Kapitel 6

Differentiation im $\mathbb{R}^n$

1 Topologie im $\mathbb{R}^n$

Hier wiederholen wir kurz die topologischen Grundbegriffe im $\mathbb{R}^n$. Das griechische Wort $\tau\acute{o}\pi o\varsigma$ bedeutet soviel wie Ort oder Lage. Mathematisch geht es in der Topologie um Mengen mit einem Konvergenzbegriff, und stetige Abbildungen zwischen diesen Mengen. Unser Interesse gilt natürlich hauptsächlich Teilmengen des $\mathbb{R}^n$, wir beginnen aber mit dem allgemeinen Begriff des metrischen Raums. Zum einen werden uns metrische Räume später vielfach begegnen, zum anderen wird so klarer, auf welche Eigenschaften des $\mathbb{R}^n$ es hier ankommt.

Definition 1.1 (Metrischer Raum) *Ein metrischer Raum ist eine Menge X mit einer Funktion $d : X \times X \to [0, \infty)$, die für alle $x, y, z \in X$ folgende Eigenschaften hat:*

> Positivität: $d(x, y) \geq 0$ *mit Gleichheit genau wenn* $x = y$,
>
> Symmetrie: $d(y, x) = d(x, y)$
>
> Dreiecksungleichung: $d(x, z) \leq d(x, y) + d(y, z)$.

Wir nennen $d(x, y)$ auch den Abstand von x und y.

In dieser Definition kann X eine beliebige Menge sein, insbesondere muss X kein Vektorraum sein. Betrachten Sie als Beispiel die Menge X aller Bahnhöfe in Frankreich und

$$d(x, y) = \begin{cases} \text{minimale Fahrzeit von } x \text{ nach } y \text{ über Paris} & \text{für } x \neq y, \\ 0 & \text{für } x = y. \end{cases} \tag{1.1}$$

Viele interessante metrische Räume sind normierte Vektorräume.

Definition 1.2 (Norm) *Eine Norm auf dem reellen (oder komplexen) Vektorraum X ist eine Funktion $\|\cdot\| : X \to \mathbb{R}$ mit folgenden Eigenschaften:*

> Positivität: $\|x\| \geq 0$, *mit Gleichheit genau wenn* $x = 0$.
>
> Halblinearität: $\|\lambda x\| = |\lambda| \cdot \|x\|$ *für alle* $\lambda \in \mathbb{R}$, $x \in X$.
>
> Dreiecksungleichung: $\|x + y\| \leq \|x\| + \|y\|$ *für alle* $x, y \in X$.

Das wichtigste Beispiel ist natürlich die euklidische Norm auf dem $\mathbb{R}^n$, also

$$|x| = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} \qquad \text{für } x = (x_1, \ldots, x_n) \in \mathbb{R}^n. \tag{1.2}$$

Die Positivität und die Halblinearität ergeben sich dabei leicht, für die Dreiecksungleichung haben wir die Ungleichung von Cauchy-Schwarz gebraucht, siehe Satz 5.2 in Kapitel 2. Andere Normen auf $\mathbb{R}^n$ sind zum Beispiel die 1-Norm und die Maximumsnorm

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \quad \text{und} \quad \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|. \tag{1.3}$$

Jeder normierte Vektorraum $(X, \|\cdot\|)$ wird zu einem metrischen Raum, indem wir den Abstand von zwei Punkten x, y erklären durch

$$d(x,y) = \|x - y\| \quad \text{für } x, y \in X. \tag{1.4}$$

Denn offensichtlich gilt $d(x,y) \geq 0$ mit Gleichheit nur für $x = y$, sowie

$$\begin{aligned} d(y,x) &= \|y - x\| = \|(-1)(x-y)\| = |(-1)|\,\|x - y\| = d(x,y), \\ d(x,z) &= \|x - z\| = \|(x-y) + (y-z)\| \leq \|x - y\| + \|y - z\| = d(x,y) + d(y,z). \end{aligned}$$

Insbesondere ist $\mathbb{R}^n$ ein metrischer Raum mit dem üblichen euklidischen Abstandsbegriff.

Definition 1.3 *Sei X ein metrischer Raum. Die offene Kugel um x_0 mit Radius $r > 0$ ist*

$$B_r(x_0) = \{x \in X : d(x, x_0) < r\}.$$

Bezüglich der Euklidischen Norm auf $\mathbb{R}^n$ gilt also wie gewohnt

$$B_r(x_0) = \{x \in \mathbb{R}^n : |x - x_0| < r\}.$$

Es ist instruktiv, sich die Kugeln $B_r(x_0)$ für die französische Eisenbahnmetrik aus (1.1) sowie die Kugeln $B_1(0)$ für die Normen $\|\cdot\|_1$ und $\|\cdot\|_\infty$ auf $\mathbb{R}^n$ zu überlegen.

Definition 1.4 *Sei X eine metrischer Raum. Eine Menge $\Omega \subset X$ heißt offen, falls zu jedem $x \in \Omega$ ein $\varepsilon > 0$ existiert mit $B_\varepsilon(x) \subset \Omega$.*

Beispiel 1.1 Die Kugel $B_r(x_0)$ ist offen im Sinn der Definition 1.4. Sei nämlich $x \in B_r(x_0)$ gegeben. Dann ist $\varepsilon = r - d(x, x_0) > 0$ und für $y \in B_\varepsilon(x)$ folgt aus der Dreiecksungleichung

$$d(y, x_0) \leq d(y, x) + d(x, x_0) < \varepsilon + d(x, x_0) = r,$$

also $B_\varepsilon(x) \subset B_r(x_0)$, was zu zeigen war.

Satz 1.1 *Für die offenen Teilmengen eines metrischen Raums X gilt:*

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

(b) *Der Durchschnitt von endlich vielen offenen Mengen ist offen.*

(c) *Die Vereinigung von beliebig vielen offenen Mengen ist offen.*

BEWEIS: Aussage (a) ist klar. Für (b) sei $x \in \bigcap_{i=1}^{N} \Omega_i$, wobei $\Omega_1, \dots, \Omega_N$ endlich viele offene Teilmengen von X sind. Dann gibt es $\varepsilon_i > 0$ mit $B_{\varepsilon_i}(x) \subset \Omega_i$. Es folgt $\varepsilon = \min_{1 \le i \le N} \varepsilon_i > 0$ sowie $B_\varepsilon(x) \subset B_{\varepsilon_i}(x) \subset \Omega_i$ für jedes i, das heißt $B_\varepsilon(x) \subset \bigcap_{i=1}^{N} \Omega_i$.

Für (c) sei nun $x \in \bigcup_{\lambda \in \Lambda} \Omega_\lambda$, wobei Λ eine beliebige Indexmenge ist. Dann ist $x \in \Omega_{\lambda_0}$ für (mindestens) ein $\lambda_0 \in \Lambda$. Da Ω_{λ_0} offen, gibt es ein $\varepsilon > 0$ mit $B_\varepsilon(x) \subset \Omega_{\lambda_0}$, also erst recht $B_\varepsilon(x) \subset \bigcup_{\lambda \in \Lambda} \Omega_\lambda$. □

Ein abzählbarer Schnitt von offenen Mengen ist nicht notwendig offen, zum Beispiel ist

$$\bigcap_{n=1}^{\infty} B_{\frac{1}{n}}(0) = \{0\}$$

nicht offen in $\mathbb{R}^n$. Eine offene Menge $\Omega \subset X$ mit $x \in \Omega$ nennt man auch offene Umgebung von x. Insbesondere wird die offene Kugel $B_\varepsilon(x)$ als ε-Umgebung von x bezeichnet.

Lemma 1.1 *In einem metrischen Raum X gibt es zu zwei Punkten $x, y \in X$ mit $x \neq y$ ein $\varepsilon > 0$ mit $B_\varepsilon(x) \cap B_\varepsilon(y) = \emptyset$.*

BEWEIS: Sei $z \in B_\varepsilon(x) \cap B_\varepsilon(y)$. Dann folgt $d(x,y) \le d(x,z) + d(z,y) < 2\varepsilon$. Also ist die Behauptung richtig für jedes $\varepsilon \le \frac{1}{2} d(x,y)$. □

Definition 1.5 *Sei X ein metrischer Raum. Die Folge $(x_k)_{k \in \mathbb{N}}$ von Punkten $x_k \in X$ konvergiert gegen $x \in X$, falls gilt:*

Für alle $\varepsilon > 0$ gibt es ein $K \in \mathbb{R}$ mit $x_k \in B_\varepsilon(x)$ für alle $k > K$.

Äquivalent dazu ist $d(x_k, x) \to 0$ mit $k \to \infty$.

