17
Differentialrechnung mehrerer Variabler

In diesem Abschnitt behandeln wir die Differentialrechnung von Funktionen mit mehreren unabhängigen Veränderlichen. Die Funktionswerte mögen dabei Skalare oder allgemeiner Vektoren in einem endlichdimensionalen, normierten Vektorraum sein.

Wie auch bei Funktionen $f : D \to \mathbb{R}$ in einer Variablen, d. h. $D \subset \mathbb{R}$, geht der Differenzierbarkeit von Funktionen in mehreren Veränderlichen die Stetigkeit voraus, die wir im ersten Band in Abschn. 9.1 allgemein definiert hatten. Wie in Definition 9.6 kann die Stetigkeit als Folgenstetigkeit definiert werden. Ist $D \subset \mathbb{R}^n$ und $\boldsymbol{f} : D \to \mathbb{R}^m$, dann heißt $\boldsymbol{f}$ stetig an einem inneren Punkt $\boldsymbol{x}^0 \in D$, wenn

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}^0} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}^0)$$

für alle gegen $\boldsymbol{x}^0$ konvergenten Vektorfolgen aus D gilt. Das ε-δ-Kriterium (Satz 9.7) der Stetigkeit lautet

$$\forall \varepsilon > 0: \quad \exists \delta > 0: \quad \forall \boldsymbol{x} \in D: \quad \|\boldsymbol{x} - \boldsymbol{x}^0\|_{\mathbb{R}^n} < \delta \quad \Rightarrow \quad \|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}^0)\|_{\mathbb{R}^m} < \varepsilon\,,$$

wobei wir an den Normen die jeweiligen Räume notiert haben. Verwenden wir in $\mathbb{R}^n$ und $\mathbb{R}^m$ jeweils die euklidische Norm (in endlichdimensionalen Räumen sind alle Normen äquivalent), dann lautet das ε-δ-Kriterium

$$\begin{aligned}&\forall \varepsilon > 0: \quad \exists \delta > 0: \quad \forall \boldsymbol{x} \in D:\\ &\left(\sum_{i=1}^{n} \left(x_i - x_i^0\right)^2\right)^{1/2} < \delta \quad \Rightarrow \quad \left(\sum_{i=1}^{m} \left(f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}^0)\right)^2\right)^{1/2} < \varepsilon\,.\end{aligned}$$

Mehr noch als im Fall von reellen Funktionen einer Veränderlichen ist der Nachweis der Stetigkeit einer Funktion mehrerer Variablen mit Hilfe des ε-δ-Kriterium eine Kunst, die in der Regel nur für sehr einfache Funktionen gelingt.

Als einfaches Beispiel soll uns die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ aus Beispiel 17.7,

$$f(x, y) := \begin{cases} \dfrac{xy}{(x^2 + y^2)^2}\,; & \boldsymbol{x} \neq \boldsymbol{0}\,,\\[2mm] 0\,; & \boldsymbol{x} = \boldsymbol{0}\end{cases}$$

Mathematik in den Ingenieur- und Naturwissenschaften 2, 5. Aufl.
Rainer Ansorge, Hans J. Oberle, Kai Rothe und Thomas Sonar.

dienen, wobei wir $\boldsymbol{x} = (x, y)^{\mathrm{T}}$ geschrieben haben. Diese Funktion ist im Punkt $\boldsymbol{x}^0 = (x^0, y^0)^{\mathrm{T}} = \mathbf{0}$ *nicht* stetig, was wir mit dem Folgenkriterium zeigen wollen. Der Funktionswert bei $\boldsymbol{x}^0$ ist nach Definition $f(\boldsymbol{x}^0) = f(\mathbf{0}) = 0$, aber für die Folge

$$\boldsymbol{x}_n := \begin{pmatrix} 1/n \\ 1/n \end{pmatrix},$$

die offenbar gegen $\boldsymbol{x}^0 = \mathbf{0}$ konvergiert, folgt

$$f(\boldsymbol{x}_n) = \frac{x_n y_n}{\left(x_n^2 + y_n^2\right)^2} = \frac{\frac{1}{n^2}}{\left(\frac{1}{n^2} + \frac{1}{n^2}\right)^2} = \frac{n^4}{4n^2} = \frac{1}{4}n^2 \longrightarrow \infty\,, \quad (n \to \infty)\,.$$

In Abschn. 9.2 des ersten Bandes haben wir den Ableitungsbegriff sowohl für skalare, wie auch für vektorwertige Funktionen *einer* reellen Variablen kennengelernt.

Ist $f : \mathbb{R} \supset D \to \mathbb{R}$ eine skalare Funktion, so ist die Ableitung an einem inneren Punkt x_0 des Definitionsbereichs D, vgl. Definition 9.1, definiert durch den Grenzwert

$$f'(x_0) := \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}\,.$$

Dieser Wert gibt die Steigung der Tangente an den Funktionsgraphen

$$\mathrm{graph}(f) := \{(x, f(x))^{\mathrm{T}} : x \in D\}$$

im Punkt $(x_0, f(x_0))^{\mathrm{T}}$ an.

Ist $\boldsymbol{f} : \mathbb{R} \supset D \to \mathbb{R}^m$ nun eine vektorwertige Funktion einer skalaren, unabhängigen Variablen x, also $\boldsymbol{f} = (f_1, \ldots, f_m)^{\mathrm{T}}$, so lässt sich die Ableitung (bei gleicher Definition wie oben) einfach komponentenweise berechnen:

$$\boldsymbol{f}'(x_0) = \left(f_1'(x_0), \ldots, f_m'(x_0)\right)^{\mathrm{T}}\,.$$

Geometrisch beschreibt die Ableitung $\boldsymbol{f}'(x_0)$ in diesem Fall den Geschwindigkeitsvektor der *Kurve* $x \mapsto \boldsymbol{f}(x)$ im Punkt x_0. Die unabhängige Variable wird hierbei als Zeit interpretiert.

Für die Übertragung des Ableitungsbegriffs auf Funktionen mit mehreren unabhängigen Variablen ist diese Interpretation jedoch nicht sinnvoll. Die unabhängige Variable $\boldsymbol{x}$ spielt dann eher die Rolle eines Ortsvektors. Dagegen ist die folgende Interpretation der Ableitung $\boldsymbol{f}'(x_0)$ in diesem Fall nützlich:

Ist $f : \mathbb{R} \supset D \to \mathbb{R}$ eine skalare, differenzierbare Funktion und ist $x_0 \in D^0$ ein innerer Punkt des Definitionsbereichs D, so ist $f'(x_0)$ die eindeutig bestimmte Zahl $a \in \mathbb{R}$, für die die affin-lineare Funktion (Tangente) $\ell(x) := a(x - x_0) + f(x_0)$ die Funktion f in der Nähe von x_0 „am besten" approximiert, d. h., genau für $a = f'(x_0)$ gilt

$$\lim_{x \to x_0} \frac{f(x) - \ell(x)}{|x - x_0|} = \lim_{x \to x_0} \frac{f(x) - f(x_0) - a(x - x_0)}{|x - x_0|} = 0\,.$$

Mit Hilfe des Landau-Symbols, vgl. Definition 9.20, lässt sich diese Beziehung auch folgendermaßen schreiben:

$$f(x) - f(x_0) - a(x - x_0) = o(x - x_0)\,.$$

Diese Charakterisierung der Ableitung $f'(x_0)$ lässt sich nun ohne großen Mühe auch auf Funktionen mit Vektorargumenten übertragen.

Bevor wir diesen Weg in Abschn. 17.2 weiter verfolgen werden, sehen wir uns zunächst eine andere, naheliegende Vorgehensweise an. Hierbei friert man zur Differentiation nach dem Vektor $\boldsymbol{x} \in \mathbb{R}^n$ alle Komponenten von $\boldsymbol{x}$ bis auf eine Komponente ein und differenziert nun nach dieser verbliebenen skalaren Variablen x_i. Auf diese Weise erhält man nun n verschiedene Ableitungen, nämlich für jede der Variablen x_i, $i = 1, \ldots, n$, nach denen differenziert wird. Welche Bedeutung haben diese *partiellen* Ableitungen und in welchen Zusammenhang stehen sie zu der oben beschriebenen Approximationseigenschaft an die Funktion f?

17.1 Partielle Ableitungen

Gegeben sei eine Funktion $f : \mathbb{R}^n \supset D \to \mathbb{R}$ in n Variablen $(x_1, \ldots, x_n)$.

Halten wir alle Variablen $x_1, \ldots, x_{i-1}$ und $x_{i+1}, \ldots, x_n$ fest und differenzieren nun nach der verbleibenden Variablen x_i, so ergeben sich die **partiellen Ableitungen** von f.

Definition (17.1)
Sei $D \subset \mathbb{R}^n$ offen, $f : D \to \mathbb{R}$ und $\boldsymbol{x}^0 \in D$.

a) f heißt in $\boldsymbol{x}^0$ nach der i-ten Koordinate x_i **partiell differenzierbar**, falls der Grenzwert

$$\begin{aligned}\frac{\partial f}{\partial x_i}(\boldsymbol{x}^0) &:= \lim_{t\to 0} \frac{f\left(\boldsymbol{x}^0 + t\boldsymbol{e}_i\right) - f(\boldsymbol{x}^0)}{t} \\ &= \lim_{t\to 0} \frac{f\left(x_1^0, \ldots, x_i^0 + t, \ldots, x_n^0\right) - f\left(x_1^0, \ldots, x_i^0, \ldots, x_n^0\right)}{t}\end{aligned}$$

existiert. $\boldsymbol{e}_i$ bezeichnet hierbei den i-ten Einheitsvektor, $i = 1, \ldots, n$.
Der Grenzwert $\frac{\partial f}{\partial x_i}(\boldsymbol{x}^0)$ heißt die **partielle Ableitung** von f nach der Variablen x_i im Punkt $\boldsymbol{x}^0 \in D$, vgl. Abb. 17.1.

b) Ist f in allen Punkten $\boldsymbol{x}^0 \in D$ partiell differenzierbar nach x_i, so heißt f **partiell differenzierbar nach der Koordinate $\boldsymbol{x_i}$** und $\frac{\partial f}{\partial x_i}$ bezeichnet die Abbildung $\boldsymbol{x}^0 \mapsto \frac{\partial f}{\partial x_i}(\boldsymbol{x}^0)$.
Trifft dies ferner für alle Koordinaten x_i, $i = 1, \ldots, n$, zu, so heißt die Funktion f **partiell differenzierbar**.
Sind darüber hinaus sämtliche partiellen Ableitungen $\frac{\partial f}{\partial x_i}$, $i = 1, \ldots, n$, auf dem Definitionsbereich D stetig, so heißt f **stetig partiell differenzierbar**, oder eine **$\boldsymbol{C^1}$-Funktion** auf D.

