) + \|h\|\psi_1(h) \end{aligned}$$

mit Funktionen φ_1, ψ_1, die im Nullpunkt stetig sind und $\varphi_1(0_V) = \psi_1(0_V) = 0$ erfüllen. Durch Multiplikation dieser beiden Gleichungen erhalten wir

$$(fg)(a+h) = (fg)(a) + f(a)dg(a)(h) + g(a)df(a)(h) + \rho(h) \tag{6.11}$$

mit dem Restterm

$$\begin{aligned} \rho(h) = {} & df(a)(h)dg(a)(h) + g(a)\|h\|\varphi_1(h) + f(a)\|h\|\psi_1(h) + \\ & dg(a)(h)\|h\|\varphi_1(h) + df(a)(h)\|h\|\psi_1(h) + \|h\|^2\varphi_1(h)\psi_1(h). \end{aligned}$$

Jeder der sechs Summanden des Restterms läuft für $h \to 0_V$ auch nach Division durch $\|h\|$ noch gegen Null. Für die hinteren fünf Summanden ist dies klar auf Grund der Stetigkeit von $\varphi_1(h)$ und $\psi_1(h)$ im Nullpunkt und wegen der Stetigkeit von linearen Abbildungen auf endlich-dimensionalen $\mathbb{R}$-Vektorräumen. Für den ersten Summanden ist zu beachten, dass die linearen Abbildungen $df(a)$ und $dg(a)$ für alle $h \in V$ die Ungleichungen $\|df(a)(h)\| \leq \|df(a)\| \cdot \|h\|$ und $\|dg(a)(h)\| \leq \|dg(a)\| \cdot \|h\|$ erfüllen, wobei $\|df(a)\|$ und $\|dg(a)\|$ die Operatornormen von $df(a)$ und $dg(a)$ bezüglich der Normen auf V und W bezeichnen. Es folgt

$$\|df(a)(h)dg(a)(h)\| \leq \|df(a)\|\|dg(a)\|\|h\|^2.$$

Der Term rechts konvergiert für $h \to 0_V$ auch nach Division durch $\|h\|$ noch gegen 0, also gilt dasselbe für den Ausdruck $\|df(a)(h)dg(a)(h)\|$. Damit ist insgesamt nachgewiesen, dass fg in a differenzierbar ist und $d(fg)(a) = f(a)dg(a) + g(a)df(a)$ gilt. □

(6.12) 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

$$dg(a) = nf(a)^{n-1}df(a).$$

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 $dh(a) = nf(a)^{n-1}df(a)$. Durch Anwendung der Produktregel erhalten wir

$$\begin{aligned} dg(a) = d(fh)(a) &= f(a)dh(a) + h(a)df(a) = nf(a)^n df(a) + f(a)^n df(a) \\ &= (n+1)f(a)^n df(a). \end{aligned}$$

□

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

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

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 $dg(x,y,z) = 3f(x,y,z)^2 df(x,y,z)$ und somit

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

Auch die folgende Ableitungsregel lässt sich ohne Schwierigkeiten auf den mehrdimensionalen Fall übertragen.

(6.13) Satz *(Kettenregel)*

Seien X, Y, Z endlich-dimensionale normierte $\mathbb{R}$-Vektorräume und $U \subseteq X$, $V \subseteq Y$ offene Teilmengen. Seien $f : U \to Y$ und $g : V \to Z$ Abbildungen mit $f(U) \subseteq V$. 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 $g \circ f : U \to Z$ in a differenzierbar, und es gilt

$$d(g \circ f)(a) \quad = \quad dg(f(a)) \circ df(a).$$

Beweis: Sei $b = f(a)$, $U_a = \{h \in X \mid a+h \in U\}$ und $V_b = \{k \in Y \mid b+k \in V\}$. Wie beim Beweis der vorhergehenden beiden Regeln schreiben wir die Funktionen in der Form