Der Grenzwert ist eindeutig bestimmt, denn wäre $y \neq x$ ebenfalls Grenzwert von (x_k), so wählen wir $\varepsilon > 0$ wie in Lemma 1.1 und erhalten für k hinreichend groß den Widerspruch

$$x_k \in B_\varepsilon(x) \cap B_\varepsilon(y) = \emptyset.$$

Definition 1.6 *Eine Teilmenge A eines metrischen Raums X heißt abgeschlossen, wenn folgende Implikation stets gilt:*

$$x_k \in A, \quad x_k \to x \quad \Rightarrow \quad x \in A.$$

Satz 1.2 *In einem metrischen Raum X gilt für jede Menge $M \subset X$:*

$$M \text{ offen} \quad \Leftrightarrow \quad X \backslash M \text{ abgeschlossen}.$$

BEWEIS: Kapitel 2, Satz 5.4. □

Folgerung 1.1 *Für die abgeschlossenen Teilmengen eines metrischen Raums X gilt:*

a) *$\emptyset, X$ sind abgeschlossen.*

b) *Die Vereinigung von endlich vielen abgeschlossenen Mengen ist abgeschlossen.*

c) *Der Durchschnitt von beliebig vielen abgeschlossenen Mengen ist abgeschlossen.*

Beweis: Folgt aus Satz 1.1 und Satz 1.2. □

Die Vereinigung von unendlich vielen abgeschlossenen Mengen ist nicht notwendig abgeschlossen, zum Beispiel $\bigcup_{n=1}^{\infty}[\frac{1}{n},1] = (0,1] \subset \mathbb{R}$.

Beispiel 1.2 (Relativtopologie) Ist (X,d) metrischer Raum, so ist jede Teilmenge $M \subset X$ selbst ein metrischer Raum mit der induzierten Abstandsfunktion

$$d_M : M \times M \to [0,\infty),\ d_M(x,y) = d(x,y). \tag{1.5}$$

Zum Beispiel ist die Sphäre $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$ ein metrischer Raum mit dem euklidischen Abstand $d_{\mathbb{S}^{n-1}}(x,y) = |x-y|$. Für die Kugeln bezüglich der induzierten Abstandsfunktion gilt

$$B_r^M(x) = \{y \in M : d_M(y,x) < r\} = \{y \in M : d(y,x) < r\} = B_r(x) \cap M.$$

Ist $\tilde{U}$ offen in X, so ist $\tilde{U} \cap M$ offen in (M,d_M). Denn für $x \in \tilde{U} \cap M$ gilt $B_\varepsilon(x) \subset \tilde{U}$ für geeignetes $\varepsilon > 0$, also $B_\varepsilon^M(x) \subset \tilde{U} \cap M$. Umgekehrt ist jede offene Menge $U \subset (M,d_M)$ von dieser Form: zu $x \in U$ gibt es ein $\varepsilon(x) > 0$ mit $B_{\varepsilon(x)}^M(x) \subset U$. Die Menge

$$\tilde{U} = \bigcup_{x \in U} B_{\varepsilon(x)}(x)$$

ist als Vereinigung offener Kugeln offen in X, und es gilt

$$U \subset \tilde{U} \cap M = \bigcup_{x \in U} B_{\varepsilon(x)}(x) \cap M = \bigcup_{x \in U} B_{\varepsilon(x)}^M(x) \subset U,$$

also

$$U = \widetilde{U} \cap M.$$

Entsprechendes gilt für die abgeschlossenen Mengen, das heißt eine Menge $A \subset (M,d_M)$ ist genau dann abgeschlossen, wenn sie von der Form $A = \tilde{A} \cap M$ ist mit $\tilde{A}$ abgeschlossen in X.

In der eindimensionalen Analysis wurden meist Funktionen auf einem Intervall I mit Randpunkten $a < b$ betrachtet. Im mehrdimenionalen Fall werden wir oft Kugeln $B_r(x)$ oder achsenparallele Quader $I_1 \times \ldots \times I_n$ betrachten, bisweilen aber auch kompliziertere Mengen. Dafür sind die folgenden Begriffe nützlich.

Definition 1.7 *Sei X ein metrischer Raum und $M \subset X$. Dann definieren wir*

$$\begin{aligned}
\operatorname{int} M &= \{x \in M : \exists\, \varepsilon > 0 \textit{ mit } B_\varepsilon(x) \subset M\} \quad \textit{(Menge der inneren Punkte von } M\textit{)},\\
\overline{M} &= \{x \in X : \forall\, \varepsilon > 0 \textit{ ist } B_\varepsilon(x) \cap M \neq \emptyset\} \quad \textit{(Abschluss von } M\textit{)},\\
\partial M &= \{x \in X : \forall\, \varepsilon > 0 \textit{ sind } B_\varepsilon(x) \cap M,\ B_\varepsilon(x) \cap (X \backslash M) \neq \emptyset\} \quad \textit{(Rand von } M\textit{)}.
\end{aligned}$$

Trivialerweise gilt $\operatorname{int} M \subset M \subset \overline{M}$. Außerdem ist $\operatorname{int} \Omega = \Omega$ für $\Omega \subset X$ offen sowie $\overline{M} = M$ für $M \subset X$ abgeschlossen.

Beispiel 1.3 Auf dem $\mathbb{R}^n$ mit der euklidischen Abstandsfunktion $d(x, y) = |x - y|$ gilt für die Kugel $B_r(x) = \{y \in \mathbb{R}^n : d(y, x) < r\}$:

$$\overline{B_r(x)} = \{y \in \mathbb{R}^n : d(y, x) \leq r\} \quad \text{und} \quad \partial B_r(x) = \{x \in \mathbb{R}^n : d(y, x) = r\}.$$

Als erstes zeigen wir $\overline{B_r(x)} \subset \{y \in \mathbb{R}^n : d(y, x) \leq r\}$. Zu $y \in \overline{B_r(x)}$ und jedem $\varepsilon > 0$ gibt es ein $z \in B_r(x)$ mit $d(y, z) < \varepsilon$, also $d(y, x) \leq d(y, z) + d(z, x) < \varepsilon + r$, und mit $\varepsilon \to 0$ folgt $d(y, x) \leq r$. Analog ergibt sich die Inklusion $\overline{\mathbb{R}^n \backslash B_r(x)} \subset \{x \in \mathbb{R}^n : d(y, x) \geq r\}$, denn zu $y \in \overline{\mathbb{R}^n \backslash B_r(x)}$ gibt es ein $z \in \mathbb{R}^n \backslash B_r(x)$ mit $d(y, z) < \varepsilon$, also $d(y, x) \geq d(z, x) - d(y, z) > r - \varepsilon$, und mit $\varepsilon \to 0$ folgt $d(y, x) \geq r$. Aber $\mathbb{R}^n \backslash B_r(x) \subset \overline{\mathbb{R}^n \backslash B_r(x)}$, und somit $\overline{\mathbb{R}^n \backslash B_r(x)} = \{y \in \mathbb{R}^n : d(y, x) \geq r\}$. Ist nun $d(y, x) \leq r$, so folgt für $0 < \theta < 1$ hinreichend nahe bei Eins

$$d(x + \theta(y - x), x) = \theta|y - x| < r \quad \text{und} \quad d(x + \theta(y - x), y) = (1 - \theta)|x - y| < \varepsilon,$$

das heißt $y \in \overline{B_r(x)}$ und somit $\overline{B_r(x)} = \{y \in \mathbb{R}^n : d(y, x) \leq r\}$. Wegen $\partial B_r(x) = \overline{B_r(x)} \cap \overline{(\mathbb{R}^n \backslash B_r(x))}$ folgt schließlich $\partial B_r(x) = \{y \in \mathbb{R}^n : d(y, x) = r\}$.

Satz 1.3 *Sei M Teilmenge des metrischen Raums X.*

(a) $\operatorname{int} M$ *ist offen, und es gilt die Implikation*

$$\Omega \textit{ offen}, \Omega \subset M \quad \Rightarrow \quad \Omega \subset \operatorname{int} M.$$

(b) $\overline{M}$ *ist abgeschlossen, und es gilt die Implikation*

$$A \textit{ abgeschlossen}, A \supset M \quad \Rightarrow \quad A \supset \overline{M}.$$

(c) ∂M *ist abgeschlossen und es gilt* $\partial M = \overline{M} \backslash \operatorname{int} M$.

Beweis: Für (a) sei $x \in \operatorname{int} M$, also $B_r(x) \subset M$ für ein $r > 0$. Für $y \in B_r(x)$ gilt dann $B_\varepsilon(y) \subset B_r(x) \subset M$ mit $\varepsilon = r - d(y, x) > 0$, vgl. Beispiel 1.1. Es folgt $B_r(x) \subset \operatorname{int} M$, damit ist $\operatorname{int} M$ offen. Sei nun Ω offen und $\Omega \subset M$. Zu $x \in \Omega$ gibt es ein $\varepsilon > 0$ mit $B_\varepsilon(x) \subset \Omega$, also auch $B_\varepsilon(x) \subset M$, das heißt $x \in \operatorname{int} M$.

Für (b) verwenden wir (a) und Satz 1.2. Nach Definition ist $X \backslash \overline{M} = \operatorname{int}(X \backslash M)$, also ist $\overline{M} = X \backslash \operatorname{int}(X \backslash M)$ abgeschlossen. Ist nun $A \subset X$ eine beliebige abgeschlossene Menge mit $A \supset M$, so ist $X \backslash A$ offen sowie $X \backslash A \subset X \backslash M$, also $X \backslash A \subset \operatorname{int}(X \backslash M)$ nach (a), und somit $A \supset \overline{M}$. Dies beweist (b).