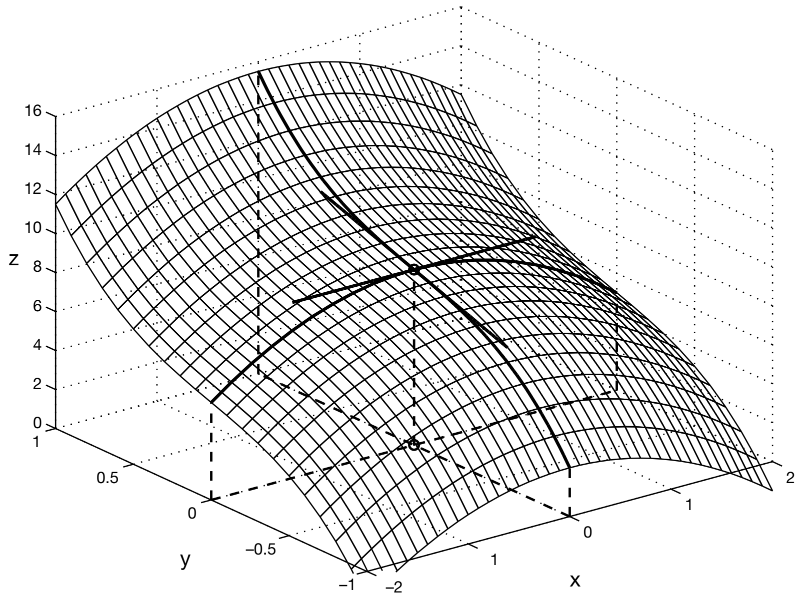

Abb. 17.1 Partielle Ableitungen einer Funktion $z = f(x, y)$.

Bemerkungen (17.2)

a) $\frac{\partial f}{\partial x_i}(\boldsymbol{x}^0)$ ist gerade die übliche eindimensionale Ableitung der „partiellen" Funktion

$$x_i \mapsto f\left(x_1^0, \dots, x_{i-1}^0, x_i, x_{i+1}^0, \dots, x_n^0\right)$$

an der Stelle x_i^0.

b) Sind die Funktionen f und g auf der offenen Menge $D \subset \mathbb{R}^n$ partiell differenzierbar, so gelten die üblichen **Differentiationsregeln**:

$$\frac{\partial}{\partial x_i}\left(\alpha f(\boldsymbol{x}) + \beta g(\boldsymbol{x})\right) = \alpha \frac{\partial f}{\partial x_i}(\boldsymbol{x}) + \beta \frac{\partial g}{\partial x_i}(\boldsymbol{x}), \quad \alpha, \beta \in \mathbb{R},$$

$$\frac{\partial}{\partial x_i}\left(f(\boldsymbol{x}) \cdot g(\boldsymbol{x})\right) = \frac{\partial f}{\partial x_i}(\boldsymbol{x}) \cdot g(\boldsymbol{x}) + f(\boldsymbol{x}) \cdot \frac{\partial g}{\partial x_i}(\boldsymbol{x}),$$

$$\frac{\partial}{\partial x_i}\left(\frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}\right) = \frac{\frac{\partial f}{\partial x_i}(\boldsymbol{x}) \cdot g(\boldsymbol{x}) - f(\boldsymbol{x}) \cdot \frac{\partial g}{\partial x_i}(\boldsymbol{x})}{g(\boldsymbol{x})^2}, \quad g(\boldsymbol{x}) \neq 0.$$

c) Für die partielle Ableitung $\frac{\partial f}{\partial x_i}(\boldsymbol{x})$ sind auch die Bezeichnungen $D_i f(\boldsymbol{x}^0)$ und $f_{x_i}(\boldsymbol{x}^0)$ gebräuchlich.

Beispiele (17.3)

a) Die Funktion $f : \mathbb{R}^3 \to \mathbb{R}$ mit $f(x, y, z) := 3xz + y\sin(x) + ze^y$ ist auf $\mathbb{R}^3$ stetig partiell differenzierbar mit

$$\frac{\partial f}{\partial x} = 3z + y\cos(x)\,, \quad \frac{\partial f}{\partial y} = \sin(x) + ze^y\,, \quad \frac{\partial f}{\partial z} = 3x + e^y$$

b) Der Schalldruck einer räumlich eindimensionalen Schallwelle ist gegeben durch die Funktion

$$p(x, t) := A\sin(\alpha x - \omega t)\,.$$

Die partielle Ableitung $\frac{\partial p}{\partial x} = \alpha A\cos(\alpha x - \omega t)$ beschreibt dann zu einem festen Zeitpunkt t die örtliche Änderung des Schalldrucks. Analog beschreibt $\frac{\partial p}{\partial t} = -\omega A\cos(\alpha x - \omega t)$ an einem festen Ort x die zeitliche Änderung des Schalldrucks.

c) Die Zustandsgleichung eines idealen Gases lautet $pV = RT$. Dabei ist p der Druck, V das Volumen und T die (absolute) Temperatur des Gases. Ferner bezeichnet R die universelle Gaskonstante.
Jede der drei Größen p, V und T lässt sich vermöge der obigen Zustandsgleichung als Funktion der beiden anderen Variablen auffassen:

$$p = p(V, T) = \frac{RT}{V}\,, \quad V = V(p, T) = \frac{RT}{p}\,, \quad T = T(p, V) = \frac{pV}{R}\,.$$

Durch Berechnung der zugehörigen partiellen Ableitungen und unter Verwendung der Zustandsgleichung lässt sich folgern:

$$\frac{\partial V}{\partial p} \cdot \frac{\partial p}{\partial T} = -\frac{\partial V}{\partial T}\,.$$

Definition (17.4)

Sei $D \subset \mathbb{R}^n$ offen, $f : D \to \mathbb{R}$ sei im Punkt $\boldsymbol{x}^0 \in D$ (nach allen Koordinaten) partiell differenzierbar. Der Vektor

$$\boldsymbol{\nabla} f(\boldsymbol{x}^0) := \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}^0), \ldots, \frac{\partial f}{\partial x_n}(\boldsymbol{x}^0)\right)^{\mathrm{T}}$$

heißt der **Gradient** der Funktion f im Punkt $\boldsymbol{x}^0$.

Dabei wird der symbolische Vektor $\boldsymbol{\nabla} := (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n})^{\mathrm{T}}$ **Nabla-Operator** genannt. Er ist nach der Form eines hebräischen Musikinstrumentes benannt.

Für den Gradienten ist mitunter auch die Schreibweise $\operatorname{grad} f(\boldsymbol{x}^0) := \boldsymbol{\nabla} f(\boldsymbol{x}^0)^{\mathrm{T}}$ als Zeilenvektor gebräuchlich. Wir werden im Folgenden beide Schreibweisen verwenden.

Bemerkung (17.5)
Sind die Funktionen $f, g\colon \mathbb{R}^n \supset D \to \mathbb{R}$ auf der offenen Menge D partiell differenzierbar, so gelten die folgenden **Differentiationsregeln**:

$$\begin{aligned}
\nabla(\alpha f + \beta g) &= \alpha \nabla f + \beta \nabla g \\
\nabla(f \cdot g) &= g \cdot \nabla f + f \cdot \nabla g \\
\nabla\left(\frac{f}{g}\right) &= \frac{g \cdot \nabla f - f \cdot \nabla g}{g^2}\,, \quad g(\boldsymbol{x}) \neq 0\,.
\end{aligned}$$

Beispiele (17.6)

a) Für die Funktion $f(x, y) := \mathrm{e}^x \cdot \sin y$ erhält man

$$\nabla f(x, y) = (\mathrm{e}^x \sin y, \mathrm{e}^x \cos y)^{\mathrm{T}} = \mathrm{e}^x (\sin y, \cos y)^{\mathrm{T}}\,.$$

b) Es bezeichne $r(\boldsymbol{x}) := \|\boldsymbol{x}\|_2 := \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$ die euklidische Norm des Vektors $\boldsymbol{x}$. Durch Differentiation ergibt sich

$$\frac{\partial r}{\partial x_i} = \frac{2x_i}{2\sqrt{x_1^2 + \ldots + x_n^2}} = \frac{x_i}{r}\,, \quad \boldsymbol{x} \neq \boldsymbol{0}$$

und damit $\nabla r = \frac{\boldsymbol{x}}{r}$, $\boldsymbol{x} \neq \boldsymbol{0}$.

c) In Verallgemeinerung des Beispiels b) erhält man mittels der Kettenregel für $k \in \mathbb{Z}$:

$$\nabla r^k = k r^{k-1} \nabla r = k r^{k-2} \boldsymbol{x}\,, \quad \boldsymbol{x} \neq \boldsymbol{0}\,.$$

d) Wir betrachten zwei Körper (Satellit und Erde) mit den Massen m und M. Der kleine Körper befinde sich in $\boldsymbol{x} \in \mathbb{R}^3 \setminus \{\boldsymbol{0}\}$, der große Körper im Ursprung. Beide Körper werden durch die Gravitation angezogen. Das Gravitationskraftfeld, dass auf den Körper in $\boldsymbol{x}$ wirkt, ist gegeben durch $\boldsymbol{K} = -\gamma m M \boldsymbol{x}/r^3$. Dabei ist γ die Gravitationskonstante. Aufgrund der Beziehung c) sehen wir nun, dass es eine Funktion Φ gibt, nämlich $\Phi(\boldsymbol{x}) := \gamma m M/r$, für die $\nabla \Phi = \boldsymbol{K}$ auf $\mathbb{R}^3 \setminus \{\boldsymbol{0}\}$ gilt. Man sagt, Φ ist ein Potential für das Gravitationsfeld $\boldsymbol{K}$.

Erstaunlicherweise lassen sich jedoch nicht alle Eigenschaften der Differentialrechnung einer Variablen ohne Weiteres auf den Fall partieller Ableitungen von Funktionen mehrerer Veränderlichen übertragen. So genügt beispielsweise die partielle Differenzierbarkeit i. Allg. nicht, um hiermit die Stetigkeit einer Funktion $f(x_1, \ldots, x_n)$ garantieren zu können. Man vergleiche dagegen Satz 9.23a) für Funktionen einer Variablen.