$$\begin{aligned} f(a+h) &= f(a) + df(a)(h) + \|h\|\psi_1(h) \\ g(b+k) &= g(b) + dg(b)(k) + \|k\|\varphi_1(k) \quad , \end{aligned}$$

wobei $\psi_1 : U_a \to Y$ und $\varphi_1 : V_b \to Z$ stetig im Nullpunkt sind und $\psi_1(0_X) = 0_Y$, $\varphi_1(0_Y) = 0_Z$ erfüllen. Sei nun $h \in U_a$ vorgegeben. Setzen wir zur Abkürzung $k = df(a)(h) + \|h\|\psi_1(h)$, dann erhalten wir

$$(g \circ f)(a+h) \quad = \quad g(f(a)+k) \quad = \quad g(f(a)) + dg(f(a))(k) + \|k\|\varphi_1(k).$$

Der zweite Summand lautet in ausgeschriebener Form

$$dg(f(a))(k) \quad = \quad dg(f(a))(df(a)(h) + \|h\|\psi_1(h)) \quad = \quad (dg(f(a)) \circ df(a))(h) + dg(f(a))\|h\|\psi_1(h).$$

Setzen wir $\rho_1(h) = dg(f(a))\|h\|\psi_1(h)$, dann läuft dieser Term für $h \to 0_X$ wegen $\|\rho_1(h)\| \leq \|dg(f(a))\|\|h\|\|\psi_1(h)\|$ auch nach Division durch $\|h\|$ noch gegen 0_Z. Den dritten Summanden von oben schätzen wir ab durch

$$\begin{aligned} \|k\| \cdot \|\varphi_1(k)\| &= \|df(a)(h) + \|h\|\psi_1(h)\| \cdot \|\varphi_1(k)\| = \|df(a)(h)\| \cdot \|\varphi_1(k)\| + \|h\| \cdot \|\psi_1(h)\| \cdot \|\varphi_1(k)\| \\ &\leq \|df(a)\| \cdot \|h\| \cdot \|\varphi_1(k)\| + \|h\| \cdot \|\psi_1(h)\| \cdot \|\varphi_1(k)\|. \end{aligned}$$

Für $h \to 0_X$ läuft $k = df(a)(h) + \|h\|\psi_1(h)$ auf Grund der Stetigkeit der linearen Abbildung $df(a)$ gegen 0_Y, und weil φ_1 in 0_Y stetig ist, läuft $\|\varphi_1(k)\|$ gegen 0_Z. Damit ist gezeigt, dass der Ausdruck $\|df(a)\| \cdot \|h\| \cdot \|\varphi_1(k)\| + \|h\| \cdot \|\psi_1(h)\| \cdot \|\varphi_1(k)\|$ auch nach Division durch $\|h\|$ für $h \to 0_X$ gegen Null läuft, und somit gilt dasselbe auch für den dritten Summanden $\|k\|\varphi_1(k)$ von oben. Insgesamt haben wir damit nachgewiesen, dass $g \circ f$ dargestellt werden kann in der Form

$$(g \circ f)(a+h) \quad = \quad (g \circ f)(a) + (dg(f(a)) \circ df(a))(h) + \|k\|\varphi_1(k) + dg(f(a))\|h\|\psi_1(h) \quad ,$$

wobei die hinteren beiden Terme auch nach Division durch $\|h\|$ noch gegen null laufen. Also ist $g \circ f$ in a differenzierbar, und die Ableitung von $g \circ f$ an der Stelle h ist gegeben durch $dg(f(a)) \circ df(a)$ □

Ist insbesondere $X = \mathbb{R}^p$, $Y = \mathbb{R}^n$ und $Z = \mathbb{R}^m$ mit $m, n, p \in \mathbb{N}$, dann können wir nach Proposition (6.7) die Ableitung $df(a)$ im Punkt a als $n \times p$-Matrix und $dg(f(a))$ als $m \times n$-Matrix auffassen. Die Matrix $d(g \circ f)(a) = dg(f(a)) \circ df(a) \in \mathscr{M}_{m\times p,\mathbb{R}}$ ist dann einfach gegeben durch das Matrixprodukt von $dg(f(a))$ und $df(a)$, denn wie wir aus der Linearen Algebra wissen, entspricht die Komposition zweier linearer Abbildungen dem Produkt der Darstellungsmatrizen.