Nach Definition gilt weiter $\partial M = \overline{M} \cap \overline{(X \backslash M)}$, also ist ∂M abgeschlossen nach (b) und Folgerung 1.1. Ferner ist ebenfalls nach Definition $X \backslash \operatorname{int} M = \overline{X \backslash M}$, folglich

$$\partial M = \overline{M} \cap (X \backslash \operatorname{int} M) = \overline{M} \backslash \operatorname{int} M.$$

□

Im Abschluss von M können noch zwei Sorten Punkte unterschieden werden, die Häufungspunkte und die isolierten Punkte.

Definition 1.8 *Ein Punkt $x \in X$ heißt*

$$\begin{aligned} \text{Häufungspunkt von } M \quad &\Leftrightarrow \quad \textit{für jedes } \varepsilon > 0 \textit{ ist } B_\varepsilon(x) \cap M \backslash \{x\} \textit{ nichtleer}, \\ \text{isolierter Punkt von } M \quad &\Leftrightarrow \quad \textit{es gibt ein } \varepsilon > 0 \textit{ mit } M \cap B_\varepsilon(x) = \{x\}. \end{aligned}$$

Ist $x \in X$ Häufungspunkt von M, so enthält $B_\varepsilon(x) \cap M\backslash\{x\}$ sogar unendlich viele Punkte. Denn würde die Menge nur aus endlich vielen Punkten $y_1, \ldots, y_N$ bestehen, so ist $\delta = \min_{1 \leq i \leq N} d(y_i, x) > 0$ und dann $B_\delta(x) \cap M\backslash\{x\} = \emptyset$, ein Widerspruch. Insbesondere können wir eine Folge $x_k \in M\backslash\{x\}$ bestimmen mit $x_k \to x$.

Definition 1.9 *Eine Teilmenge M eines metrischen Raums X heißt dicht, falls $\overline{M} = X$.*

Bekanntes Beispiel sind die rationalen Zahlen $\mathbb{Q}$ im metrischen Raum $\mathbb{R}$, beziehungsweise die rationalen Punkte $\mathbb{Q}^n$ im $\mathbb{R}^n$.

Definition 1.10 (Stetigkeit) *Sei D Teilmenge eines metrischen Raums X, und Y ein weiterer metrischer Raum. Eine Abbildung $f : D \to Y$ heißt stetig im Punkt $x_0 \in D$, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt mit*

$$d(f(x), f(x_0)) < \varepsilon \quad \text{für alle } x \in D \text{ mit } d(x, x_0) < \delta,$$

oder äquivalent mit $f(B_\delta(x_0) \cap D) \subset B_\varepsilon(f(x_0))$. Die Funktion f heißt stetig, wenn sie in jedem Punkt $x_0 \in D$ stetig ist.

Wir müssten hier eigentlich $d_X(\cdot,\cdot)$ und $d_Y(\cdot,\cdot)$ schreiben, denn im allgemeinen sind X und Y verschiedene metrische Räume, jedoch führt die einfachere Notation nicht zu Misverständnissen.

Definition 1.11 (Lipschitzstetigkeit) *Eine Abbildung $f : D \to Y$ heißt Lipschitzstetig mit Konstante $L \geq 0$, falls*

$$d(f(x), f(x')) \leq L\, d(x, x') \quad \text{für alle } x, x' \in D.$$

Beispiel 1.4 Die Abstandsfunktion von einem Punkt $x_0 \in X$, das heißt

$$f : X \to \mathbb{R},\ f(x) = d(x, x_0),$$

ist Lipschitzstetig mit Konstante $L = 1$, denn aus der Dreiecksungleichung folgt

$$f(x) = d(x, x_0) \leq d(x, x') + d(x', x_0) = d(x, x') + f(x').$$

Durch Vertauschen von x und x' folgt $|f(x) - f(x')| \leq d(x, x')$ wie gewünscht.

Die folgende Umformulierung der Stetigkeit erfreut sich zum Teil großer Beliebtheit.

Satz 1.4 (Charakterisierung stetiger Abbildungen) *Seien X, Y metrische Räume und $\Omega \subset X$ sei offen. Eine Abbildung $f : \Omega \to Y$ ist genau dann stetig, wenn für jede jede offene Menge $V \subset Y$ das Urbild $f^{-1}(V)$ offen in X ist.*

Beweis: Sei f stetig, $V \subset Y$ offen und $x_0 \in f^{-1}(V)$. Dann ist $y_0 = f(x_0) \in V$, also gilt $B_\varepsilon(y_0) \subset V$ für geeignetes $\varepsilon > 0$. Es gibt dann ein $\delta > 0$ mit $f(B_\delta(x_0) \cap \Omega) \subset B_\varepsilon(y_0)$. Wir können außerdem $B_\delta(x_0) \subset \Omega$ annehmen, und dann folgt $B_\delta(x_0) \subset f^{-1}(V)$ wie verlangt.

Umgekehrt sei $y_0 = f(x_0)$ und $\varepsilon > 0$ gegeben. Nach Voraussetzung ist dann $f^{-1}(B_\varepsilon(y_0))$ offen, das heißt es gibt ein $\delta > 0$ mit $B_\delta(x_0) \subset f^{-1}(B_\varepsilon(y_0))$ beziehungsweise $f(B_\delta(x_0)) \subset B_\varepsilon(y_0)$. □

Im Gegensatz zur Definition 1.10 der Stetigkeit, die in jedem einzelnen Punkt des Definitionsbereichs überprüft werden kann, bezieht sich die hier gegebene Eigenschaft auf die Funktion als Ganzes, sie ist nicht lokal.

Aus der Analysis 1 wissen wir, dass noch eine dritte topologische Eigenschaft von Mengen von großer Bedeutung ist.

Definition 1.12 (Folgenkompaktheit) *Eine Teilmenge M eines metrischen Raums X heißt (folgen-)kompakt, wenn gilt: jede Folge $(x_k)_{k\in\mathbb{N}}$ mit $x_k \in M$ hat eine Teilfolge $(x_{k_p})_{p\in\mathbb{N}}$, die gegen ein $x \in M$ konvergiert.*

Wir werden eine alternative Charakterisierung der Kompaktheit mittels Überdeckungen bei Gelegenheit kennenlernen. Zunächst beziehen wir uns aber stets auf Definition 1.12. Um die Kompaktheit einer Teilmenge des $\mathbb{R}^n$ festzustellen, ist das folgende Kriterium sehr nützlich, das in Kapitel 4, Satz 2.2, gezeigt wurde.

Satz 1.5 (Kompaktheit im $\mathbb{R}^n$) *Eine Menge $M \subset \mathbb{R}^n$ ist genau dann kompakt, wenn sie abgeschlossen und beschränkt ist.*

Diese Aussage ist in vielen metrischen Räumen falsch, das heißt es kann abgeschlossene und beschränkte Mengen geben, die nicht kompakt sind. Folgende Aussagen über Stetigkeit und kompakte Mengen sind oft nützlich.

Satz 1.6 (Bilder kompakter Mengen) *Seien X, Y metrische Räume, D eine Teilmenge von X und $f : D \to Y$ stetig. Ist $K \subset D$ kompakt, so gilt*

(1) *$f(K)$ ist kompakte Teilmenge von Y.*

(2) *Ist f injektiv, so ist $f^{-1} : f(K) \to X$ stetig.*

Beweis: Sei (y_k) eine Folge in $M = f(K)$, also $y_k = f(x_k)$ für $x_k \in K$. Da K kompakt, gibt es eine Teilfolge mit $x_{k_j} \to x \in K$. Aus der Stetigkeit von f folgt $y_{k_j} = f(x_{k_j}) \to f(x) \in M$.

Jetzt nehmen wir indirekt an, f^{-1} sei in $y = f(x)$ unstetig. Dann gibt es eine Folge $y_k = f(x_k)$ mit $y_k \to y$, aber $d_X(x_k, x) \geq \varepsilon$ für alle k. Da K kompakt, gibt es eine Teilfolge $x_{k_j} \to x' \in K$ und es gilt $d(x', x) \geq \varepsilon$. Aber f ist stetig in x', also $f(x') = \lim_{j\to\infty} f(x_{k_j}) = y$, im Widerspruch zur Injektivität von f. □

Satz 1.7 (Extrema) *Eine stetige Funktion auf einer kompakten Teilmenge $K \neq \emptyset$ eines metrischen Raums X ist beschränkt und nimmt ihr Infimum und Supremum an.*

Beweis: vgl. Kapitel 4, Satz 2.1 □

Beispiel 1.5 Sei $K \subset X$ kompakt. Dann gibt es zu jedem $x_0 \in X$ einen nächsten Punkt $x \in K$, das heißt

$$d(x, x_0) = \inf_{y\in K} d(y, x_0) = \operatorname{dist}(x_0, K).$$

Der Punkt x ist nicht notwendig eindeutig, betrachte etwa $K = \{1, -1\} \subset \mathbb{R}$ und $x_0 = 0$.