Beispiel (17.7)
Sei die Funktion $f\colon \mathbb{R}^2 \to \mathbb{R}$, gegeben durch

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

f ist auf ganz $\mathbb{R}^2$ partiell differenzierbar mit $f_x(0,0) = f_y(0,0) = 0$, sowie

$$\frac{\partial f}{\partial x} = \frac{y}{(x^2+y^2)^2} - 4\frac{x^2 y}{(x^2+y^2)^3}\,,$$
$$\frac{\partial f}{\partial y} = \frac{x}{(x^2+y^2)^2} - 4\frac{x y^2}{(x^2+y^2)^3}$$

für $(x,y) \neq (0,0)$.

Andererseits ist die Funktion f aber in $(x_0, y_0) = (0,0)$ nicht stetig! Wir haben bereits zu Anfang des Kapitels gesehen, dass der Grenzwert der Folge $f(1/n, 1/n)$ für $n \to \infty$ die Unstetigkeit in $(x_0, y_0) = (0,0)$ zeigt.

Um aus der partiellen Differenzierbarkeit einer Funktion auf deren Stetigkeit schließen zu können, benötigt man eine zusätzliche Eigenschaft, beispielsweise die Beschränktheit aller partiellen Ableitungen dieser Funktion. Diese Bedingung ist gerade in Beispiel 17.7 verletzt!

Satz (17.8)
Ist $f\colon \mathbb{R}^n \supset D \to \mathbb{R}$ in einer Umgebung eines inneren Punktes $\boldsymbol{x}^0 \in D$ partiell differenzierbar und sind die partiellen Ableitungen $\frac{\partial f}{\partial x_i}$, $i = 1,\ldots,n$, dort beschränkt, so ist f im Punkt $\boldsymbol{x}^0$ stetig.

Beweis.
Zu einem hinreichend kleinen $\varepsilon > 0$ und $\|\boldsymbol{x} - \boldsymbol{x}^0\|_\infty < \varepsilon$ bilden wir:

$$\begin{aligned} f(\boldsymbol{x}) - f(\boldsymbol{x}^0) = {} & \left[f(x_1,\ldots,x_n) - f\left(x_1,\ldots,x_{n-1},x_n^0\right)\right] \\ & + \left[f\left(x_1,\ldots,x_{n-1},x_n^0\right) - f\left(x_1,\ldots,x_{n-1}^0,x_n^0\right)\right] \\ & \vdots \\ & + \left[f\left(x_1,x_2^0,\ldots,x_n^0\right) - f\left(x_1^0,x_2^0,\ldots,x_n^0\right)\right]\,. \end{aligned}$$

Für jede der obigen Differenzen betrachten wir f als Funktion nur einer Variablen, nämlich: $f(x_1,\ldots,x_{n-1},\,\cdot\,)$, $f(x_1,\ldots,x_{n-2},\,\cdot\,,x_n^0),\ldots,f(\,\cdot\,,x_2^0,\ldots,x_n^0)$. Alle diese partiellen Funktionen einer Variablen sind nach Voraussetzung differenzierbar und damit auch stetig in der Nähe von x_j^0, $j = n, n-1, \ldots, 1$. Somit lässt sich auf die obigen Differenzen jeweils der erste Mittelwertsatz, vgl. Satz 10.8, anwenden und man erhält mit geeigneten Zwischenstellen $\xi_1,\ldots,\xi_n$:

$$\begin{aligned} f(\boldsymbol{x}) - f(\boldsymbol{x}^0) = {} & \frac{\partial f}{\partial x_n}(x_1,\ldots,x_{n-1},\xi_n)\,\left(x_n - x_n^0\right) \\ & + \frac{\partial f}{\partial x_{n-1}}\,\left(x_1,\ldots,x_{n-2},\xi_{n-1},x_n^0\right)\,\left(x_{n-1} - x_{n-1}^0\right) \\ & \vdots \\ & + \frac{\partial f}{\partial x_1}\,\left(\xi_1,x_2^0,\ldots,x_n^0\right)\,\left(x_1 - x_1^0\right)\,. \end{aligned}$$

Nun sind die partiellen Ableitungen nach Voraussetzung in der Umgebung $\|\boldsymbol{x} - \boldsymbol{x}^0\|_\infty < \varepsilon$ beschränkt: Man erhält daher aus der obigen Relation eine Abschätzung der Form:

$$|f(\boldsymbol{x}) - f(\boldsymbol{x}^0)| \leq C_1\,|x_1 - x_1^0| + \ldots + C_n\,|x_n - x_n^0|\,.$$

Daher folgt $f(\boldsymbol{x}) \to f(\boldsymbol{x}^0)$ für $\|\boldsymbol{x} - \boldsymbol{x}^0\|_\infty \to 0$, also die Stetigkeit von f in $\boldsymbol{x}^0$.

■

Bemerkung (17.9)
Die Voraussetzungen des Satzes 17.8 sind erfüllt, falls f **stetig** partiell differenzierbar ist. Die partiellen Ableitungen $\frac{\partial f}{\partial x_i}$, $i = 1, \ldots, n$, sind dann nämlich aufgrund der Min-Max-Eigenschaft von stetigen Funktionen, vgl. Satz 9.13, auf einem Kompaktum $\|\boldsymbol{x} - \boldsymbol{x}^0\|_\infty \leq \varepsilon$ beschränkt!

Stetig partiell differenzierbare Funktionen sind mithin immer auch stetig.

Höhere Ableitungen

Definition (17.10)
Eine skalare Funktion f sei auf einer offenen Menge $D \subset \mathbb{R}^n$ partiell differenzierbar. Die partiellen Ableitungen sind dann selbst wieder Funktionen $\frac{\partial f}{\partial x_i}: D \to \mathbb{R}$, und wir können uns fragen, ob diese wiederum partiell differenzierbar sind. Ist dies der Fall, so erhalten wir hiermit die **partiellen Ableitungen zweiter Ordnung** der Funktion f.

Induktiv definieren wir für $i_1, \ldots, i_k \in \{1, \ldots, n\}$:

$$\frac{\partial^2 f}{\partial x_j\,\partial x_i} := \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)$$

$$\frac{\partial^k f}{\partial x_{i_k}\partial x_{i_{k-1}} \ldots \partial x_{i_1}} := \frac{\partial}{\partial x_{i_k}}\left(\frac{\partial^{k-1} f}{\partial x_{i_{k-1}} \ldots \partial x_{i_1}}\right), \quad k \geq 2\,.$$

Die Funktion f heißt ***k*-fach partiell differenzierbar**, falls alle partiellen Ableitungen der Ordnung k

$$\frac{\partial^k f}{\partial x_{i_k}\partial x_{i_{k-1}} \ldots \partial x_{i_1}} = D_{i_k} D_{i_{k-1}} \ldots D_{i_1} f = f_{x_{i_1} \ldots x_{i_k}}\,, \; \{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$$

gemäß der obigen Definition auf D existieren.

Sind diese partiellen Ableitungen zudem alle stetig, so heißt die Funktion f ***k*-fach stetig partiell differenzierbar**, oder eine **C^k-Funktion** auf D, $k = 1, 2, 3 \ldots$

Darüber hinaus werden stetige Funktionen üblicherweise auch als **C^0-Funktionen** bezeichnet. Ferner heißen Funktionen, die beliebig oft stetig partiell differenzierbar sind, **C^∞-Funktionen**.

Für C^2-Funktionen ist die Reihenfolge, in der die partiellen Ableitungen zweiter Ordnung berechnet werden, unerheblich. So erhält man beispielsweise für die Funktion $f(x,y) := x^3 \sin y + x^4 y^2$

$$\frac{\partial^2 f}{\partial x\,\partial y}(x,y) = \frac{\partial}{\partial x}(x^3 \cos y + 2x^4 y) = 3x^2 \cos y + 8x^3 y$$
$$\frac{\partial^2 f}{\partial y\,\partial x}(x,y) = \frac{\partial}{\partial y}(3x^2 \sin y + 4x^3 y^2) = 3x^2 \cos y + 8x^3 y\,.$$

Tatsächlich ist aber die Stetigkeit der partiellen Ableitungen hierbei eine unentbehrliche Voraussetzung.

Satz (17.11): Vertauschbarkeitssatz von Schwarz[1)]
Ist $f : \mathbb{R}^n \supset D \to \mathbb{R}$ eine C^2-Funktion auf der offenen Menge D, so gilt für alle $i, j \in \{1, \dots, n\}$

$$\frac{\partial^2 f}{\partial x_j\,\partial x_i}(\boldsymbol{x}) = \frac{\partial^2 f}{\partial x_i\,\partial x_j}(\boldsymbol{x})\,, \quad \forall\, \boldsymbol{x} \in D\,.$$

Beweis.
Da für die Aussage des Satzes offenbar nur die Koordinaten x_i und x_j eine Rolle spielen, genügt es, o. B. d. A. den Fall $n = 2$ sowie $i = 1$ und $j = 2$ zu betrachten.

Sei also $(x^0, y^0) \in D$ und $\varepsilon > 0$ so klein gewählt, dass

$$Q := \{(x,y) : |x - x^0| \le \varepsilon \ \wedge\ |y - y^0| \le \varepsilon\} \subset D\,.$$

Für $(x,y) \in Q$ mit $x \ne x^0$ und $y \ne y^0$ berechnen wir den Ausdruck

$$A(x,y) := f(x,y) - f(x^0,y) - f(x,y^0) + f(x^0,y^0)\,.$$

Unter zweifacher Anwendung des Mittelwertsatzes in einer Variablen finden wir

$$\begin{aligned} A(x,y) &= (f(x,y) - f(x^0,y)) - (f(x,y^0) - f(x^0,y^0)) \\ &=: Z_1(y) - Z_1(y^0) \\ &= Z_1'(\eta_1)\,(y - y^0) \\ &= \left(\frac{\partial f}{\partial y}\,(x,\eta_1) - \frac{\partial f}{\partial y}\,\left(x^0,\eta_1\right)\right)\,(y - y^0) \\ &= \frac{\partial^2 f}{\partial x\,\partial y}(\xi_1,\eta_1)\,(x - x^0)\,(y - y^0) \end{aligned}$$

mit Zwischenwerten $\xi_1 = x^0 + \theta_1(x - x^0)$ und $\eta_1 = y^0 + \widetilde{\theta}_1(y - y^0)$, $0 < \theta_1, \widetilde{\theta}_1 < 1$.

1) Hermann Amandus Schwarz (1843–1921); Halle, Zürich, Göttingen, Berlin.