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

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

Die Anwendung der Kettenregel liefert

$$\begin{aligned} d(g \circ f)(x,y,z) \quad &= \quad dg(f(x,y,z)) \circ df(x,y,z) \quad = \quad dg(x+y+z) \circ df(x,y,z) \quad = \\ \begin{pmatrix} 3(x+y+z)^2 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \quad &= \quad \begin{pmatrix} 3(x+y+z)^2 & 3(x+y+z)^2 & 3(x+y+z)^2 \end{pmatrix}. \end{aligned}$$

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

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 Proposition (6.3) gegeben durch

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

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

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

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

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 $d(F \circ f)(x) = df_1(x) + ... + df_m(x)$.

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

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

und es gilt

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

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

§ 7. *Lokale Umkehrbarkeit und implizit definierte Funktionen*

Inhaltsübersicht

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 $df(a)$ 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

- $\mathscr{C}^1$-Abbildung und $\mathscr{C}^1$-Diffeomorphismus
- implizit definierte Funktion
- Umkehrregel
- Schrankensatz
- Satz über die lokale Umkehrbarkeit
- Satz über implizit definierte Funktionen

In § 6 haben wir die aus der Analysis einer Variablen bekannten Ableitungsregeln auf mehrdimensionale Funktionen übertragen, mit einer Ausnahme: der Umkehrregel. Das holen wir nun als erstes nach.

(7.1) 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 $\phi(0_{\mathbb{R}^n}) = df(a)$ 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

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

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 $\varrho(h) = \phi_{R(h)}$, wobei die Matrix $R(h) = (r_{ij}(h)) \in \mathscr{M}_{m\times n,\mathbb{R}}$ für $h \neq 0_{\mathbb{R}^n}$ durch $r_{ij}(h) = \frac{\psi_i(h)}{\|h\|_2^2} h_j$ gegeben und $R(0_{\mathbb{R}^n})$ die Nullmatrix ist. Für $h \to 0_{\mathbb{R}^n}$ konvergieren die Einträge dieser Matrix wegen $|h_j| \leq \|h\|_2$ 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 $\varrho(h)(h) = \phi_{R(h)}(h) = R(h)\cdot h$ 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 $\phi(h) = df(a) + \varrho(h)$ für $h \in U'$. □

In der folgenden Fassung der Umkehrregel ist die Formulierung der Voraussetzungen noch etwas sperrig. Sobald wir den Satz über die lokale Umkehrbarkeit bewiesen haben, werden wir dieses Problem beheben können.

> **(7.2) Proposition** 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 $\det df(a) \neq 0$. Schließlich setzen wir noch voraus, dass g in $b = f(a)$ stetig ist. Dann ist g in b differenzierbar, und es gilt $dg(b) = df(a)^{-1}$.

Beweis: Durch Anwendung der Kettenregel kann der Beweis auf $a = b = 0_{\mathbb{R}^n}$ zurückgeführt werden. Auf Grund des Lemmas existiert dann eine offene Umgebung U' von $0_{\mathbb{R}^n}$ und eine Abbildung $\phi : U' \to \mathscr{L}(\mathbb{R}^n,\mathbb{R}^n)$ mit $f(h) = \phi(h)(h)$ für alle $h \in U'$ und $\lim\limits_{h\to 0} \phi(h) = df(0_{\mathbb{R}^n})$. Wegen $\det df(0_{\mathbb{R}^n}) \neq 0$ 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 $\lim\limits_{h\to 0} \phi(h)^{-1} = df(0_{\mathbb{R}^n})^{-1}$.