Satz 1.8 (Gleichmäßige Stetigkeit) *Sei K kompakte Teilmenge eines metrischen Raums X. Dann ist jede stetige Abbildung $f : K \to Y$ gleichmäßig stetig.*

BEWEIS: vgl. Kapitel 5, Satz 1.4. Wäre f nicht gleichmäßig stetig, so gibt es ein $\varepsilon > 0$ und $x_n, x_n' \in K$ mit $d(x_n, x_n') \to 0$, aber $d(f(x_n), f(x_n')) \geq \varepsilon$. Da K kompakt, konvergiert die Folge x_n nach evtl. Auswahl einer Teilfolge gegen ein $x \in K$. Wegen $d(x_n', x) \leq d(x_n', x_n) + d(x_n, x)$ konvergiert dann auch die Folge x_n' gegen x, und es gilt aufgrund der Stetigkeit

$$\varepsilon \leq d(f(x_n), f(x_n')) \leq d(f(x_n), f(x)) + d(f(x), f(x_n')) \to 0,$$

ein Widerspruch. $\square$

2 Partielle Ableitungen

Wir kommen nun zur Differentiation von Funktionen im $\mathbb{R}^n$. Um für diese Ableitungen zu definieren, ist die einfachste und vielfach beste Idee, alle Variablen bis auf x_j als konstant aufzufassen und die resultierende Funktion der einen Variablen x_j wie üblich zu differenzieren. Auf diese Weise ergeben sich für $j = 1, \ldots, n$ die n partiellen Ableitungen der Funktion. Im Folgenden bezeichnet $e_1, \ldots, e_n$ die Standardbasis des $\mathbb{R}^n$, also $e_j = (0, \ldots, 1, \ldots, 0)$ mit der 1 an der j-ten Stelle.

Definition 2.1 *Sei $\Omega \subset \mathbb{R}^n$ offen und $f : \Omega \to \mathbb{R}^m$. Die partielle Ableitung von f nach x_j an der Stelle $x \in \Omega$ ist der Grenzwert (falls existent)*

$$\partial_j f(x) = \lim_{t \to 0} \frac{f(x + te_j) - f(x)}{t}.$$

Andere Bezeichnungen sind $\dfrac{\partial f}{\partial x_j}(x)$ und $D_j f(x)$.

Die partielle Ableitung $\partial_j f(x)$ ist einfach die Ableitung der reellen Funktion $(-\delta, \delta) \to \mathbb{R}^m$, $t \mapsto f(x + te_j)$, an der Stelle $t = 0$, oder alternativ die Ableitung der reellen Funktion

$$(x_j - \delta, x_j + \delta) \to \mathbb{R}^m, \quad y \mapsto f(x_1, \ldots, x_{j-1}, y, x_{j+1}, \ldots, x_n)$$

an der Stelle $y = x_j$. Der folgende Satz formuliert in diesem Zusammenhang wohlbekannte Ergebnisse der eindimensionalen Analysis.

Satz 2.1 (Ableitungsregeln) *Sei $\Omega \subset \mathbb{R}^n$ offen und $x \in \Omega$. Die Existenz der partiellen Ableitungen $\partial_j f(x)$ und $\partial_j g(x)$ sei vorausgesetzt. Dann gelten folgende Aussagen:*

(a) Linearität: *für $f, g : \Omega \to \mathbb{R}^m$ und $\alpha, \beta \in \mathbb{R}$ gilt*

$$\partial_j(\alpha f + \beta g)(x) = \alpha\, \partial_j f(x) + \beta\, \partial_j g(x).$$

(b) Komponentenweise Differentiation: *für $f : \Omega \to \mathbb{R}^m$ gilt, wenn eine der Seiten existiert,*

$$\partial_j f(x) = \sum_{i=1}^{m} \partial_j f_i(x)\, e_i.$$

(c) Produktregel: *für* $f : \Omega \to \mathbb{R}^m$ *und* $g : \Omega \to \mathbb{R}$ *gilt*

$$\partial_j (fg)(x) = (\partial_j f)(x) g(x) + f(x)(\partial_j g)(x).$$

(d) Quotientenregel: *für* $f : \Omega \to \mathbb{R}^m$ *und* $g : \Omega \to \mathbb{R}$ *mit* $g(x) \neq 0$ *gilt*

$$\partial_j \left(\frac{f}{g}\right)(x) = \frac{(\partial_j f)(x) g(x) - f(x)(\partial_j g)(x)}{g(x)^2}.$$

(e) Kettenregel: *ist* $f : \Omega \to I \subset \mathbb{R}$ *und* $\varphi : I \to \mathbb{R}$ *differenzierbar in* $f(x)$, *so folgt*

$$\partial_j (\varphi \circ f)(x) = \varphi'(f(x)) \partial_j f(x).$$

Beispiel 2.1 Wir betrachten die Euklidische Abstandsfunktionen vom Nullpunkt

$$r : \mathbb{R}^n \to \mathbb{R},\ r(x) = |x| = \sqrt{x_1^2 + \ldots + x_n^2}.$$

In $x \neq 0$ existieren die partiellen Ableitungen, und zwar gilt mit der Kettenregel

$$\partial_j r(x) = \frac{2x_j}{2\sqrt{x_1^2 + \ldots + x_n^2}} = \frac{x_j}{r} \qquad \text{für } r = r(x) = |x|.$$

Im Nullpunkt sind die partiellen Ableitungen dagegen nicht definiert, denn $r(0 + te_i) = |t|$ ist in $t = 0$ nicht differenzierbar. Die Funktion $\partial_j r$ ist in $x \neq 0$ ihrerseits partiell differenzierbar, und wir erhalten die zweiten partiellen Ableitungen

$$\partial_i (\partial_j r)(x) = \frac{(\partial_i x_j) r - x_j \partial_i r}{r^2} = \frac{1}{r}\left(\delta_{ij} - \frac{x_i x_j}{r^2}\right).$$

Ist $\varphi : (0, \infty) \to \mathbb{R}$ zweimal differenzierbar, so berechnen wir weiter für $f = \varphi \circ r : \mathbb{R}^n \to \mathbb{R}$

$$\begin{aligned}
\partial_j f(x) &= \varphi'(r) \partial_j r = \varphi'(r) \frac{x_j}{r}, \\
\partial_i (\partial_j f)(x) &= \varphi''(r) \partial_i r \partial_j r + \varphi'(r) \partial_i (\partial_j r) = \varphi''(r) \frac{x_i x_j}{r^2} + \frac{\varphi'(r)}{r}\left(\delta_{ij} - \frac{x_i x_j}{r^2}\right).
\end{aligned}$$

Der Operator $\Delta = \sum_{i=1}^n \partial_i^2$ heißt Laplaceoperator, die Lösungen der Gleichung $\Delta f = 0$ heißen harmonische Funktionen. Wir können jetzt die rotationssymmetrischen harmonischen Funktionen ausrechnen, und zwar erhalten wir

$$0 \overset{!}{=} \Delta f(x) = \varphi''(r) + \frac{n-1}{r} \varphi'(r) = r^{1-n} \big(r^{n-1} \varphi'(r)\big)'.$$

Diese Gleichung hat die Lösungen, mit Integrationskonstanten $a, b \in \mathbb{R}$,

$$\varphi(r) = \begin{cases} a\,\dfrac{r^{2-n}}{2-n} + b & \text{für } n \geq 3 \\ a \log r + b & \text{für } n = 2. \end{cases}$$

Für $n = 3$ ist f das Newtonsche Gravitationspotential.

Im vorangegangenen Beispiel traten zweite partielle Ableitungen auf. Ist für $f : \Omega \to \mathbb{R}^m$ die Ableitungsfunktion $\partial_j f : \Omega \to \mathbb{R}^m$ definiert und ihrerseits in $x \in \Omega$ nach x_i partiell differenzierbar, so setzen wir

$$\partial_i \partial_j f(x) := \partial_i(\partial_j f)(x) \qquad \text{(alternative Notation } D^2_{ij} f(x) \text{ oder } \frac{\partial^2 f}{\partial x_i \partial x_j}(x)). \tag{2.1}$$

Entsprechend für Ableitungen beliebiger Ordnung: ist für $i_1, \ldots, i_k \in \{1, \ldots, n\}$ die Ableitungsfunktion $\partial_{i_1} \ldots \partial_{i_k} f : \Omega \to \mathbb{R}^n$ definiert und in $x \in \Omega$ nach x_i partiell differenzierbar, so setzen wir induktiv

$$\partial_i \partial_{i_1} \ldots \partial_{i_k} f(x) = \partial_i(\partial_{i_1} \ldots \partial_{i_k} f)(x). \tag{2.2}$$

Die folgenden Klassen von Funktionen spielen in der Analysis eine wichtige Rolle. Wir werden sehen, dass die Eigenschaft der Stetigkeit der partiellen Ableitungen in vielen Anwendungen wesentlich ist.