Eine andere Zusammenfassung der Terme in $A(x, y)$ ergibt

$$\begin{aligned}
A(x, y) &= (f(x, y) - f(x, y^0)) - (f(x^0, y) - f(x^0, y^0)) \\
&=: Z_2(x) - Z_2(x^0) \\
&= Z_2'(\xi_2)(x - x^0) \\
&= \left(\frac{\partial f}{\partial x}(\xi_2, y) - \frac{\partial f}{\partial x}(\xi_2, y^0) \right)(x - x^0) \\
&= \frac{\partial^2 f}{\partial y\, \partial x}(\xi_2, \eta_2)(x - x^0)(y - y^0)
\end{aligned}$$

mit Zwischenwerten $\xi_2 = x^0 + \theta_2(x - x^0)$ und $\eta_2 = y^0 + \widetilde{\theta}_2(y - y^0)$, $0 < \theta_2, \widetilde{\theta}_2 < 1$.
Da $x \neq x^0$ und $y \neq y^0$ vorausgesetzt wurde, folgt

$$\frac{\partial^2 f}{\partial x\, \partial y}(\xi_1, \eta_1) = \frac{\partial^2 f}{\partial y\, \partial x}(\xi_2, \eta_2)\,.$$

Bilden wir nun den Grenzwert dieser Ausdrücke für $(x, y) \to (x^0, y^0)$, so erhält man aufgrund der Stetigkeit der zweiten partiellen Ableitungen

$$\frac{\partial^2 f}{\partial x\, \partial y}(x^0, y^0) = \frac{\partial^2 f}{\partial y\, \partial x}(x^0, y^0)\,.$$

■

Folgerung (17.12)
Ist f eine C^k-Funktion, $k \geq 2$, so kann man die Reihenfolge der Differentiationen zur Berechnung der partiellen Ableitungen von f bis zur k-ten Ordnung beliebig vertauschen

$$\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \dots \partial x_{i_1}} = \frac{\partial^k f}{\partial x_{j_k} \partial x_{j_{k-1}} \dots \partial x_{j_1}}\,,$$

wobei $(j_1, \dots, j_k)$ eine beliebige Permutation von $(i_1, \dots, i_k)$ ist.

Beispiele (17.13)

a) Dass die Stetigkeit der partiellen Ableitungen eine wichtige Voraussetzung für Vertauschbarkeit der Ableitungsreihenfolge ist, zeigt das folgende Beispiel. Die Funktion

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

ist auf $\mathbb{R}^2$ stetig und zweifach partiell differenzierbar. Eine explizite Berechnung der partiellen Ableitungen erster und zweiter Ordnung ergibt jedoch

$$\begin{aligned}
f_{xy}(0, 0) &= \frac{\partial^2 f}{\partial y\, \partial x}(0, 0) = -1 \\
f_{yx}(0, 0) &= \frac{\partial^2 f}{\partial x\, \partial y}(0, 0) = +1\,.
\end{aligned}$$

b) Will man für die Funktion

$$f(x, y, z) := y^2 z \sin(x^3) + (\cosh y + 17\, e^{x^2}) z^2$$

die partielle Ableitung dritter Ordnung f_{xyz} berechnen, so kann man aufgrund des Satzes von Schwarz zunächst nach z differenzieren:

$$f_{xyz} = \frac{\partial^2}{\partial x\, \partial y} [y^2 \sin(x^3)] + 2z \cdot \frac{\partial^2}{\partial x\, \partial y} [\cosh y + 17 e^{x^2}] = 6yx^2 \cos(x^3)\,.$$

Viele physikalische und ingenieurwissenschaftliche Probleme führen auf sogenannte **partielle Differentialgleichungen zweiter Ordnung**. Dies sind Gleichungen, in denen neben der gesuchten Funktion (mit physikalischer Bedeutung Druck, Temperatur, …) auch deren partielle Ableitungen nach Ort und Zeit bis zur zweiten Ordnung auftreten. Eine besondere Bedeutung kommt dabei dem sogenannten **Laplace**[2]**–Operator** in den Ortsvariablen $x_1, \ldots, x_n$ zu, der folgendermaßen definiert wird

$$\Delta := \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}\,. \tag{17.1}$$

Für eine skalare Funktion $u = u(\boldsymbol{x}, t)$, die vom Ort $\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$ und der Zeit t abhängt, ist also

$$\Delta u = \sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2} = u_{x_1 x_1} + \ldots + u_{x_n x_n}\,. \tag{17.2}$$

So lassen sich beispielsweise die Transversalschwingungen einer Membran, die Ausbreitung von Schallwellen und auch andere Schwingungsprobleme durch eine skalare Funktion $u = u(\boldsymbol{x}, t)$ beschreiben, die der sogenannten **Wellengleichung**

$$\Delta u(\boldsymbol{x}, t) - \frac{1}{c^2} u_{tt}(\boldsymbol{x}, t) = 0 \tag{17.3}$$

genügt. Die Konstante c ist hierbei die Wellengeschwindigkeit. Die Ableitungen des Laplace-Operators wirken dabei wie in (17.2) nur auf die Ortsvariablen $\boldsymbol{x}$.

Die Temperaturverteilung in einem isotropen, starren Körper wird durch die sogenannte **Wärmeleitungsgleichung**

$$\Delta u(\boldsymbol{x}, t) - \frac{1}{k} u_t(\boldsymbol{x}, t) = \varphi(\boldsymbol{x}, t) \tag{17.4}$$

beschrieben. Hierbei bezeichnet $u(\boldsymbol{x}, t)$ die (normierte) Temperatur des Körpers am Ort $\boldsymbol{x}$ und zur Zeit t. $\varphi(\boldsymbol{x}, t)$ steht für die Dichte von zusätzlichen Wärmequellen, und k bezeichnet die innere Temperaturleitfähigkeit. Genauer gilt $k = \lambda/(\rho c)$ mit λ: Wärmeleitfähigkeit, ρ: Dichte und c: spezifische Wärme. Wiederum wirken die Ableitungen des Laplace-Operators nur auf die Ortsvariablen.

2) Pierre Simon Marquis de Laplace (1749–1827); Paris.

Betrachtet man die beiden Gln. (17.3) und (17.4) im homogenen und stationären Fall, also $\varphi = 0$ und $u_t = 0$, so gehen beide Gleichungen in die sogenannte **Laplace-Gleichung** oder **Potentialgleichung**

$$\Delta u(\boldsymbol{x}) = 0 \tag{17.5}$$

für die Funktion $u = u(\boldsymbol{x})$ über. Die Lösungen dieser partiellen Differentialgleichung heißen **harmonische Funktionen**.

Beispiele (17.14)

a) Wir berechnen den Laplace-Operator Δu für eine Funktion $u = u(r)$, die nur vom Abstand r des Punktes $\boldsymbol{x}$ vom Ursprung abhängt.
Mit $r = \|\boldsymbol{x}\|_2$ und Beispiel 17.6b) findet man

$$\begin{aligned}
\Delta u(r) &= \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2} u(r) = \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left[u'(r)\frac{x_i}{r}\right] \\
&= \sum_{i=1}^{n} \left\{ u''(r)\frac{x_i^2}{r^2} + u'(r)\frac{r - x_i^2/r}{r^2} \right\} \\
&= \frac{u''(r)}{r^2}\left(\sum_{i=1}^{n} x_i^2\right) + \frac{u'(r)}{r^3}\sum_{i=1}^{n}\left(r^2 - x_i^2\right) \\
&= u''(r) + \frac{n-1}{r}u'(r)\,.
\end{aligned}$$

b) Mit Hilfe der Darstellung von $\Delta u(r)$ aus a) folgt nun direkt (einsetzen!), dass die Funktion

$$u(r) = \begin{cases} \ln r\,, & \text{für} \quad n = 2 \\ r^{2-n}\,, & \text{für} \quad n > 2 \end{cases}$$

eine radialsymmetrische Lösung der Laplace-Gleichung auf $\mathbb{R}^n \setminus \{\mathbf{0}\}$ bildet.

c) Ebenso folgt mit a), dass die Funktion

$$u(\boldsymbol{x}, t) := \frac{1}{r}\cos(r - ct)\,, \qquad r = \|\boldsymbol{x}\|_2\,, \qquad \boldsymbol{x} \in \mathbb{R}^3 \setminus \{\mathbf{0}\}\,,$$

die Wellengleichung (17.3) löst.

Vektorwertige Funktionen

Im Folgenden betrachten wir vektorwertige Funktionen $\boldsymbol{f}\colon D \to \mathbb{R}^m$, wobei $D \subset \mathbb{R}^n$ offen sei und $m,\ n > 1$.

Definition (17.15)
Die Funktion $\boldsymbol{f}$ heißt **partiell differenzierbar** in $\boldsymbol{x}^0 \in D$, falls für alle $i = 1, \dots, n$ die Grenzwerte

$$\frac{\partial \boldsymbol{f}}{\partial x_i}(\boldsymbol{x}^0) := \lim_{t \to 0} \frac{\boldsymbol{f}\left(\boldsymbol{x}^0 + t\boldsymbol{e}_i\right) - \boldsymbol{f}(\boldsymbol{x}^0)}{t}$$

existieren.

Die partiellen Ableitungen lassen sich demnach komponentenweise berechnen

$$\frac{\partial \boldsymbol{f}}{\partial x_i}\,(\boldsymbol{x}^0) = \left(\frac{\partial f_1}{\partial x_i}\,(\boldsymbol{x}^0), \ldots, \frac{\partial f_m}{\partial x_i}\,(\boldsymbol{x}^0)\right)^{\mathrm{T}} . \tag{17.6}$$

Sind alle partiellen Ableitungen $\frac{\partial f_i}{\partial x_j}$ auf D erklärt und dort stetig, so heißt $\boldsymbol{f}$ **stetig partiell differenzierbar**, oder eine $\boldsymbol{C^1}$**-Funktion**.

Analog nennt man $\boldsymbol{f}$ eine $\boldsymbol{C^k}$**-Funktion**, $k \in \mathbb{N}_0 \cup \{\infty\}$, wenn jede Komponentenfunktion f_i von $\boldsymbol{f}$ eine C^k-Funktion ist, vgl. Definition 17.10.

