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*

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

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

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

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

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

$\square$

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

Beispiel 3.1 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 3.2 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 $f(x_0+h) = A(x_0 + h) = Ax_0 + Ah = f(x_0) + Ah$.

Beispiel 3.3 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

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

und es folgt

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

was zu zeigen war.

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

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

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

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

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

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

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

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

Denn nach Beispiel 3.3 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

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

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:

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

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 $y_0 = f(x_0)$

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

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 $h \neq 0$

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

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

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

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 ⇒ 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

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

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. $\square$

Kapitel 7

Anwendungen der Differentialrechnung

1 Schrankensatz

Ein Grundproblem in der Analysis ist es, Informationen über die Ableitung in Eigenschaften der Funktion zu übersetzen. Für Funktionen einer Variablen, also $f : [a, b] \to \mathbb{R}$, stehen uns dazu zwei Argumente zur Verfügung:

a) der Mittelwertsatz (siehe Kapitel 4.2):

$$f(b) - f(a) = f'(\tau)(b - a) \quad \text{für ein } \tau \in (a, b);$$

b) der Hauptsatz der Differential- und Integralrechnung (siehe Kapitel 5.2):

$$f(b) - f(a) = \int_a^b f'(t)\, dt.$$

Im Vergleich ist der Mittelwertsatz insofern stärker, als er mit minimalen Voraussetzungen auskommt – f muss differenzierbar in (a, b) und stetig in den Endpunkten a,b sein. Dagegen verlangt unsere Version des Hauptsatzes, dass f stetig differenzierbar auf $[a, b]$ ist. Tasächlich reicht es, wenn f stetig auf $[a, b]$ und stückweise C^1 ist, das heißt es gibt eine Zerlegung $a = t_0 < \ldots < t_N = b$, so dass $f\ |_{[t_{k-1},t_k]}$ stetig differenzierbar ist für $k = 1, \ldots, N$. Wir können dann nämlich den Hauptsatz auf jedem Teilintervall anwenden und erhalten

$$f(b) - f(a) = \sum_{k=1}^{N} \big(f(t_k) - f(t_{k-1})\big) = \sum_{k=1}^{N} \int_{t_{k-1}}^{t_k} f'(t)\, dt = \int_a^b f'(t)\, dt,$$

wobei f' in jedem der Punkte t_k einen links- und rechtsseitigen Grenzwert hat, die nicht notwendig gleich sind. Ein Nachteil des Mittelwertsatzes ist, dass er für vektorwertige Funktionen in der Regel nicht gilt, zum Beispiel ist für $f(t) = e^{it}$

$$f(2\pi) - f(0) = 0 \quad \text{aber} \quad f'(\tau) = ie^{i\tau} \neq 0 \text{ für alle } \tau \in (0, 2\pi).$$

Der Mittelwertsatz kann natürlich auf jede Komponente einzeln angewandt werden, aber die Zwischenstellen werden in der Regel verschieden sein, und dass ist unpraktisch. Im Folgenden arbeiten wir deshalb mit dem Hauptsatz, und nehmen die stärkere Voraussetzung C^1 oder stückweise C^1 in Kauf. Es stellt sich heraus, dass die C^1-Bedingung für die Anwendungen richtig ist.

Wie können diese eindimensionalen Werkzeuge nun für Funktionen mehrerer Variabler $f : \Omega \to \mathbb{R}$ mit $\Omega \subset \mathbb{R}^n$ eingesetzt werden? Die einfache Antwort heißt: indem f längs Kurven $\gamma : [a,b] \to \Omega$, $\gamma = \gamma(t)$, ausgewertet wird.

Lemma 1.1 *Sei $\gamma : I = [a,b] \to \Omega$ stetig und stückweise C^1. Dann gilt für $f \in C^1(\Omega, \mathbb{R}^m)$*

$$f\big(\gamma(b)\big) - f\big(\gamma(a)\big) = \int_a^b Df\big(\gamma(t)\big)\gamma'(t)\,dt. \tag{1.1}$$

Beweis: Nach Folgerung 3.1 ist die Funktion $f \circ \gamma$ stückweise C^1, und es folgt aus dem Hauptsatz der Differential- und Integralrechnung und der Kettenregel

$$f\big(\gamma(b)\big) - f\big(\gamma(a)\big) = \int_a^b \frac{d}{dt} f\big(\gamma(t)\big)\,dt = \int_a^b Df\big(\gamma(t)\big)\gamma'(t)\,dt.$$