Definition 2.2 (C^k-Räume) *Sei $\Omega \subset \mathbb{R}^n$ offen und $k \in \mathbb{N}_0 \cup \{\infty\}$. Wir bezeichnen mit $C^k(\Omega, \mathbb{R}^m)$ die Menge aller k-mal stetig differenzierbaren Funktionen auf Ω mit Werten im $\mathbb{R}^m$, das heißt alle partiellen Ableitungen $\partial_{i_1} \ldots \partial_{i_j} f$ der Ordnung $j \leq k$ (bzw. $j < \infty$ im Fall $k = \infty$) sind definiert und stetig auf Ω. Im reellwertigen Fall, also $m = 1$, setzen wir $C^k(\Omega, \mathbb{R}) = C^k(\Omega)$.*

Wir wollen nun zeigen, dass die Operatoren ∂_i und ∂_j auf C^2-Funktionen vertauschen. Aus der Existenz der partiellen Ableitungen $\partial_i \partial_j f$ und $\partial_j \partial_i f$ allein folgt das nicht, wie das Beispiel

$$f(x,y) = \begin{cases} xy \dfrac{x^2 - y^2}{x^2 + y^2} & \text{für } (x,y) \neq (0,0) \\ 0 & \text{für } (x,y) = (0,0) \end{cases} \tag{2.3}$$

zeigt. Für diese Funktion ist $\partial_1 \partial_2 f(0,0) = 1$, aber $\partial_2 \partial_1 f(0,0) = -1$.

Satz 2.2 (Symmetrie der 2. Ableitung, H. A. Schwarz) *Sei $\Omega \subset \mathbb{R}^n$ offen. Ist $f \in C^2(\Omega)$, so vertauschen für $1 \leq i, j \leq n$ die Ableitungen nach x_i und x_j:*

$$\partial_i \partial_j f = \partial_j \partial_i f \qquad \text{auf } \Omega.$$

Beweis: Nach Definition ist $\partial_j f(x)$ Grenzwert der Differenzenquotienten

$$\Delta_j^t f(x) = \frac{f(x + te_j) - f(x)}{t} \qquad \text{für } t \to 0.$$

Für $\Delta_i^s(\Delta_j^t f)(x) = \frac{1}{s}\Big(\Delta_j^t f(x + se_i) - \Delta_j^t f(x)\Big)$ folgt

$$\lim_{s \to 0} \Big(\lim_{t \to 0} \Delta_i^s(\Delta_j^t f)(x) \Big) = \partial_i(\partial_j f)(x).$$

Das Problem besteht darin, die beiden Grenzwerte zu vertauschen. Der Differenzenquotient vertauscht immerhin mit der partiellen Ableitung:

$$\partial_i(\Delta_j^t f)(x) = \frac{1}{t}\big(\partial_i f(x + te_j) - \partial_i f(x)\big) = \Delta_j^t(\partial_i f)(x).$$

Wir verwenden nun den Mittelwertsatz der Differentialrechnung. Ist g nach x_i partiell differenzierbar, so gilt für ein $\alpha \in [0,1]$:

$$\Delta_i^s g(x) = \frac{g(x+se_i) - g(x)}{s} = \partial_i g(x + \alpha s e_i).$$

Wir wenden das an auf $g = \Delta_j^t f$ und auf $g = \partial_i f$, wobei $\alpha, \beta \in [0,1]$ von s, t abhängen:

$$\Delta_i^s(\Delta_j^t f)(x) = \partial_i(\Delta_j^t f)(x + \alpha s e_i) = \Delta_j^t(\partial_i f)(x + \alpha s e_i) = \partial_j(\partial_i f)(x + \alpha s e_i + \beta t e_j).$$

Da nach Voraussetzung $\partial_j \partial_i f$ stetig in x ist, folgt die Behauptung, indem wir erst $t \to 0$ und dann $s \to 0$ gehen lassen. □

Bemerkung. Tatsächlich haben wir gezeigt: existieren $\partial_i f$, $\partial_j f$, $\partial_j \partial_i f$ und ist $\partial_j \partial_i f$ stetig in x, so existiert auch $\partial_i \partial_j f(x)$ und es gilt $\partial_i \partial_j f(x) = \partial_j \partial_i f(x)$.

Folgerung 2.1 *Für eine Funktion $f \in C^k(\Omega, \mathbb{R}^m)$ vertauschen die partiellen Ableitungen bis zur Ordnung k, das heißt für jede Permutation $\sigma \in S_k$ gilt*

$$\partial_{i_{\sigma(1)}} \ldots \partial_{i_{\sigma(k)}} f = \partial_{i_1} \ldots \partial_{i_k} f.$$

Beweis: Nach Satz 2.2 können benachbarte Operatoren ∂_i, ∂_j vertauscht werden. Die symmetrische Gruppe wird durch Vertauschungen erzeugt (siehe Lineare Algebra). □

Der Begriff der partiellen Ableitung allein ist nicht geeignet, um die mehrdimensionale Differentialrechnung zu entwickeln. Ein entscheidender Mangel ist, dass aus der Existenz der partiellen Ableitungen $\partial_1 f, \ldots, \partial_n f$ in $x \in \Omega$ nicht die Stetigkeit von f im Punkt x folgt.

Beispiel 2.2 Sei $\Omega = \mathbb{R}^2$ und

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

Dann gilt $f(x,0) = 0 = f(0,y)$, insbesondere $\partial_1 f(0,0) = 0 = \partial_2 f(0,0)$. Aber für $c(t) = (t,t)$ gilt $f(c(t)) = 1/2$ für alle $t \neq 0$, das heißt f ist nicht stetig im Nullpunkt.

Insbesondere können wir im allgemeinen keine Kettenregel für $f \circ c$ formulieren, da die Funktion $f \circ c$ nicht einmal stetig sein muss. Die Definition der partiellen Ableitungen macht explizit von den Koordinaten auf $\mathbb{R}^n$ Gebrauch. Es wäre denkbar, dass sich ein besserer Ableitungsbegriff ergibt, wenn alle Richtungen gleichberechtigt betrachtet werden. Dies führt auf den Begriff der Richtungsableitung.

Definition 2.3 (Richtungsableitung) *Sei $\Omega \subset \mathbb{R}^n$ offen und $f : \Omega \to \mathbb{R}^m$. Für $x \in \Omega$ und $v \in \mathbb{R}^n$ heißt der Grenzwert (falls existent)*

$$\partial_v f(x) = \lim_{t \to 0} \frac{f(x+tv) - f(x)}{t}$$

Richtungsableitung von f an der Stelle x in Richtung v. Dies ist die gewöhnliche Ableitung der Funktion $t \mapsto f(x+tv)$ an der Stelle $t = 0$.

Beispiel 2.3 Die Richtungsableitung von $r(x) = |x|$ in $x \in \mathbb{R}^n \backslash \{0\}$ in Richtung $v \in \mathbb{R}^n$ ist

$$\partial_v r(x) = \frac{d}{dt} \sqrt{|x|^2 + 2t\langle x, v\rangle + |v|^2}\, |_{t=0} = \left\langle \frac{x}{|x|}, v \right\rangle.$$

Leider reicht aber selbst die Existenz aller Richtungsableitungen von f in $x \in \Omega$ nicht aus, damit f auch stetig im Punkt x ist.

Beispiel 2.4 Betrachte jetzt auf $\Omega = \mathbb{R}^2$ die Funktion

$$f(x,y) = \begin{cases} \dfrac{2xy^2}{x^2+y^4} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

Dann existieren im Punkt (0,0) alle Richtungsableitungen, denn für $v = (a,b) \neq (0,0)$ ist

$$\partial_v f(0,0) = \lim_{t \to 0} \frac{2ab^2}{a^2 + t^2 b^4} = \begin{cases} 2b^2/a & \text{für } a \neq 0 \\ 0 & \text{für } a = 0. \end{cases}$$

Dennoch ist f im Nullpunkt unstetig, denn für $c(t) = (t^2, t)$ gilt $f(c(t)) = 1$ für alle $t \neq 0$.

3 Die Ableitung

In Analysis 1 hatten wir gesehen, dass die Differenzierbarkeit einer Funktion f im Punkt x_0 gleichbedeutend mit der Existenz einer affinlinearen Funktion ist, die in x_0 mit f in erster Ordnung übereinstimmt. Dieser Ableitungsbegriff hat sich für Funktionen mehrerer Variabler als der richtige durchgesetzt.

Definition 3.1 (Ableitung) *Sei $\Omega \subset \mathbb{R}^n$ offen und $f : \Omega \to \mathbb{R}^m$. Die lineare Abbildung $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ heißt Ableitung von f im Punkt x_0, falls gilt:*

$$\lim_{x \to x_0} \frac{f(x) - \big(f(x_0) + A(x - x_0)\big)}{|x - x_0|} = 0. \tag{3.1}$$

f heißt dann differenzierbar in x_0 mit Ableitung $Df(x_0) = A$.

Mit der Substitution $h = x - x_0$ erhalten wir die äquivalente Fassung

$$\lim_{h \to 0} \frac{f(x_0 + h) - \big(f(x_0) + A\,h\big)}{|h|} = 0. \tag{3.2}$$

Satz 3.1 (Berechnung der Ableitung) *Die Funktion $f : \Omega \to \mathbb{R}^m$ sei in $x_0 \in \Omega$ differenzierbar. Dann besitzt f in x_0 die Richtungsableitungen*

$$\partial_v f(x_0) = Df(x_0)v \quad \textit{für alle } v \in \mathbb{R}^n, \tag{3.3}$$

und $Df(x_0)$ hat bezüglich der Standardbasen die Matrixdarstellung (Jacobimatrix)

$$(\partial_j f_i(x_0)) = \begin{pmatrix} \partial_1 f_1(x_0) & \ldots & \ldots & \partial_n f_1(x_0) \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \partial_1 f_m(x_0) & \ldots & \ldots & \partial_n f_m(x_0) \end{pmatrix} \in \mathbb{R}^{m \times n}. \tag{3.4}$$

Insbesondere ist die Ableitung durch (3.1) eindeutig bestimmt.