Fasst man die partiellen Ableitungen $\frac{\partial \boldsymbol{f}}{\partial x_i}$ (Spaltenvektoren) zu einer Matrix zusammen, so erhält man die sogenannte **Jacobi**[3)]**–Matrix**, auch **Funktionalmatrix** genannt, der Funktion $\boldsymbol{f}$

$$\boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}^0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}\,(\boldsymbol{x}^0)\,, & \ldots\,, & \frac{\partial f_1}{\partial x_n}\,(\boldsymbol{x}^0) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}\,(\boldsymbol{x}^0)\,, & \ldots\,, & \frac{\partial f_m}{\partial x_n}\,(\boldsymbol{x}^0) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\nabla} f_1(\boldsymbol{x}^0)^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\nabla} f_m(\boldsymbol{x}^0)^{\mathrm{T}} \end{pmatrix} . \tag{17.7}$$

Definition (17.16)
Im Fall $m = n$ nennt man eine Funktion $\boldsymbol{f} : \mathbb{R}^n \supset D \to \mathbb{R}^n$ ein **Vektorfeld** auf D. Ist $\boldsymbol{f}$ zudem eine C^k-Funktion, so heißt $\boldsymbol{f}$ ein $\boldsymbol{C^k}$**-Vektorfeld**, $k \in \mathbb{N}_0 \cup \{\infty\}$.

Beispiele für Vektorfelder sind etwa die Geschwindigkeits- und Beschleunigungsfelder strömender Flüssigkeiten oder Gase, elektromagnetische Felder oder der Temperaturgradient eines bestimmten Volumens (Körpers).

Ist $f : \mathbb{R}^n \supset D \to \mathbb{R}$ eine skalare, partiell differenzierbare Funktion, so ist der Gradient $\boldsymbol{\nabla} f$ ein Vektorfeld auf D, das sogenannte **Gradientenfeld** von $\boldsymbol{f}$.

Für Vektorfelder sind neben der Jacobi-Matrix zwei weitere Ausdrücke in den partiellen Ableitungen erster Ordnung von besonderer Bedeutung, die Divergenz und die Rotation. Wir geben im Folgenden die formalen Definitionen und die resultierenden Rechenregeln an. Auf die anschauliche Bedeutung dieser Größen werden wir erst später im Zusammenhang mit der Integration von Vektorfeldern eingehen können.

Definition (17.17)
Für ein partiell differenzierbares Vektorfeld $\boldsymbol{f} : \mathbb{R}^n \supset D \to \mathbb{R}^n$, D offen, definieren wir die **Divergenz** durch

$$\operatorname{div} \boldsymbol{f}\,(\boldsymbol{x}^0) := \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}\,(\boldsymbol{x}^0)\,.$$

Formale Schreibweise: $\operatorname{div} \boldsymbol{f}\,(\boldsymbol{x}) = \boldsymbol{\nabla}^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{x}) = \langle \boldsymbol{\nabla}, \boldsymbol{f}(\boldsymbol{x}) \rangle$.

3) Carl Gustav Jacobi (1804–1851); Königsberg, Berlin.

Bemerkungen (17.18)

a) Für partiell differenzierbare Vektorfelder $\boldsymbol{f}, \boldsymbol{g}\colon D \to \mathbb{R}^n$ und eine partiell differenzierbare skalare Funktion $\varphi\colon D \to \mathbb{R}$ gelten die folgenden **Rechenregeln**:
Linearität: $\operatorname{div}(\alpha \boldsymbol{f} + \beta \boldsymbol{g}) = \alpha \operatorname{div} \boldsymbol{f} + \beta \operatorname{div} \boldsymbol{g}$,
Produktregel: $\operatorname{div}(\varphi \boldsymbol{f}) = \langle \boldsymbol{\nabla} \varphi, \boldsymbol{f} \rangle + \varphi \operatorname{div} \boldsymbol{f}$.

b) Ist $f : D \to \mathbb{R}$ eine C^2-Funktion, so findet man für den Laplace-Operator von f die Darstellung

$$\Delta f = \operatorname{div}(\boldsymbol{\nabla} f) .$$

Beispiele (17.19)

a) Für $\boldsymbol{x} \in \mathbb{R}^n$, $r = \|\boldsymbol{x}\|_2$ und $k \in \mathbb{N}$ berechnet man mittels Produktregel und Beispiel 17.6

$$\begin{aligned} \operatorname{div}\left(\frac{\boldsymbol{x}}{r^k}\right) &= \left\langle \boldsymbol{\nabla} \frac{1}{r^k} , \boldsymbol{x} \right\rangle + \frac{1}{r^k} \operatorname{div} \boldsymbol{x} \\ &= \left\langle -k r^{-(k+2)} \boldsymbol{x} , \boldsymbol{x} \right\rangle + \frac{1}{r^k} n = \frac{n-k}{r^k} . \end{aligned}$$

Speziell für $k = n = 3$ folgt, dass alle Vektorfelder der Form

$$\boldsymbol{K}(\boldsymbol{x}) = c \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|^3} , \quad \boldsymbol{x} \in \mathbb{R}^3 , \quad c = \text{const.} ,$$

sogenannte **zentrale Kraftfelder**, divergenzfrei sind, vgl. Beispiel 17.6c).

b) Mit Hilfe der Produktregel und des obigen Beispiels lässt sich $\Delta u(r)$ für eine radialsymmetrische Funktion u auch folgendermaßen berechnen, vgl. auch Beispiel 17.14

$$\begin{aligned} \Delta u(r) = \operatorname{div}(\boldsymbol{\nabla} u(r)) &= \operatorname{div}\left(u'(r) \frac{\boldsymbol{x}}{r} \right) \\ &= \left\langle \boldsymbol{\nabla} u'(r) , \frac{\boldsymbol{x}}{r} \right\rangle + u'(r) \operatorname{div}\left(\frac{\boldsymbol{x}}{r} \right) \\ &= \left\langle u''(r) \frac{\boldsymbol{x}}{r} , \frac{\boldsymbol{x}}{r} \right\rangle + u'(r) \frac{n-1}{r} \\ &= u''(r) + \frac{n-1}{r} u'(r) . \end{aligned}$$

Definition (17.20)

Für ein partiell differenzierbares Vektorfeld $\boldsymbol{f} : \mathbb{R}^3 \supset D \to \mathbb{R}^3$, D offen, definieren wir die **Rotation** durch

$$\operatorname{rot} \boldsymbol{f}(\boldsymbol{x}^0) := \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3} , \quad \frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1} , \quad \frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2} \right)^{\mathrm{T}} \Bigg|_{\boldsymbol{x}^0} .$$

Formale Schreibweise:

$$\operatorname{rot} \boldsymbol{f} = \boldsymbol{\nabla} \times \boldsymbol{f} = \begin{vmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \\ \dfrac{\partial}{\partial x_1} & \dfrac{\partial}{\partial x_2} & \dfrac{\partial}{\partial x_3} \\ f_1 & f_2 & f_3 \end{vmatrix} .$$

Bemerkungen (17.21)

a) Für partiell differenzierbare Vektorfelder $\boldsymbol{f}, \boldsymbol{g}\colon D \to \mathbb{R}^3$ und eine partiell differenzierbare skalare Funktion $\varphi\colon D \to \mathbb{R}$ gelten die folgenden **Rechenregeln**:
Linearität: $\operatorname{rot}(\alpha \boldsymbol{f} + \beta \boldsymbol{g}) = \alpha \operatorname{rot} \boldsymbol{f} + \beta \operatorname{rot} \boldsymbol{g}$
Produktregel: $\operatorname{rot}(\varphi \boldsymbol{f}) = (\boldsymbol{\nabla}\varphi) \times \boldsymbol{f} + \varphi \operatorname{rot} \boldsymbol{f}$.

b) Für eine C^2-Funktion $\varphi\colon D \to \mathbb{R}$, $D \subset \mathbb{R}^3$ offen, folgt aufgrund des Vertauschbarkeitssatzes von Schwarz:

$$\operatorname{rot}(\boldsymbol{\nabla}\varphi) = \boldsymbol{0} .$$

Gradientenfelder sind also stets rotationsfrei!

Beispiele (17.22)

a) Für $\boldsymbol{x} \in \mathbb{R}^3 \setminus \{\boldsymbol{0}\}$, $r := \|\boldsymbol{x}\|_2$ und $k \in \mathbb{N}$ findet man mit Hilfe der Produktregel:

$$\begin{aligned}\operatorname{rot}\left(\frac{\boldsymbol{x}}{r^k}\right) &= \boldsymbol{\nabla}\left(\frac{1}{r^k}\right) \times \boldsymbol{x} + \frac{1}{r^k} \operatorname{rot} \boldsymbol{x} \\ &= -\frac{k}{r^{k+2}} \boldsymbol{x} \times \boldsymbol{x} + \frac{1}{r^k}\, \boldsymbol{0} = \boldsymbol{0} .\end{aligned}$$

Insbesondere sind also zentrale Kraftfelder rotationsfrei! Man vergleiche hierzu auch das Beispiel 17.19.

b) Für einen festen Vektor $\boldsymbol{a} \in \mathbb{R}^3 \setminus \{\boldsymbol{0}\}$ beschreibt die Funktion $\boldsymbol{v}(\boldsymbol{x}) := \boldsymbol{a} \times \boldsymbol{x}$ das Geschwindigkeitsfeld der Drehung des $\mathbb{R}^3$ um die Achse $\boldsymbol{a}$ mit der konstanten Winkelgeschwindigkeit $\omega = \|\boldsymbol{a}\|$.
Die explizite Berechnung der Rotation ergibt $\operatorname{rot} \boldsymbol{v}(\boldsymbol{x}) = 2\boldsymbol{a}$.

17.2 Das vollständige Differential

Für die Differentialrechnung mehrerer Veränderlicher ist insbesondere die eingangs erwähnte Approximationseigenschaft

$$f(x) = f(x_0) + f'(x_0)\,(x - x_0) + o(x - x_0)$$

wesentlich. Sie gestattet es, die Funktion f in der Nähe des Punktes x_0 durch eine affin-lineare Funktion, im eindimensionalen Fall eine Gerade, anzunähern. Diese Eigenschaft lässt sich unmittelbar auf den Fall mehrerer unabhängiger Variablen übertragen:

Definition (17.23)

Sei $D \subset \mathbb{R}^n$ offen, $\boldsymbol{x}^0 \in D$ und $\boldsymbol{f}\colon D \to \mathbb{R}^m$. Die Funktion $\boldsymbol{f}$ heißt im Punkt $\boldsymbol{x}^0$ **differenzierbar**, auch **vollständig** oder **total differenzierbar**, falls es eine lineare Abbildung

$$\boldsymbol{\ell}\colon \mathbb{R}^n \to \mathbb{R}^m , \quad \boldsymbol{\ell}(\boldsymbol{x} - \boldsymbol{x}^0) = \boldsymbol{A}\,(\boldsymbol{x} - \boldsymbol{x}^0)$$

mit einer Matrix $\boldsymbol{A} \in \mathbb{R}^{(m,n)}$ gibt, für die die folgende Approximationseigenschaft gilt

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}^0) + \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^0) + o(\|\boldsymbol{x} - \boldsymbol{x}^0\|) ,$$

d. h.