□

Im Beweis trat das Integral einer vektorwertigen Funktion auf. Dieses kann komponentenweise erklärt werden, das heißt für $v \in C^0([a,b], \mathbb{R}^m)$ ist

$$\int_a^b v(t)\,dt = \int_a^b \Big(\sum_{i=1}^m v_i(t) e_i\Big)\,dt = \sum_{i=1}^m \Big(\int_a^b v_i(t)\,dt\Big) e_i.$$

Alternativ kann man nachprüfen, dass die in Kapitel 5 gegebene Definition des Integrals mittels Riemannscher Summen ohne Änderungen auch für Funktionen mit Werten in $\mathbb{R}^m$ funktioniert. Man kann sich so oder so davon überzeugen, dass der Hauptsatz der Differential- und Integralrechnung ganz analog für vektorwertige Funktionen gilt.

Definition 1.1 *Ein metrischer Raum X heißt wegweise zusammenhängend, falls es zu je zwei Punkten x_0, $x_1 \in X$ eine stetige Abbildung $c : [0,1] \to X$ gibt mit $c(0) = x_0$, $c(1) = x_1$.*

Eine solche Abbildung nennt man auch einen Weg (Englisch: *path*) von x_0 nach x_1.

Lemma 1.2 *Sei $\Omega \subset \mathbb{R}^n$ offen und wegweise zusammenhängend. Dann gibt es einen stetigen, stückweise affinen Weg $\gamma : [0,1] \to \Omega$, d.h. einen Polygonzug, mit $\gamma(0) = x_0$, $\gamma(1) = x_1$.*

Beweis: Sei $c \in C^0([0,1], \Omega)$ mit $c(0) = x_0$ und $c(1) = x_1$ gegeben. Betrachte die Zerlegung $t_k = k/N$ für $k = 0, \ldots, N$ des Intervalls $[0,1]$, und definiere einen Polygonzug $\gamma : [0,1] \to \mathbb{R}^n$ mit $\gamma(t_k) = c(t_k)$ für alle k durch

$$\gamma(t) = \frac{t_k - t}{t_k - t_{k-1}}\, c(t_{k-1}) + \frac{t - t_{k-1}}{t_k - t_{k-1}}\, c(t_k) \qquad \text{für } t \in [t_{k-1}, t_k].$$

Wir wollen zeigen, dass dieser Polygonzug für N hinreichend groß ganz in Ω verläuft. Da $c([0,1])$ kompakt ist und $\mathbb{R}^n \backslash \Omega$ abgeschlossen, gibt es ein $\varrho > 0$ mit

$$|x - c(t)| \geq \varrho \text{ für alle } t \in [0,1],\ x \in \mathbb{R}^n \backslash \Omega.$$

Denn sonst gibt es Folgen $x_j \in c([0,1])$, $y_j \in \mathbb{R}^n \backslash \Omega$ mit $|x_j - y_j| \to 0$. Nach Übergang zu einer Teilfolge gilt $x_j \to x \in c([0,1])$, und weiter $y_j \to x$, also $x \in \mathbb{R}^n \backslash \Omega$, ein Widerspruch. Da c gleichmäßig stetig auf $[0,1]$ ist, gilt

$$\operatorname{osc}(c,\delta) = \sup\{|c(t)-c(s)| : 0 \leq s,t \leq 1,\ |t-s| \leq \delta\} \to 0 \quad \text{mit } \delta \to 0.$$

Daraus folgt, dass γ gleichmäßig gegen c konvergiert mit $N \to \infty$: für $t \in [0,1]$ wählen wir $k \in \{1,\dots,N\}$ mit $t \in [t_{k-1}, t_k]$ und schätzen wie folgt ab:

$$|\gamma(t)-c(t)| \leq \frac{t_k - t}{t_k - t_{k-1}}|c(t_{k-1})-c(t)| + \frac{t - t_{k-1}}{t_k - t_{k-1}}|c(t_k)-c(t)| \leq \operatorname{osc}(c, \frac{1}{N}) \to 0 \quad \text{mit } N \to \infty.$$

Für N groß ist also $|\gamma(t) - c(t)| < \varrho$ für alle $t \in [0,1]$, und damit $\gamma([0,1]) \subset \Omega$. □