Beweis: Für $v = 0$ sind beide Seiten von (3.3) nach Definition gleich Null. Für $v \neq 0$ setzen wir $Df(x_0) = A$, und erhalten für $t \to 0$ aus (3.1)

$$\begin{aligned}\left|\frac{f(x_0+tv)-f(x_0)}{t} - Av\right| &= \frac{\big|f(x_0+tv) - \big(f(x_0)+A(tv)\big)\big|}{|t|} \\ &= \frac{\big|f(x_0+tv) - \big(f(x_0)+A(tv)\big)\big|}{|tv|}\,|v| \to 0.\end{aligned}$$

Setzen wir $v = e_j$ und berechnen die partielle Ableitung komponentenweise, siehe Satz 2.1, so ergibt sich

$$Df(x_0)e_j = \partial_j f(x_0) = \sum_{i=1}^m \partial_j f_i(x_0) e_i.$$

□

Im allgemeinen ist der Nachweis der Differenzierbarkeit direkt anhand der Definition 3.1 nicht so einfach. Hier ein paar Beispiele.

Beispiel 3.1 Für eine symmetrische Bilinearform $b : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ betrachten wir die zugehörige quadratische Form $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = b(x,x)$. Wir behaupten, dass f in $x_0 \in \mathbb{R}^n$ differenzierbar ist mit Ableitung

$$Df(x_0) : \mathbb{R}^n \to \mathbb{R},\ Df(x_0)v = 2\,b(v,x_0) \qquad \text{für alle } v \in \mathbb{R}^n.$$

Die rechte Seite definiert eine lineare Abbildung von $\mathbb{R}^n$ nach $\mathbb{R}$. Wir berechnen

$$f(x_0+h) = b(x_0+h, x_0+h) = f(x_0) + 2\,b(h,x_0) + b(h,h).$$

Nun gilt die Abschätzung

$$|b(h,h)| \leq \sum_{i,j=1}^n |b(e_i,e_j)|\,|h_i|\,|h_j| \leq C|h|^2 \quad \text{ mit } C = \sum_{i,j=1}^n |b(e_i,e_j)|,$$

und es folgt

$$\frac{f(x_0+h) - \big(f(x_0) + 2b(h,x_0)\big)}{|h|} = \frac{b(h,h)}{|h|} \to 0 \quad \text{ mit } h \to 0,$$

was zu zeigen war.

Beispiel 3.2 Die Funktion $f : I = (a,b) \to \mathbb{R}^m$ habe in $x_0 \in I$ die Ableitung $f'(x_0) \in \mathbb{R}^m$ im Sinne von Analysis 1. Dann ist f differenzierbar in x_0 im Sinne von Definition 3.1 mit

$$Df(x_0) : \mathbb{R} \to \mathbb{R}^m, \quad Df(x_0)h = f'(x_0)h.$$

Denn es gilt für $h \neq 0$

$$\frac{\big|f(x_0+h) - \big(f(x_0) + f'(x_0)h\big)\big|}{|h|} = \left|\frac{f(x_0+h)-f(x_0)}{h} - f'(x_0)\right| \to 0 \quad \text{mit } h \to 0.$$

Beispiel 3.3 Sei $\Omega \subset \mathbb{R}^n$ offen und $A \in L(\mathbb{R}^n, \mathbb{R}^m)$. Dann ist die Einschränkung

$$f : \Omega \to \mathbb{R}^m,\ f(x) = Ax \qquad \text{für alle } x \in \Omega,$$

in allen $x_0 \in \Omega$ differenzierbar mit Ableitung $Df(x_0) = A$. Dies folgt sofort wegen $f(x_0+h) = A(x_0 + h) = Ax_0 + Ah = f(x_0) + Ah$.

Definition 3.2 (Gradient) *Sei $\Omega \subset \mathbb{R}^n$ offen und $f : \Omega \to \mathbb{R}$ differenzierbar in $x \in \Omega$. Der Gradient von f im Punkt x ist der Vektor*

$$\operatorname{grad} f(x) = \sum_{j=1}^{n} \partial_j f(x) e_j = \begin{pmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{pmatrix} \in \mathbb{R}^n.$$

Der Gradient ist der eindeutig bestimmte Vektor mit der Eigenschaft

$$\langle \operatorname{grad} f(x), v \rangle = Df(x)v \qquad \text{für alle } v \in \mathbb{R}^n. \tag{3.5}$$

Dabei ist $\langle \cdot, \cdot \rangle$ das Standardskalarprodukt. Ist $\operatorname{grad} f(x) = 0$, so heißt x kritischer Punkt von f. Ist x nicht kritisch, so ist die Richtung von $\operatorname{grad} f(x)$ diejenige, in der f am stärksten ansteigt. Denn für $v \in \mathbb{R}^n$ mit $|v| = 1$ folgt aus der Ungleichung von Cauchy-Schwarz

$$\partial_v f(x) = \langle \operatorname{grad} f(x), v \rangle \leq |\operatorname{grad} f(x)|,\ \text{Gleichheit genau wenn } v = \frac{\operatorname{grad} f(x)}{|\operatorname{grad} f(x)|}. \tag{3.6}$$

Beispiel 3.4 Der Gradient der Funktion $f(x) = \varphi(r)$ mit $r(x) = |x|$ ist nach Beispiel 2.1

$$\operatorname{grad} f(x) = \varphi'(r)\frac{x}{r} \qquad \text{für } x \neq 0.$$

Beispiel 3.5 Sei $b : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ eine symmetrische Bilinearform, und $B \in L(\mathbb{R}^n, \mathbb{R}^n)$ der zugehörige Endomorphismus bzgl. des Standardskalarprodukts, genauer

$$b(x, y) = \langle x, By \rangle \qquad \text{für alle } x, y \in \mathbb{R}^n.$$

Bezüglich der Standardbasis gilt $B_{ij} = b(e_i, e_j)$. Für $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = b(x, x)$, folgt

$$\operatorname{grad} f(x) = 2Bx \qquad \text{für alle } x \in \mathbb{R}^n.$$

Denn nach Beispiel 3.1 gilt für alle $v \in \mathbb{R}^n$

$$\langle \operatorname{grad} f(x), v \rangle = Df(x)v = 2b(v, x) = 2\langle v, Bx \rangle.$$

Man kann sich eine reellwertige Funktion f auf $\Omega \subset \mathbb{R}^n$ stets als Höhenfunktion einer Landschaft über der Grundfläche Ω vorstellen. Dazu betrachtet man ihren Graph

$$G = \{(y, f(y)) : y \in \Omega\} \subset \Omega \times \mathbb{R} \subset \mathbb{R}^{n+1}.$$

Aus der Differenzierbarkeit von f im Punkt x folgt, dass der Graph im Punkt $p = (x, f(x))$ eine Tangentialebene besitzt. Um dies zu sehen, betrachten wir für $\lambda > 0$ die Mengen $G_{p,\lambda} = \frac{1}{\lambda}(G - p)$, das heißt G wird um $-p$ verschoben – dabei landet p im Nullpunkt – und dann mit dem Faktor $\frac{1}{\lambda}$ gestreckt. Wir erwarten, dass die $G_{p,\lambda}$ für $\lambda \to 0$ gegen die

Tangentialhyperebene konvergieren. Da G Graph über Ω ist, ist $G_{p,\lambda}$ Graph über der Menge $\Omega_{x,\lambda} = \frac{1}{\lambda}(\Omega - x)$, und zwar gilt

$$G_{p,\lambda} = \Big\{\Big(\frac{y-x}{\lambda}, \frac{f(y)-f(x)}{\lambda}\Big) : y \in \Omega\Big\} = \Big\{\Big(z, \frac{f(x+\lambda z)-f(x)}{\lambda}\Big) : z \in \Omega_{x,\lambda}\}.$$

Für $z \in \mathbb{R}^n$ ist $x + \lambda z \in \Omega$ und folglich $z \in \Omega_{x,\lambda}$ für $\lambda > 0$ hinreichend klein. Aus (3.3) folgt nun für die Graphenfunktionen der $G_{p,\lambda}$

$$\lim_{\lambda \searrow 0} \frac{f(x+\lambda z)-f(x)}{\lambda} = Df(x)z.$$

Es ist damit gerechtfertigt, die Menge $T_pG = \{(z, Df(x)z) : z \in \mathbb{R}^n\}$ als Tangentialraum des Graphen im Punkt $(x, f(x))$ zu definieren. Als Bild der linearen Abbildung $z \mapsto (z, Df(x)z)$ ist T_pG ein linearer Unterraum mit der Basis $(e_1, \partial_1 f(x)), \ldots, (e_n, \partial_n f(x))$. Es ist leicht zu sehen, dass der Vektor

$$\nu = \frac{(-\operatorname{grad} f(x), 1)}{\sqrt{1 + |\operatorname{grad} f(x)|^2}}$$

senkrecht auf den Tangentialraum steht und Länge Eins hat.