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}^0} \frac{\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}^0) - \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}^0)}{\|\boldsymbol{x}-\boldsymbol{x}^0\|} = \boldsymbol{0}\,.$$

Die lineare Abbildung $\boldsymbol{\ell}$ heißt dann das **Differential** der Funktion $\boldsymbol{f}$ im Punkte $\boldsymbol{x}^0$ und wird mit $\mathbf{d}\boldsymbol{f}(\boldsymbol{x}^0)$ bezeichnet.

Die zugehörige Matrix $\boldsymbol{A}$ ist gerade die in (17.7) definierte **Jacobi-Matrix**, bzw. **Funktionalmatrix** $A = \boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}^0)$.

Bemerkung (17.24)
Im Fall einer skalaren Funktion, also $m = 1$, ist $\boldsymbol{A} = \boldsymbol{a}$ ein Zeilenvektor und $\boldsymbol{a}(\boldsymbol{x}-\boldsymbol{x}^0)$ ein Skalarprodukt: $\langle \boldsymbol{a}^{\mathrm{T}}, \boldsymbol{x}-\boldsymbol{x}^0\rangle$. Wie wir sehen werden, ist $\boldsymbol{a}$ dann gerade der Gradient, $\boldsymbol{a} = \boldsymbol{\nabla} f(\boldsymbol{x}^0)^{\mathrm{T}}$.

Der folgende Satz zeigt, dass das Differential und damit die Jacobi-Matrix eindeutig bestimmt ist, und gibt den Zusammenhang zu den partiellen Ableitungen von $\boldsymbol{f}$ an.

Satz (17.25)
Sei $\boldsymbol{f} : \mathbb{R}^n \supset D \to \mathbb{R}^m$, D offen und $\boldsymbol{x}^0 \in D$.

a) Ist $\boldsymbol{f}$ in $\boldsymbol{x}^0$ differenzierbar, so ist $\boldsymbol{f}$ auch stetig in $\boldsymbol{x}^0$.
b) Ist $\boldsymbol{f}$ in $\boldsymbol{x}^0$ differenzierbar, so ist das Differential und damit auch die Jacobi-Matrix eindeutig bestimmt und durch (17.7) gegeben.
c) Ist $\boldsymbol{f}$ eine C^1-Funktion auf D, vgl. Definition 17.15, so ist $\boldsymbol{f}$ auf D auch (vollständig) differenzierbar.

Beweis.
zu a): Ist $\boldsymbol{f}$ in $\boldsymbol{x}^0$ differenzierbar, so folgt $\lim_{\boldsymbol{x}\to\boldsymbol{x}^0} \|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}^0) - \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}^0)\| = 0$ und damit auch mittels Dreiecksungleichung

$$\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}^0)\| \le \|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}^0) - \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}^0)\| + \|\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}^0)\| \to 0$$

für $\boldsymbol{x} \to \boldsymbol{x}^0$, also $\lim_{\boldsymbol{x}\to\boldsymbol{x}^0} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}^0)$.

zu b): Mit $\boldsymbol{x} = \boldsymbol{x}^0 + t\boldsymbol{e}_i$, $|t| < \varepsilon$ und $i \in \{1, \ldots, n\}$ folgt

$$\begin{aligned}
&\frac{\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}^0) - \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}^0)}{\|\boldsymbol{x}-\boldsymbol{x}^0\|_\infty} \\
&= \frac{\boldsymbol{f}\left(\boldsymbol{x}^0 + t\boldsymbol{e}_i\right) - \boldsymbol{f}(\boldsymbol{x}^0)}{|t|} - \frac{t\boldsymbol{A}\boldsymbol{e}_i}{|t|} \\
&= \frac{t}{|t|} \cdot \left(\frac{\boldsymbol{f}\left(\boldsymbol{x}^0 + t\boldsymbol{e}_i\right) - \boldsymbol{f}(\boldsymbol{x}^0)}{t} - \boldsymbol{A}\boldsymbol{e}_i \right) \\
&\to \boldsymbol{0} \qquad (t \to 0)\,.
\end{aligned}$$

Damit folgt aber auch

$$\lim_{t\to 0} \frac{\boldsymbol{f}\left(\boldsymbol{x}^0 + t\boldsymbol{e}_i\right) - \boldsymbol{f}(\boldsymbol{x}^0)}{t} = \boldsymbol{A}\boldsymbol{e}_i\,, \quad i = 1,\ldots,n\,.$$

Nach Definition 17.15 ist $\boldsymbol{f}$ also partiell differenzierbar, und für die partiellen Ableitungen gilt:

$$\frac{\partial f_k}{\partial x_i}(\boldsymbol{x}^0) = a_{ki}\,, \quad i = 1,\ldots,n\,, \quad k = 1,\ldots,m\,.$$

Damit ist $\boldsymbol{A}$ gerade durch die Jacobi-Matrix (17.7) gegeben.

zu c): Da $\boldsymbol{f}$ nach Definition 17.23 genau dann differenzierbar ist, falls jede Komponentenfunktion f_k differenzierbar ist, genügt es o. B. d. A., die Behauptung für eine skalare Funktion $f : D \to \mathbb{R}$ zu zeigen.
Analog zum Beweis von Satz 17.8 findet man durch das Einschieben sich aufhebender Terme und Anwendung des Mittelwertsatzes

$$\begin{aligned} f(\boldsymbol{x}) - f(\boldsymbol{x}^0) &= f(x_1,\ldots,x_n) - f\left(x_1,\ldots,x_{n-1},x_n^0\right) \\ &\quad + f\left(x_1,\ldots,x_{n-1},x_n^0\right) - f\left(x_1,\ldots,x_{n-1}^0,x_n^0\right) \\ &\quad \vdots \\ &\quad + f\left(x_1,x_2^0,\ldots,x_n^0\right) - f\left(x_1^0,x_2^0,\ldots,x_n^0\right) \\ &= \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\boldsymbol{\xi}_k)\left(x_k - x_k^0\right) \end{aligned}$$

mit $\boldsymbol{\xi}_k := (x_1,\ldots,x_{k-1},x_k^0 + \Theta_k(x_k - x_k^0),x_{k+1}^0,\ldots,x_n^0)^{\mathrm{T}},\ 0 < \Theta_k < 1$.
Damit wird:

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^0) - \nabla f(\boldsymbol{x}^0)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^0) = \sum_{k=1}^{n}\left(\frac{\partial f}{\partial x_k}(\boldsymbol{\xi}_k) - \frac{\partial f}{\partial x_k}(\boldsymbol{x}^0)\right)\left(x_k - x_k^0\right)$$

und mittels Dreiecksungleichung

$$\begin{aligned} &\frac{\left|f(\boldsymbol{x}) - f(\boldsymbol{x}^0) - \nabla f(\boldsymbol{x}^0)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^0)\right|}{\|\boldsymbol{x} - \boldsymbol{x}^0\|_\infty} \\ &\quad = \left|\sum_{k=1}^{n}\left(\frac{\partial f}{\partial x_k}(\boldsymbol{\xi}_k) - \frac{\partial f}{\partial x_k}(\boldsymbol{x}^0)\right)\frac{x_k - x_k^0}{\|\boldsymbol{x} - \boldsymbol{x}^0\|_\infty}\right| \\ &\quad \le \sum_{k=1}^{n}\left|\frac{\partial f}{\partial x_k}(\boldsymbol{\xi}_k) - \frac{\partial f}{\partial x_k}(\boldsymbol{x}^0)\right| \to 0\,, \quad \text{für} \quad \boldsymbol{x} \to \boldsymbol{x}^0\,. \end{aligned}$$

Nach Definition 17.23 ist f damit also differenzierbar und für die Jacobi-Matrix gilt $\boldsymbol{J} f(\boldsymbol{x}^0) = \boldsymbol{\nabla} f(\boldsymbol{x}^0)^{\mathrm{T}}$. ■

Beispiele (17.26)

a) Für $f(x_1, x_2) := x_1 \mathrm{e}^{2x_2}$ erhält man die Jacobi-Matrix:

$$\boldsymbol{J} f(x_1, x_2) = \boldsymbol{\nabla} f(x_1, x_2)^{\mathrm{T}} = \mathrm{e}^{2x_2}\,(1\ ,\ 2x_1)\,.$$

b) Für $\boldsymbol{f}(x_1, x_2, x_3) := \begin{pmatrix} x_1 x_2 x_3 \\ \sin(x_1 + 2x_2 + 3x_3) \end{pmatrix}$ erhält man die Jacobi-Matrix:

$$\boldsymbol{J}\boldsymbol{f}(x_1, x_2, x_3) = \begin{pmatrix} x_2 x_3 & x_1 x_3 & x_1 x_2 \\ \cos s & 2\cos s & 3\cos s \end{pmatrix} \in \mathbb{R}^{(2,3)}\,.$$

Dabei bezeichnet $s := x_1 + 2x_2 + 3x_3$.

c) Für eine affin-lineare Funktion $\boldsymbol{f}(\boldsymbol{x}) := \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ mit $\boldsymbol{A} \in \mathbb{R}^{(m,n)}$ und $\boldsymbol{b} \in \mathbb{R}^m$ erhält man wegen $\frac{\partial \boldsymbol{f}}{\partial x_i} = \boldsymbol{A}\boldsymbol{e}_i$, $i = 1, \ldots, n$,

$$\boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\,, \quad \forall\, \boldsymbol{x} \in \mathbb{R}^n\,.$$

d) Für ein quadratisches Polynom $f(\boldsymbol{x}) := \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + c$ mit $\boldsymbol{A} \in \mathbb{R}^{(n,n)}$, $\boldsymbol{b} \in \mathbb{R}^n$ und $c \in \mathbb{R}$ erhält man mit der verallgemeinerten Produktregel, vgl. hierzu Band 1, Satz 9.23,

$$\begin{aligned} \frac{\partial f}{\partial x_i} &= \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{e}_i + b_i \\ &= \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}})\boldsymbol{e}_i + b_i\,. \end{aligned}$$

Damit ergibt sich für die Jacobi-Matrix

$$\boldsymbol{J} f(\boldsymbol{x}) = \boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}}) + \boldsymbol{b}^{\mathrm{T}}\,.$$

Bemerkung (17.27)

Für eine differenzierbare Funktion $\boldsymbol{f} : \mathbb{R}^n \supset D \to \mathbb{R}^m$ gilt nach Definition 17.23

$$\begin{aligned} \boldsymbol{f}(\boldsymbol{x} + \Delta\boldsymbol{x}) &= \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}) \cdot \Delta\boldsymbol{x} + o(\|\Delta\boldsymbol{x}\|) \\ &= \boldsymbol{f}(\boldsymbol{x}) + \sum_{k=1}^{n} \frac{\partial \boldsymbol{f}}{\partial x_k}(\boldsymbol{x})\Delta x_k + o(\|\Delta\boldsymbol{x}\|)\,. \end{aligned}$$

Vernachlässigt man für kleine Vektoren $\Delta\boldsymbol{x}$ den Fehler $o(\|\Delta\boldsymbol{x}\|)$, so ergibt sich „in erster Näherung“ für die Änderung in $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$:

$$\Delta\boldsymbol{y} := \boldsymbol{f}(\boldsymbol{x} + \Delta\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}) \approx \sum_{k=1}^{n} \frac{\partial \boldsymbol{f}}{\partial x_k}(\boldsymbol{x})\Delta x_k\,. \tag{17.8}$$

Man vergleiche hierzu auch Band 1, Abschn. 10.4, insbesondere Bemerkung 10.39.