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 $\lim\limits_{k\to 0} \phi(g(k))^{-1} = df(0_{\mathbb{R}^n})^{-1}$ folgt. Also können wir g in der Form

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

darstellen, wobei die Abbildung $\rho(k) = \phi(g(k))^{-1} - df(0_{\mathbb{R}^n})^{-1}$ 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 $g(k) = g(0_{\mathbb{R}^n}) + df(0_{\mathbb{R}^n})^{-1}(k) + \psi(k)$ 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 $dg(0_{\mathbb{R}^n}) = df(0_{\mathbb{R}^n})^{-1}$. □

Wenden wir uns nun dem Konzept der lokalen Umkehrbarkeit zu. 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.

(7.3) 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 eine $\mathscr{C}^1$-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.

(7.4) 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 $\gamma_U = \sup\{\|df(x)\| \mid x \in U\}$, 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 $\|df(x)\|$ bezeichnet die entsprechende Operatornorm von $df(x) \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$.

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 $df(a) = (df_1(a), ..., df_m(a))$ und somit $|df_i(a)(v)| \leq \max\{|df_j(a)(v)| \mid 1 \leq j \leq m\} = \|df(a)(v)\| \leq \gamma_U \|v\|$, 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 (5.5) für Richtungsableitungen gibt es für $1 \leq j \leq m$ jeweils ein $a_j \in [x_1, x_2]$ mit $df_j(a_j)(v) = \partial_v f_j(a_j) = f_j(x_2) - f_j(x_1)$, wobei $v = x_2 - x_1$ ist. Es folgt $|f_j(x_1) - f_j(x_2)| \leq |df_j(a_i)| \leq \gamma_U \|v\|$ und somit $\|f(x_1) - f(x_2)\| \leq \gamma_U \|v\|$. □

(7.5) 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 $df(a) \in \mathscr{L}(V, W)$ 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 $df(a)$ 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 $\kappa_1 : W \to \mathbb{R}^n$ von $\mathbb{R}$-Vektorräumen, die nach Folgerung (2.20) 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 $\tilde{f} = \kappa_1 \circ f \circ \kappa^{-1}$ 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

$$\begin{aligned} d\tilde{f}(\kappa(a)) &= d(\kappa_1 \circ f \circ \kappa^{-1})(\kappa(a)) = d\kappa_1((f \circ \kappa^{-1})(\kappa(a))) \circ d(f \circ \kappa^{-1})(\kappa(a)) = \\ & d\kappa_1(f(a)) \circ df(\kappa^{-1}(\kappa(a))) \circ d(\kappa^{-1})(\kappa(a)) = \kappa_1 \circ df(a) \circ \kappa^{-1}. \end{aligned}$$

Mit $a \mapsto df(a)$ ist auch die Abbildung $\kappa(a) \mapsto \kappa \circ df(a) \circ \kappa^{-1}$ 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 $df(a)$ 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 $df(0_{\mathbb{R}^n})^{-1} \circ f$ ersetzt werden kann, haben wir die Aussage auf den Fall $df(0_{\mathbb{R}^n}) = \mathrm{id}_{\mathbb{R}^n}$ 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 (1.22) 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 $\|\mathrm{id}_{\mathbb{R}^n} - df(x)\| \leq \frac{1}{2}$ für alle $x \in \bar{B}_{2r}(0_{\mathbb{R}^n})$ gilt; dies ist wegen $df(0_{\mathbb{R}^n}) = \mathrm{id}_{\mathbb{R}^n}$ und auf Grund der Stetigkeit der Ableitungsfunktion df möglich. Für alle x in diesem abgeschlossenen Ball gilt dann $d\varphi_w(x) = \mathrm{id}_{\mathbb{R}^n} - df(x)$, und mit dem Schrankensatz (7.4) 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 (3.9) 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 in Satz (7.2) zeigt nun, dass g auch differenzierbar ist, mit der Ableitungsfunktion gegeben durch $dg(w) = df((f|_{\tilde{U}})^{-1}(w))^{-1}$. 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.