Im Beweis der Differentiationsregeln brauchen wir folgende Abschätzung für lineare Abbildungen $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, vgl. Beispiel 1.13 in Kapitel 3:

$$(3.7) \qquad |A(x)| \leq |A|\,|x| \quad \text{für alle } x \in \mathbb{R}^n.$$

Zum Beweis verwenden wir die Matrixdarstellung bezüglich der Standardbasen

$$A(x) = \sum_{j=1}^n x_j A(e_j) \quad \text{mit } A(e_j) = \sum_{i=1}^m a_{ij} e_i.$$

Aus der Dreiecksungleichung in $\mathbb{R}^m$ und der Ungleichung von Cauchy-Schwarz folgt

$$|A(x)|^2 \leq \Big(\sum_{j=1}^n |x_j|\,|A(e_j)|\Big)^2 \leq \Big(\sum_{j=1}^n x_j^2\Big)\Big(\sum_{j=1}^n |A(e_j)|^2\Big) = |A|^2\,|x|^2.$$

Hier wurde $\sum_{j=1}^n |A(e_j)|^2 = \sum_{j=1}^n \sum_{i=1}^m a_{ij}^2 = |A|^2$ benutzt. Als Folgerung ergibt sich, dass jede lineare Abbildung $A : \mathbb{R}^n \to \mathbb{R}^m$ Lipschitzstetig ist, genauer

$$|A(x) - A(y)| = |A(x-y)| \leq |A|\,|x-y|.$$

Wir bemerken, dass in Abschätzung (3.7) die Euklidische Norm $|A|$ durch die im allgemeinen kleinere Operatornorm $\|A\| = \sup_{|x|=1} |A(x)|$ ersetzt werden kann; das ist offenbar die optimale Konstante in (3.7). Da die Operatornorm schwieriger auszurechnen ist, arbeiten wir jedoch in der Regel mit der Euklidischen Norm.

Die Abschätzung (3.7) ist im allgemeinen nicht gültig, wenn $\mathbb{R}^n$ durch einen unendlichdimenionalen normierten Raum ersetzt wird, und lineare Abbildungen sind dann nicht automatisch stetig.

Satz 3.2 (Differenzierbarkeit $\Rightarrow$ Stetigkeit) *Sei $\Omega \subset \mathbb{R}^n$. Ist $f : \Omega \to \mathbb{R}^m$ differenzierbar in x_0, so ist f stetig in x_0.*

Beweis: Wie soeben besprochen, sind affinlineare Funktionen stetig auf $\mathbb{R}^n$. Es reicht daher zu zeigen, dass die Funktion $\varphi(x) = f(x) - \big(f(x_0) + Df(x_0)(x - x_0)\big)$ stetig in x_0 ist. Aber $\varphi(x_0) = 0$, und nach Definition der Differenzierbarkeit gilt

$$\varphi(x) = |x - x_0| \, \frac{\varphi(x)}{|x - x_0|} \to 0 \quad \text{mit } x \to x_0.$$

$\square$

Wir müssen jetzt die Differentiationsregeln erarbeiten.

Satz 3.3 (Kettenregel) *Seien $f : U \to \mathbb{R}^m$ und $g : V \to \mathbb{R}^p$ mit $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ offen und $f(U) \subset V$. Sind f in x_0 und g in $f(x_0)$ differenzierbar, so ist auch $g \circ f$ in x_0 differenzierbar und es gilt die Kettenregel*

$$D(g \circ f)(x_0) = Dg\big(f(x_0)\big)Df(x_0).$$

Für die zugehörigen Jacobimatrizen bedeutet das mit $y_0 = f(x_0)$

$$\frac{\partial (g \circ f)_i}{\partial x_k}(x_0) = \sum_{j=1}^{m} \frac{\partial g_i}{\partial y_j}(y_0) \frac{\partial f_j}{\partial x_k}(x_0) \quad \textit{für } 1 \leq i \leq p,\ 1 \leq k \leq n.$$

Beweis: Sei $y_0 = f(x_0)$, $Df(x_0) = A$, $Dg(y_0) = B$. Wir definieren für hinreichend kleine $\xi \in \mathbb{R}^n \backslash \{0\}$, $\eta \in \mathbb{R}^m \backslash \{0\}$ die Funktionen

$$\varepsilon_f(\xi) = \frac{f(x_0 + \xi) - \big(f(x_0) + A\xi\big)}{|\xi|} \quad \text{und} \quad \varepsilon_g(\eta) = \frac{g(y_0 + \eta) - \big(g(y_0) + B\eta\big)}{|\eta|}.$$

Mit $\varepsilon_f(0) = 0$ und $\varepsilon_g(0) = 0$ sind beide Funktionen nach Voraussetzung im Nullpunkt stetig. Offensichtliche Kandidatin für die Ableitung von $g \circ f$ in x_0 ist BA, also berechnen wir

$$\begin{aligned}
& \frac{(g \circ f)(x_0 + \xi) - \big((g \circ f)(x_0) + BA\xi\big)}{|\xi|} \\
= \; & \frac{g\big(y_0 + A\xi + |\xi|, \varepsilon_f(\xi)\big) - \big(g(y_0) + BA\xi\big)}{|\xi|} \\
= \; & \frac{g(y_0) + B\eta + |\eta| \, \varepsilon_g(\eta) - \big(g(y_0) + BA\xi\big)}{|\xi|} \qquad \text{wobei } \eta = A\xi + |\xi| \, \varepsilon_f(\xi) \\
= \; & B\varepsilon_f(\xi) + \frac{|\eta|}{|\xi|} \, \varepsilon_g(\eta).
\end{aligned}$$

Wegen $|B\varepsilon_f(\xi)| \leq |B||\varepsilon_f(\xi)|$ und $|\eta| \leq (|A| + |\varepsilon_f(\xi)|)|\xi| \leq C|\xi|$ konvergiert die rechte Seite wie gewünscht gegen Null. $\square$

Beispiel 3.6 Spezialfall ist die Verkettung $f \circ c$ einer Kurve $c : (a,b) \to \Omega \subset \mathbb{R}^n$ und einer Funktion $f : \Omega \to \mathbb{R}$. Ist c differenzierbar in $t \in (a,b)$ und f differenzierbar in $c(t)$, so folgt

$$\frac{d(f \circ c)}{dt}(t) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(c(t)) \frac{dc_j}{dt}(t),$$

beziehungsweise in vektorieller Form

$$(f \circ c)'(t) = Df(x)c'(t) = \langle \operatorname{grad} f(x), c'(t) \rangle \qquad \text{wobei } x = c(t).$$

Ist $f \circ c$ konstant, so folgt $\operatorname{grad} f(x) \perp c'(t)$. Anschaulich bedeutet das, der Gradient von f steht senkrecht auf die Niveaumengen $\{x \in \Omega : f(x) = \text{const.}\}$. Im zweidimensionalen Fall kann man sich die Niveaumengen als Höhenlinien vorstellen.

Weitere Regeln für den Umgang mit der Ableitung sind die folgenden.

Satz 3.4 (Ableitungsregeln) *Sei $\Omega \subset \mathbb{R}^n$ offen und $x \in \Omega$. Die Existenz der Ableitungen $Df(x)$ und $Dg(x)$ sei jeweils vorausgesetzt. Dann gelten folgende Aussagen:*

(a) Linearität: *für $f, g : \Omega \to \mathbb{R}^m$ und $\alpha, \beta \in \mathbb{R}$ gilt*

$$D(\alpha f + \beta g)(x) = \alpha Df(x) + \beta Dg(x).$$

(b) Produktregel: *für $f : \Omega \to \mathbb{R}^m$ und $g : \Omega \to \mathbb{R}$ gilt*

$$D(fg)(x) = Df(x)g(x) + f(x)Dg(x).$$

(c) Quotientenregel: *für $f : \Omega \to \mathbb{R}^m$ und $g : \Omega \to \mathbb{R}$ mit $g(x) \neq 0$ gilt*

$$D\left(\frac{f}{g}\right)(x) = \frac{Df(x)g(x) - f(x)Dg(x)}{g(x)^2}.$$

Beweis: Wir setzen $Df(x) = A$, $Dg(x) = B$ sowie für $h \neq 0$

$$\varepsilon_f(h) = \frac{f(x+h) - \big(f(x) + Ah\big)}{|h|} \qquad \text{und} \qquad \varepsilon_g(h) = \frac{g(x+h) - \big(g(x) + Bh\big)}{|h|}.$$

Nach Voraussetzung gilt $\varepsilon_f(h) \to 0$, $\varepsilon_g(h) \to 0$ mit $h \to 0$. Mit der jeweils behaupteten Ableitung ist nun für $h \to 0$ der Grenzwert in (3.2) nachzuprüfen. Für (a) gilt

$$\frac{(\alpha f + \beta g)(x+h) - \big((\alpha f + \beta g)(x) + (\alpha A + \beta B)h\big)}{|h|} = \alpha\varepsilon_f(h) + \beta\varepsilon_g(h) \to 0.$$