Beispiel (17.28)

Die Oberfläche eines quaderförmigen Werkstücks mit den Kantenlängen x, y und z ist gegeben durch $S := 2(xy + xz + yz)$.

Es werde gemessen: $x = 10\,\text{cm},\ \ y = 12\,\text{cm},\ \ z = 20\,\text{cm}$. Die Messgenauigkeit betrage hierbei: $|\Delta x|,\ |\Delta y|,\ |\Delta z| \leq 0.1$ cm.

Mit diesen Daten erhält man für die Oberfläche $S = 1120\,\text{cm}^2$ und für den (absoluten) Fehler in S ergibt sich nach (17.8) in erster Näherung

$$\begin{aligned} &\Delta S \approx 2(y+z)\Delta x + 2(x+z)\Delta y + 2(x+y)\Delta z \\ \Rightarrow \quad &|\Delta S| \leq (64+60+44)\cdot 0.1\,\text{cm}^2 = 16.8\,\text{cm}^2\,. \end{aligned}$$

Damit hat man $S = (1120 \pm 17)\,\text{cm}^2$.

Bemerkung (17.29)
Ausgehend von der Näherung (17.8) $\boldsymbol{\Delta y} \approx \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\boldsymbol{x})\Delta x_k$ schreibt man für die Ableitung einer Funktion $\boldsymbol{f}\colon \mathbb{R}^n \supset D \to \mathbb{R}^m$ im Punkt $\boldsymbol{x}^0 \in D$ (offen)

$$\mathbf{d}\boldsymbol{f}(\boldsymbol{x}^0) = \frac{\partial \boldsymbol{f}}{\partial x_1}(\boldsymbol{x}^0)\,\mathrm{d}x_1 + \ldots + \frac{\partial \boldsymbol{f}}{\partial x_n}(\boldsymbol{x}^0)\,\mathrm{d}x_n\,. \tag{17.9}$$

Die Größen $\mathrm{d}x_1, \ldots, \mathrm{d}x_n$ heißen die **Differentiale** der Koordinaten $x_1, \ldots, x_n$. Ferner ist $\mathbf{d}\boldsymbol{f}(\boldsymbol{x}^0)$ das (vollständige) Differential der Funktion $\boldsymbol{f}$ im Punkt $\boldsymbol{x}^0$, vgl. Definition 17.23.

Interpretiert man die Differentiale $\mathrm{d}x_i$ nun als **Projektionen**

$$\mathrm{d}x_i(\boldsymbol{x} - \boldsymbol{x}^0) := x_i - x_i^0 \qquad (i\text{-te Koordinate})\,,$$

so wird aus der rechten Seite von (17.9) gerade eine lineare Abbildung mit

$$\mathbf{d}\boldsymbol{f}(\boldsymbol{x}^0)(\boldsymbol{x} - \boldsymbol{x}^0) = \sum_{i=1}^{n} \frac{\partial \boldsymbol{f}}{\partial x_i}(\boldsymbol{x}^0)\ \left(x_i - x_i^0\right) = \boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}^0)\cdot(\boldsymbol{x} - \boldsymbol{x}^0)\,.$$

Mit dieser Interpretation stimmt (17.9) also genau mit der Definition 17.23 überein.

Im folgenden Satz fassen wir die wichtigsten Differentiationsregeln zusammen.

Satz (17.30): Differentiationsregeln

a) **Linearität**
Sind $\boldsymbol{f}, \boldsymbol{g}\colon D \to \mathbb{R}^m,\ D \subset \mathbb{R}^n$ (offen), in $\boldsymbol{x}^0 \in D$ differenzierbare Funktionen, so ist auch $(\alpha\boldsymbol{f} + \beta\boldsymbol{g})$ in $\boldsymbol{x}^0$ differenzierbar (α, β reell), und es gilt für die Differentiale $\mathbf{d}(\alpha\boldsymbol{f} + \beta\boldsymbol{g})(\boldsymbol{x}^0) = \alpha\,\mathbf{d}\boldsymbol{f}(\boldsymbol{x}^0) + \beta\,\mathbf{d}\boldsymbol{g}(\boldsymbol{x}^0)$ bzw. für die Jacobi-Matrizen

$$\boldsymbol{J}(\alpha\boldsymbol{f} + \beta\boldsymbol{g})(\boldsymbol{x}^0) = \alpha\boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}^0) + \beta\boldsymbol{J}\boldsymbol{g}(\boldsymbol{x}^0)\,.$$

b) **Kettenregel**
Ist $\boldsymbol{f}\colon D_f \to \mathbb{R}^m$ in $\boldsymbol{x}^0 \in D_f \subset \mathbb{R}^n$ (offen) differenzierbar und ist $\boldsymbol{g}\colon D_g \to \mathbb{R}^k$ differenzierbar in $\boldsymbol{y}^0 := \boldsymbol{f}(\boldsymbol{x}^0) \in D_g \subset \mathbb{R}^m$ (offen), so ist auch die Hintereinanderausführung $\boldsymbol{g} \circ \boldsymbol{f}$ in $\boldsymbol{x}^0$ differenzierbar. Für die Differentiale gilt $\mathbf{d}(\boldsymbol{g} \circ \boldsymbol{f})(\boldsymbol{x}^0) = \mathbf{d}\boldsymbol{g}(\boldsymbol{y}^0) \circ \mathbf{d}\boldsymbol{f}(\boldsymbol{x}^0)$ und für die Jacobi-Matrizen analog

$$\boldsymbol{J}(\boldsymbol{g} \circ \boldsymbol{f})(\boldsymbol{x}^0) = \boldsymbol{J}\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}^0))\cdot\boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}^0)\,.$$

Beweis.
zu a): Dies folgt unmittelbar mit den Grenzwertsätzen aus der Definition 17.23.
zu b): Aufgrund der Voraussetzungen gelten die Approximationseigenschaften:

$$\begin{aligned} \boldsymbol{f}(\boldsymbol{x}) &= \boldsymbol{f}(\boldsymbol{x}^0) + \boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}^0)\,(\boldsymbol{x}-\boldsymbol{x}^0) + \boldsymbol{r}_1(\boldsymbol{x}) \\ \boldsymbol{g}(\boldsymbol{y}) &= \boldsymbol{g}(\boldsymbol{y}^0) + \boldsymbol{J}\boldsymbol{g}(\boldsymbol{y}^0)\,(\boldsymbol{y}-\boldsymbol{y}^0) + \boldsymbol{r}_2(\boldsymbol{y})\,, \end{aligned} \tag{17.10}$$

mit

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}^0} \frac{\boldsymbol{r}_1(\boldsymbol{x})}{\|\boldsymbol{x}-\boldsymbol{x}^0\|} = \boldsymbol{0}\,, \qquad \lim_{\boldsymbol{y}\to\boldsymbol{y}^0} \frac{\boldsymbol{r}_2(\boldsymbol{y})}{\|\boldsymbol{y}-\boldsymbol{y}^0\|} = \boldsymbol{0}\,.$$

Insbesondere folgt aus der ersten Beziehung:

$$\begin{aligned} \|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}^0)\| &= \|\boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}^0)\,(\boldsymbol{x}-\boldsymbol{x}^0) + \boldsymbol{r}_1(\boldsymbol{x})\,\| \\ &\le \|\boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}^0)\|\;\|\boldsymbol{x}-\boldsymbol{x}^0\| + \|\boldsymbol{r}_1(\boldsymbol{x})\| \end{aligned}$$

$$\Rightarrow \quad \frac{\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}^0)\|}{\|\boldsymbol{x}-\boldsymbol{x}^0\|} \le \|\boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}^0)\| + \frac{\|\boldsymbol{r}_1(\boldsymbol{x})\|}{\|\boldsymbol{x}-\boldsymbol{x}^0\|}\,.$$

Hiermit sieht man, dass $\frac{\|\boldsymbol{f}(\boldsymbol{x})-\boldsymbol{f}(\boldsymbol{x}^0)\|}{\|\boldsymbol{x}-\boldsymbol{x}^0\|}$ für $\boldsymbol{x}\to\boldsymbol{x}^0$ beschränkt ist.
Setzt man nun die erste der Beziehungen (17.10) in die zweite ein, so erhält man:

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x})) = \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}^0)) &+ \boldsymbol{J}\boldsymbol{g}(\boldsymbol{y}^0)\cdot\boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}^0)\cdot(\boldsymbol{x}-\boldsymbol{x}^0) \\ &+ \underbrace{\boldsymbol{J}\boldsymbol{g}(\boldsymbol{y}^0)\cdot\boldsymbol{r}_1(\boldsymbol{x}) + \boldsymbol{r}_2(\boldsymbol{y})}_{=:\,\boldsymbol{r}(\boldsymbol{x})}\,. \end{aligned}$$