(7.6) Folgerung Sei $U \subseteq \mathbb{R}^n$ offen und $f : U \to \mathbb{R}^n$ eine injektive, stetig differenzierbare Abbildung mit der Eigenschaft, dass $df(x)$ 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 (2.13) 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 (2.14) definiert. Beispielsweise ist die Jacobi-Matix von ρ_{kug} in jeden Punkt des Definitionsbereichs gegeben durch die Matrix

$$d\rho_{\text{kug}}(r, \vartheta, \varphi) = \begin{pmatrix} \sin(\vartheta)\cos(\varphi) & r\cos(\vartheta)\cos(\varphi) & -r\sin(\vartheta)\sin(\varphi) \\ \sin(\vartheta)\sin(\varphi) & r\cos(\vartheta)\sin(\varphi) & r\sin(\vartheta)\cos(\varphi) \\ \cos(\vartheta) & -r\sin(\vartheta) & 0 \end{pmatrix}$$

mit der Determinante $\det d\rho_{\text{kug}}(r, \vartheta, \varphi) = r^2 \sin(\vartheta)$. Weil $\sin(\vartheta) > 0$ für $\vartheta \in]0, \pi[$ gilt, ist die Ableitung in jedem Punkt des Definitionsbereichs invertierbar.

Die Voraussetzungen der Umkehrregel können nun einfacher formuliert werden.

(7.7) Satz *(Umkehrregel)*
Sei $U \subseteq \mathbb{R}^n$ offen und $f : U \to \mathbb{R}^n$ eine $\mathscr{C}^1$-Abbildung. Sei $a \in U$ ein Punkt mit $\det df(a) \neq 0$. Dann existieren offene Umgebungen $U_1 \subseteq U$ von a und $V_1 \subseteq \mathbb{R}^n$ von $b = f(a)$, so dass durch $f_1 = f|_{U_1}$ ein $\mathscr{C}^1$-Diffeomorphismus $U_1 \to V_1$ definiert ist. Die Ableitung der Umkehrabbildung $f_1^{-1} : V_1 \to U_1$ erfüllt dabei

$$d(f_1^{-1})(y) \quad = \quad df(f_1^{-1}(y))^{-1}$$

für alle $y \in V_1$. Insbesondere gilt also $d(f_1^{-1})(b) = df(a)^{-1}$.

Beweis: Die Existenz der Umgebung U_1 und V_1 ergeben sich direkt aus dem Satz (7.5) über die lokale Umkehrbarkeit. Dies zeigt, dass die Voraussetzungen der Umkehrregel in der Fassung von Proposition (7.2) erfüllt sind. Aus dieser wiederum ergibt sich die Differenzierbarkeit der Umkehrabbildung und der angegebene Wert der Ableitung. □

Wir zeigen, wie die Umkehrregel auf die Polarkoordinaten-Abbildung angewendet werden kann. Sei $(r, \varphi) \in \mathbb{R}^+ \times \mathbb{R}$ und $(x, y) = \rho_{\text{pol}}(r, \varphi)$, also $x = r\cos(\varphi)$ und $y = r\sin(\varphi)$. Offenbar ist die Funktion ρ_{pol} auf ihrem gesamten Definitionsbereich eine $\mathscr{C}^1$-Abbildung, und die Ableitung an der Stelle (r, φ) ist gegeben durch

$$d\rho_{\text{pol}}(r, \varphi) \quad = \quad \begin{pmatrix} \cos(\varphi) & -r\sin(\varphi) \\ \sin(\varphi) & r\cos(\varphi) \end{pmatrix}.$$

Es gilt $\det d\rho_{\text{pol}}(r, \varphi) = r^2(\cos(\varphi)^2 + \sin(\varphi)^2) = r^2 > 0$. Dies zeigt, dass der Satz über die lokale Umkehrbarkeit angewendet werden kann. Dieser liefert offene Umgebungen U von (r, φ) und V von (x, y) mit der Eigenschaft, dass $\rho_{\text{pol}}|_U$ zu einem Homömorphismus zwischen U und V wird. Mit der Umkehrregel können wir die Ableitung der Umkehrfunktion $\psi = (\rho_{\text{pol}}|_U)^{-1}$ an der Stelle (x, y) ausrechnen. Es gilt