Satz 1.1 (Konstanzsatz) *Sei $\Omega \subset \mathbb{R}^n$ offen und wegweise zusammenhängend. Dann gilt:*

$$Df(x) = 0 \quad \textit{für alle } x \in \Omega \quad \Rightarrow \quad f \textit{ ist konstant.}$$

Beweis: Zu $x_0, x_1 \in \Omega$ gibt es nach Lemma 1.2 einen Polygonzug $\gamma : [0,1] \to \Omega$ mit $\gamma(0) = x_0$ und $\gamma(1) = x_1$. Da γ stetig und stückweise C^1, folgt aus Lemma 1.1

$$f(x_1) - f(x_0) = \int_0^1 Df(\gamma(t))\gamma'(t)\,dt = 0.$$

□

Definition 1.2 (Konvexität einer Teilmenge) *Eine Menge $M \subset \mathbb{R}^n$ heißt* konvex, *falls folgende Implikation gilt:*

$$x_0, x_1 \in M \quad \Rightarrow \quad (1-t)x_0 + tx_1 \in M \quad \textit{für alle } t \in [0,1].$$

Satz 1.2 (Schrankensatz) *Sei $\Omega \subset \mathbb{R}^n$ offen und konvex, und $f \in C^1(\Omega, \mathbb{R}^m)$. Es gebe ein $L < \infty$ mit $|Df(x)| \leq L$ für alle $x \in \Omega$. Dann folgt*

$$|f(x_1) - f(x_0)| \leq L\,|x_1 - x_0| \quad \textit{für alle } x_0, x_1 \in \Omega.$$

Beweis: Für jede stetige Funktion $\varphi : I = [a,b] \to \mathbb{R}^m$ gilt die Ungleichung

$$\Big|\int_a^b \varphi\Big| \leq \int_a^b |\varphi|. \tag{1.2}$$

Dies folgt durch Anwendung der Dreiecksungleichung auf die Riemannschen Summen. Sei nun $\gamma(t) = (1-t)x_0 + tx_1$ für $0 \leq t \leq 1$. Aus (1.1) und (3.7) folgt, da $\gamma'(t) = x_1 - x_0$,

$$|f(x_1) - f(x_0)| = \left|\int_0^1 Df\big(\gamma(t)\big)(x_1 - x_0)\,dt\right| \leq \int_0^1 |Df\big(\gamma(t)\big)(x_1 - x_0)|\,dt \leq L\,|x_1 - x_0|.$$

□

Wir halten noch eine etwas andere Formulierung fest, die später zum Beispiel bei der Lösung gewöhnlicher Differentialgleichungen benutzt wird.

Folgerung 1.1 *Sei $\Omega \subset \mathbb{R}^n$ offen und $f \in C^1(\Omega, \mathbb{R}^m)$. Dann gibt es zu jeder kompakten Menge $K \subset \Omega$ eine Konstante $L < \infty$ mit*

$$|f(x) - f(y)| \leq L\,|x - y| \quad \textit{für alle } x, y \in K.$$

Beweis: Angenommen nicht, dann gibt es zu jedem $k \in \mathbb{N}$ Punkte $x_k, y_k \in K$ mit

$$|f(x_k) - f(y_k)| > k\,|x_k - y_k| \qquad \text{für } k = 1, 2, \ldots.$$

Da f stetig auf der kompakten Menge K ist, gibt es ein $M < \infty$ mit $|f(x)| \leq M$ für alle $x \in K$, nach Satz 2.1 in Kapitel 4. Weiter können wir nach Wahl einer Teilfolge und Umnummerierung annehmen, dass $x_k \to x \in K$ mit $k \to \infty$. Aber

$$|x_k - y_k| < \frac{1}{k}\,|f(x_k) - f(y_k)| \leq \frac{2M}{k} \to 0 \quad \text{mit } k \to \infty,$$

also folgt $y_k \to x$ mit $k \to \infty$. Wähle nun ein $r > 0$ mit $\overline{B_r(x)} \subset \Omega$. Da Df stetig ist, gibt es wieder nach Satz 2.1 in Kapitel 4 ein $L < \infty$ mit

$$|Df(y)| \leq L \qquad \text{für alle } y \in \overline{B_r(x)}.$$

Für hinreichend große k gilt $x_k, y_k \in B_r(x)$, also liefert Satz 1.2

$$k|x_k - y_k| < |f(x_k) - f(y_k)| \leq L|x_k - y_k|,$$

ein Widerspruch für k hinreichend groß. $\square$