Für (b) berechnen wir mit etwas mehr Mühe

$$\begin{aligned}
& \frac{(fg)(x+h) - \big((fg)(x) + (Ag(x) + f(x)B)\,h\big)}{|h|} \\
= \; & \frac{\big(f(x) + Ah + \varepsilon_f(h)|h|\big)\big(g(x) + Bh + \varepsilon_g(h)|h|\big) - \big(f(x)g(x) + g(x)Ah + f(x)Bh\big)}{|h|} \\
= \; & \frac{1}{|h|}(Ah)(Bh) + \varepsilon_f(h)\,(g(x) + Bh + \varepsilon_g(h)|h|) + \varepsilon_g(|h|)(f(x) + Ah).
\end{aligned}$$

Wie in (3.7) bemerkt gilt $|Ah| \leq |A||h|$ sowie $|Bh| \leq |B||h|$, also geht die rechte Seite mit $h \to 0$ gegen Null. In (c) können wir $m = 1$ und $f \equiv 1$ annehmen, denn sonst schreiben wir

$f/g = f(1/g)$ und verwenden (b). Es gilt

$$
\begin{aligned}
& \frac{1}{|h|} \left(\frac{1}{g(x+h)} - \Big(\frac{1}{g(x)} - \frac{Bh}{g(x)^2} \Big) \right) \\
= \;& \frac{1}{|h|} \frac{1}{g(x)g(x+h)} \Big(g(x) - \big(g(x) + Bh + \varepsilon_g(h)|h| \big) + \frac{g(x+h)}{g(x)} Bh \Big) \\
= \;& \frac{1}{g(x)g(x+h)} \left(\Big(\frac{g(x+h)}{g(x)} - 1 \Big) \frac{Bh}{|h|} - \varepsilon_g(h) \right).
\end{aligned}
$$

Wegen $g(x) \neq 0$ und $g(x+h) \to g(x)$ mit $h \to 0$ nach Satz 3.2 geht die rechte Seite wieder gegen Null mit $h \to 0$. □

Die koordinatenfreie Notation ermöglicht oft ein gutes geometrisches Verständnis für die Ableitung als lineare Abbildung. Allerdings ist beim Umgang mit Zeilen- und Spaltenvektoren Vorsicht geboten, zum Beispiel kann bei der Produktregel Verwirrung entstehen, wenn eine der beteiligten Funktionen vektorwertig sind. Erst recht wird die Notation kompliziert, wenn zweite oder höhere Ableitungen zu bilden sind. Im Zweifelsfall sollte man auf die partiellen Ableitungen zurückgreifen.

Satz 3.5 (komponentenweise Differentiation) *$f : \Omega \to \mathbb{R}^m$ ist genau dann in $x_0 \in \Omega$ differenzierbar, wenn alle Komponenten $f_i : \Omega \to \mathbb{R}$, $i = 1, \ldots, m$, in x_0 differenzierbar sind. Ist $P_i : \mathbb{R}^n \to \mathbb{R}$ Projektion auf die i-te Koordinate, so gilt $Df_i(x_0) = P_i\, Df(x_0)$.*

Beweis: Es gilt nach Definition

$$
Df(x_0) = A \quad \Leftrightarrow \quad \lim_{x \to x_0} \frac{f(x) - \big(f(x_0) + A(x - x_0) \big)}{|x - x_0|} = 0.
$$

Die Konvergenz im $\mathbb{R}^n$ ist gleichbedeutend mit der Konvergenz aller Komponenten. Durch Anwendung von P_i ergibt sich daher weiter die äquivalente Formulierung

$$
\lim_{x \to x_0} \frac{f_i(x) - \big(f_i(x_0) + P_i A(x - x_0) \big)}{|x - x_0|} = 0 \quad \text{für alle } i = 1, \ldots, m. \tag{3.8}
$$

Aus $Df(x_0) = A$ folgt somit $Df_i(x_0) = P_i A$. Ist umgekehrt $Df_i(x_0) = A_i$ für $i = 1, \ldots, m$, so definieren wir $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ durch $A(x) = \sum_{i=1}^m A_i(x) e_i$. Dann ist $P_i A = A_i$, also gilt (3.8) und somit $Df(x_0) = A$. □

Wie besprochen kann aus der Existenz der partiellen Ableitungen nicht auf die Differenzierbarkeit geschlossen werden, ja nicht einmal auf die Stetigkeit. Das ist schade, denn die partiellen Ableitungen sind so schön einfach auszurechnen, während der Nachweis der Differenzierbarkeit anhand der Definition 3.1 etwas Aufwand erfordert. Der folgende Satz liefert ein zentrales, hinreichendes Kriterium für die Differenzierbarkeit einer Funktion.

Satz 3.6 (stetig partiell differenzierbar $\Rightarrow$ differenzierbar) *Sei $\Omega \subset \mathbb{R}^n$ offen. Die Funktion $f : \Omega \to \mathbb{R}^m$ sei in allen $x \in \Omega$ nach $x_1, \ldots, x_n$ partiell differenzierbar. Sind die Ableitungen $\partial_i f : \Omega \to \mathbb{R}^m$ in $x_0 \in \Omega$ stetig, so ist f in x_0 differenzierbar.*

Beweis: Wegen Satz 3.5 können wir $m = 1$ annehmen. Mit Satz 3.1 kennen wir bereits die einzig mögliche Kandidatin für die Ableitung, und zwar

$$
A : \mathbb{R}^n \to \mathbb{R}, \quad Ah = \sum_{k=1}^n \partial_k f(x_0) h_k.
$$

Zu $\varepsilon > 0$ gibt es ein $\delta > 0$ mit $B_\delta(x_0) \subset \Omega$, so dass

$$|\partial_k f(x) - \partial_k f(x_0)| < \varepsilon \quad \text{für alle } x \in B_\delta(x_0).$$

Für $|h| < \delta$ betrachten wir die Punkte $x_k = x_0 + \sum_{i=1}^k h_i e_i$ mit $1 \leq k \leq n$, und erhalten aus dem Mittelwertsatz der Differentialrechnung

$$f(x_k) - f(x_{k-1}) = f(x_{k-1} + h_k e_k) - f(x_{k-1}) = \partial_k f(x_{k-1} + s_k h_k e_k) h_k$$

für geeignete $s_k \in [0,1]$. Es folgt, da $|x_{k-1} + s_k h_k e_k - x_0| \leq |h| < \delta$,

$$\begin{aligned}
\frac{\big|f(x_0+h) - \big(f(x_0) + Ah\big)\big|}{|h|} &= \frac{1}{|h|} \Big| \sum_{k=1}^n \big(f(x_k) - f(x_{k-1}) - \partial_k f(x_0) h_k\big) \Big| \\
&= \frac{1}{|h|} \Big| \sum_{k=1}^n \Big(\partial_k f(x_{k-1} + s_k h_k e_k) - \partial_k f(x_0)\Big) h_k \Big| \\
&\leq \sum_{k=1}^n \Big| \partial_k f(x_{k-1} + s_k h_k e_k) - \partial_k f(x_0) \Big| < n\varepsilon.
\end{aligned}$$

□

Für das Rechnen mit C^k-Funktionen gelten die folgenden Regeln. Besonders angenehm ist der Raum $C^\infty(\Omega)$, der sogar unter Differentiation abgeschlossen ist.

Folgerung 3.1 *Sei $k \in \mathbb{N}_0 \cup \{\infty\}$.*

(a) *Mit $f, g \in C^k(\Omega, \mathbb{R}^m)$ gilt $\alpha f + \beta g \in C^k(\Omega, \mathbb{R}^m)$ für $\alpha, \beta \in \mathbb{R}$.*

(b) *Aus $f, g \in C^k(\Omega)$ folgt $fg \in C^k(\Omega)$, sowie $f/g \in C^k(\Omega)$ falls $g \neq 0$ auf Ω.*

(c) *Sind $f \in C^k(U, \mathbb{R}^m)$, $g \in C^k(V, \mathbb{R}^p)$ mit $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ offen und $f(U) \subset V$, so ist $g \circ f \in C^k(U, \mathbb{R}^p)$.*

Beweis: Im Fall $k = 0$ sind die Aussagen wohlbekannt. Die Behauptungen (a) und (b) folgen nun aus den Rechenregeln für die partielle Ableitung, siehe Satz 2.1, mit Induktion über k. Sind zum Beispiel $f, g \in C^k(\Omega)$ für ein $k \geq 1$, so gilt induktiv $\partial_j(fg) = (\partial_j f)g + f(\partial_j g) \in C^{k-1}(\Omega)$, also $fg \in C^k(\Omega)$.

Für $k \geq 1$ sind die Abbildungen f und g aus (c) differenzierbar nach Satz 3.6. Dann ist $g \circ f$ ebenfalls differenzierbar wegen der Kettenregel, Satz 3.3, mit partiellen Ableitungen $\partial_k (g \circ f)_i = \sum_{j=1}^m (\partial_j g_i) \circ f \, \partial_k f_j$. Nun ist $(\partial_j g_i) \circ f \in C^{k-1}(U)$ nach Induktionsannahme sowie $\partial_k f_j \in C^{k-1}(U)$ nach Voraussetzung, also ist $\partial_k (g \circ f)_i \in C^{k-1}(U)$ mit der Produktregel aus (b) und folglich $g \circ f$ von der Klasse C^k. □