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

also $r = \sqrt{x^2 + y^2}$. Durch Umformen erhalten wir außerdem $\cos(\varphi) = \frac{x}{r}$ und $\sin(\varphi) = \frac{y}{r}$. Mit der Umkehrregel erhalten wir

$$d\psi(x, y) \; = \; d\rho_{\text{pol}}(r, \varphi)^{-1} \; = \; \begin{pmatrix} \cos(\varphi) & -r\sin(\varphi) \\ \sin(\varphi) & r\cos(\varphi) \end{pmatrix}^{-1} \; = \; \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\frac{1}{r}\sin(\varphi) & \frac{1}{r}\cos(\varphi) \end{pmatrix} = \; \begin{pmatrix} \dfrac{x}{\sqrt{x^2+y^2}} & \dfrac{y}{\sqrt{x^2+y^2}} \\ -\dfrac{y}{x^2+y^2} & \dfrac{x}{x^2+y^2} \end{pmatrix}.$$

Wenden wir uns nun dem zweiten großen Thema dieses Kapitels zu, den implizit definierten Funktionen.

(7.8) 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.

(7.9) 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

$$\partial_X f(x,y) : X \to Z\,,\ v \mapsto df(x,y)(v,0) \qquad \text{und} \qquad \partial_Y f(x,y) : Y \to Z\,,\ w \mapsto df(x,y)(0,w).$$

Nach Definition gilt für alle $(x,y) \in U$, $v \in X$ und $w \in Y$ jeweils

$$\partial_X f(x,y)(v) + \partial_Y f(x,y)(w) \quad = \quad df(x,y)(v,0) + df(x,y)(0,w) \quad = \quad df(x,y)(v,w).$$

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 $df(x,y)$. 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.

(7.10) 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

$$(u_1,w) = \Phi(u,v) \quad \Leftrightarrow \quad (u_1,w) = (u, \phi(u) + \psi(v)) \quad \Leftrightarrow \quad (u_1 = u) \wedge (w = \phi(u) + \psi(v)) \quad \Leftrightarrow$$
$$(u = u_1) \wedge (w = \phi(u_1) + \psi(v)) \quad \Leftrightarrow \quad (u = u_1) \wedge (\psi(v) = w - \phi(u_1)) \quad \Leftrightarrow$$
$$(u = u_1) \wedge (v = \psi^{-1}(w) - \psi^{-1}(\phi(u_1))) \quad \Leftrightarrow \quad (u,v) = (u_1, \psi^{-1}(w) - (\psi^{-1} \circ \phi)(u_1)).$$

Dies zeigt, dass $\Psi(u_1,w) = (u_1, \psi^{-1}(w) - (\psi^{-1} \circ \phi)(u_1))$ tatsächlich die Umkehrabbildung von Φ ist. □

(7.11) 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 (7.5) 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

$$f(x,y) = z \quad \Leftrightarrow \quad (x, f(x,y)) = (x,z) \quad \Leftrightarrow \quad \Phi(x,y) = (x,z) \quad \Leftrightarrow \quad (x,y) = \Psi(x,z)$$
$$\Leftrightarrow \quad (x,y) = (x, h(x,z)) \quad \Leftrightarrow \quad y = h(x,z).$$

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 $f(x,y) = 0_Y \Leftrightarrow y = g(x)$ 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 $d\Phi_X(x,y) = \Phi_X$ für alle (x,y) in U. Für die Ableitung von $d\Phi$ an der Stelle (a,b) und $(u,v) \in X \times Y$ erhalten wir

$$d\Phi(a,b)(u,v) \quad = \quad (d\Phi_X(a,b)(u,v), d\Phi_Y(a,b)(u,v)) \quad = \quad (\Phi_X(u,v), df(a,b)(u,v)) \quad =$$

$$(u, \partial_X f(a,b)(u) + \partial_Y f(a,b)(v)).$$

Weil $\partial_Y f(a,b)$ invertierbar ist, erhalten wir mit Lemma (7.10) die Invertierbarkeit von $d\Phi(a,b)$. 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. □

(7.12) Folgerung Die Funktion $g : U' \to U''$ aus dem Satz hat an der Stelle a die Ableitung

$$dg(a) \quad = \quad -\partial_Y f(a,b)^{-1} \circ \partial_X f(a,b).$$

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 $dg_1(x) = (\mathrm{id}_X, dg(x))$ für alle $x \in U'$. Die Kettenregeln liefert nun

$$d\Phi(x) \quad = \quad df(g_1(x)) \circ dg_1(x) \quad = \quad df(x, g(x)) \circ (\mathrm{id}_X, dg(x)).$$

Wir stellen nun $d\Phi$ mit Hilfe der mehrdimensionalen partiellen Ableitungen von f dar. Für einen beliebigen Vektor $v \in X$ gilt

$$d\Phi(x)(v) \quad = \quad df(x,g(x))((\mathrm{id}_X, dg(x))(v)) \quad = \quad df(x,g(x))(v, dg(x)(v)) \quad =$$

$$\partial_X f(x,g(x))(v) + \partial_Y f(x,g(x))(dg(x)(v)) \quad = \quad \partial_X f(x,g(x))(v) + (\partial_Y f(x,g(x)) \circ dg(x))(v)$$

also $d\Phi(x) = \partial_X f(x,g(x)) + \partial_Y f(x,g(x)) \circ dg(x)$. Nun ist die Funktion g gerade so definiert, dass $\Phi(x) = f(x,g(x)) = 0$ für alle $x \in U'$ gilt. Folglich ist $d\Phi(x) = 0$ für alle $x \in U'$. Wegen $b = g(a)$ und auf Grund der Invertierbarkeit von $\partial_Y f$ im Punkt (a,b) erhalten wir

$$\partial_X f(a,b) + \partial_Y f(a,b) \circ dg(a) = 0 \quad \Leftrightarrow \quad \partial_Y f(a,b) \circ dg(a) = -\partial_X f(a,b)$$
$$\Leftrightarrow \quad dg(a) = -\partial_Y f(a,b)^{-1} \circ \partial_X f(a,b).$$ □

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

$$\begin{array}{rcl} x^3+y^3+z^3 &=& 7 \\ xy+yz+xz &=& -2 \end{array} \quad \Leftrightarrow \quad \begin{array}{rcl} y &=& g(x) \\ z &=& h(x) \end{array}$$ 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

$$df(x,y,z) \quad = \quad \begin{pmatrix} 3x^2 & 3y^2 & 3z^2 \\ y+z & x+z & x+y \end{pmatrix}$$

also insbesondere

$$df(2,-1,0) = \begin{pmatrix} 12 & 3 & 0 \\ -1 & 2 & 1 \end{pmatrix} \quad , \quad \partial_X f(2,-1,0) = \begin{pmatrix} 12 \\ -1 \end{pmatrix} \quad , \quad \partial_Y f(2,-1,0) = \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix}.$$

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 $dk(2)$ erhalten wir nach Folgerung (7.12) durch

$$dk(2) \quad = \quad -\partial_Y f(2,-1,0)^{-1} \circ \partial_X f(2,-1,0) \quad = \quad -\tfrac{1}{3}\begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 12 \\ -1 \end{pmatrix}$$

$$= \quad -\tfrac{1}{3}\begin{pmatrix} 1 & 0 \\ -2 & 3 \end{pmatrix}\begin{pmatrix} 12 \\ -1 \end{pmatrix} \quad = \quad \begin{pmatrix} -4 \\ 9 \end{pmatrix}.$$

Es gilt also $g'(2) = -4$ und $h'(2) = 9$.

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.