Hieraus folgt für $\boldsymbol{x}\ne\boldsymbol{x}^0$ und $\boldsymbol{y} := \boldsymbol{f}(\boldsymbol{x})$:

$$\begin{aligned} \frac{\boldsymbol{r}(\boldsymbol{x})}{\|\boldsymbol{x}-\boldsymbol{x}^0\|} &= \boldsymbol{J}\boldsymbol{g}(\boldsymbol{y}^0)\frac{\boldsymbol{r}_1(\boldsymbol{x})}{\|\boldsymbol{x}-\boldsymbol{x}^0\|} + \frac{\boldsymbol{r}_2(\boldsymbol{y})}{\|\boldsymbol{x}-\boldsymbol{x}^0\|} \\ &= \begin{cases} \boldsymbol{J}\boldsymbol{g}(\boldsymbol{y}^0)\dfrac{\boldsymbol{r}_1(\boldsymbol{x})}{\|\boldsymbol{x}-\boldsymbol{x}^0\|}\,, & \text{falls} \quad \boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{y}^0\,, \\ \boldsymbol{J}\boldsymbol{g}(\boldsymbol{y}^0)\dfrac{\boldsymbol{r}_1(\boldsymbol{x})}{\|\boldsymbol{x}-\boldsymbol{x}^0\|} + \dfrac{\boldsymbol{r}_2(\boldsymbol{y})}{\|\boldsymbol{y}-\boldsymbol{y}^0\|}\cdot\dfrac{\|\boldsymbol{f}(\boldsymbol{x})-\boldsymbol{f}(\boldsymbol{x}^0)\|}{\|\boldsymbol{x}-\boldsymbol{x}^0\|}\,, & \text{sonst.} \end{cases} \end{aligned}$$

Wegen der Stetigkeit von $\boldsymbol{f}$ in $\boldsymbol{x}^0$ geht auch $\boldsymbol{y}\to\boldsymbol{y}^0$ für $\boldsymbol{x}\to\boldsymbol{x}^0$. Damit ergibt sich:

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}^0} \frac{\boldsymbol{r}(\boldsymbol{x})}{\|\boldsymbol{x}-\boldsymbol{x}^0\|} = \boldsymbol{0}\,.$$

Insgesamt ist hiermit gezeigt, dass $(\boldsymbol{g}\circ\boldsymbol{f})$ in $\boldsymbol{x}^0$ differenzierbar ist und dass für die Jacobi-Matrix $\boldsymbol{J}(\boldsymbol{g}\circ\boldsymbol{f})(\boldsymbol{x}^0) = \boldsymbol{J}\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}^0))\cdot\boldsymbol{J}\boldsymbol{f}(\boldsymbol{x}^0)$ gilt. ■

Bemerkung (17.31)
Mitunter wird die Kettenregel auf den folgenden Spezialfall angewendet:

Es sei $\boldsymbol{c}\colon I \to D \subset \mathbb{R}^n$ eine in $t_0 \in I$ differenzierbare Kurve. Dabei ist $I \subset \mathbb{R}$ ein Intervall und $D \subset \mathbb{R}^n$ eine offene Menge. Ferner sei $f : D \to \mathbb{R}$ eine in $\boldsymbol{x}^0 := \boldsymbol{c}(t_0)$ differenzierbare skalare Funktion.

Aufgrund der Kettenregel ist dann auch die Hintereinanderausführung

$$f \circ \boldsymbol{c}\colon I \to \mathbb{R}\,, \quad (f \circ \boldsymbol{c})(t) = f(c_1(t), \ldots, c_n(t))$$

in t_0 differenzierbar und für die Ableitung gilt

$$(f \circ \boldsymbol{c})'(t_0) = \boldsymbol{J}f(\boldsymbol{c}(t_0)) \cdot \boldsymbol{J}\boldsymbol{c}(t_0) = \boldsymbol{\nabla} f(\boldsymbol{c}(t_0))^{\mathrm{T}} \boldsymbol{c}'(t_0) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\boldsymbol{c}(t_0)) \cdot c_k'(t_0)\,.$$

Beispiel (17.32)
Ein Massenpunkt bewege sich mit der zeitlich veränderlichen Masse $m(t) = m_0 - \alpha t$ und der (skalaren) Geschwindigkeit $v(t) = v_0 + \beta t^3$. Wir fragen nach der zeitlichen Änderung der kinetischen Energie $T = \frac{1}{2} m v^2$.

Natürlich könnte man die obigen Funktionen m und v einsetzen und die resultierende Funktion $T(t)$ direkt differenzieren. Man kann aber stattdessen auch die Kettenregel anwenden:

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \boldsymbol{\nabla} T(m, v)^{\mathrm{T}} \cdot \begin{pmatrix} m'(t) \\ v'(t) \end{pmatrix} = \left(\frac{1}{2} v^2,\ mv\right) \begin{pmatrix} -\alpha \\ 3\beta t^2 \end{pmatrix} = 3\beta m v t^2 - \frac{1}{2}\alpha v^2\,.$$

Wenn man hierin nun die Funktionen m und v einsetzt, ergibt sich:

$$\frac{\mathrm{d}T}{\mathrm{d}t} = 3\beta(m_0 - \alpha t)\,\left(v_0 + \beta t^3\right) t^2 - \frac{\alpha}{2}\,\left(v_0 + \beta t^3\right)^2\,.$$

Richtungsableitungen

Bei der Definition der partiellen Ableitungen in Definition 17.1 haben wir eine Funktion f von mehreren Variablen „längs einer Koordinatenrichtung“ eingeschränkt und diese partielle Funktion einer Variablen $t \mapsto f(\boldsymbol{x}^0 + t\boldsymbol{e}_i)$ im Punkt $t = 0$ differenziert.

Es ist klar, dass man die Einschränkung auf eindimensionale Teilbereiche des Definitionsbereichs nicht nur längs der Koordinatenrichtungen $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ vornehmen muss, sondern allgemeiner eine Funktion von mehreren Veränderlichen auch „längs beliebiger Richtungen“ $t \mapsto f(\boldsymbol{x}^0 + t\boldsymbol{v})$ differenzieren kann. Dies ist im Übrigen ein Spezialfall der in Bemerkung 17.31 beschriebenen Situation.

Definition (17.33)
Für eine Funktion $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ offen, $\boldsymbol{x}^0 \in D$ und einen Vektor $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ heißt

$$D_{\boldsymbol{v}} f(\boldsymbol{x}^0) := \lim_{t \to 0} \frac{f(\boldsymbol{x}^0 + t\boldsymbol{v}) - f(\boldsymbol{x}^0)}{t}$$

die **Richtungsableitung**, auch **Gâteaux**[4]**–Ableitung**, von f in Richtung $\boldsymbol{v}$.

4) René Gâteaux (1889–1914); Bar-le-Duc, Rom.

Bemerkungen (17.34)

a) Die obige Definition lässt sich auf Funktionale $f\colon V \supset D \to \mathbb{R}$ von allgemeinen normierten Vektorräumen V übertragen. In diesem Zusammenhang wird die Richtungsableitung auch mit $\delta f(\boldsymbol{x}^0, \boldsymbol{v})$ bezeichnet und **erste Variation** von f im Punkt $\boldsymbol{x}^0$ und in Richtung $\boldsymbol{v}$ genannt.

b) Ist der Richtungsvektor $\boldsymbol{v}$ ein Einheitsvektor, also $\|\boldsymbol{v}\| = 1$, so beschreibt $D_{\boldsymbol{v}} f(\boldsymbol{x}^0)$ den Anstieg bzw. die Steigung der Funktion f im Punkt $\boldsymbol{x}^0$ in Richtung $\boldsymbol{v}$.

c) Wählt man als Richtungsvektoren $\boldsymbol{v}$ die kanonischen Einheitsvektoren $\boldsymbol{e}_i$ in den Koordinatenrichtungen, so erhält man für die Richtungsableitungen gerade die partiellen Ableitungen von f:

$$D_{\boldsymbol{e}_i} f(\boldsymbol{x}^0) = \frac{\partial f}{\partial x_i}(\boldsymbol{x}^0)\,,\ i = 1, \ldots, n\,.$$

d) Ist f im Punkt $\boldsymbol{x}^0$ differenzierbar, so gilt

$$D_{\boldsymbol{v}} f(\boldsymbol{x}^0) = \frac{\mathrm{d}}{\mathrm{d}t}(f \circ \boldsymbol{c})(0)$$

mit $\boldsymbol{c}(t) := \boldsymbol{x}^0 + t\boldsymbol{v}$. Da $\boldsymbol{c}$ als affin-lineare Funktion differenzierbar ist, existieren sämtliche Richtungsableitungen von f in $\boldsymbol{x}^0$ und mit der Kettenregel folgt

$$D_{\boldsymbol{v}} f(\boldsymbol{x}^0) = \boldsymbol{\nabla} f(\boldsymbol{x}^0)^{\mathrm{T}} \boldsymbol{c}'(0) = \boldsymbol{\nabla} f(\boldsymbol{x}^0)^{\mathrm{T}} \boldsymbol{v}\,. \tag{17.11}$$

Es sei nochmals darauf hingewiesen, vgl. Satz 17.25, dass die Voraussetzung der Differenzierbarkeit von f erfüllt ist, falls f stetig partiell differenzierbar ist. Die bloße partielle Differenzierbarkeit von f genügt hierfür nicht.

Satz (17.35): Eigenschaften des Gradienten

Sei $f: D \to \mathbb{R}$ mit $D \subset \mathbb{R}^n$ offen im Punkt $\boldsymbol{x}^0 \in D$ differenzierbar.

a) Der Gradientenvektor $\boldsymbol{\nabla} f(\boldsymbol{x}^0) \in \mathbb{R}^n$ steht senkrecht auf der **Niveaumenge**

$$N_f(\boldsymbol{x}^0) := \{\boldsymbol{x} \in D : f(\boldsymbol{x}) = f(\boldsymbol{x}^0)\}\,.$$

Im Fall $n = 2$ nennt man die Niveaumengen auch **Höhenlinien**, im Fall $n = 3$ **Äquipotentialflächen**, vgl. Abb. 17.2.

b) $\boldsymbol{\nabla} f(\boldsymbol{x}^0)$ gibt die **Richtung des steilsten Anstiegs** von f im Punkt $\boldsymbol{x}^0$ an.

Beweis.

zu a): Sei $\boldsymbol{c}(t)$, $-\varepsilon < t < \varepsilon$, eine differenzierbare Kurve mit $\boldsymbol{c}(0) = \boldsymbol{x}^0$, die ganz in der Niveaumenge $N_f(\boldsymbol{x}^0)$ verläuft. Die Existenz solcher Kurven wird später für differenzierbare und in $\boldsymbol{x}^0$ reguläre Funktionen mit Hilfe des Satzes über implizite Funktionen sichergestellt.
Nach Definition der Niveaumenge $N_f(\boldsymbol{x}^0)$ gilt dann

$$f(\boldsymbol{c}(t)) = f(\boldsymbol{x}^0) = \text{const.}\,, \quad \forall\, t \in\,]-\varepsilon, \varepsilon[\,.$$

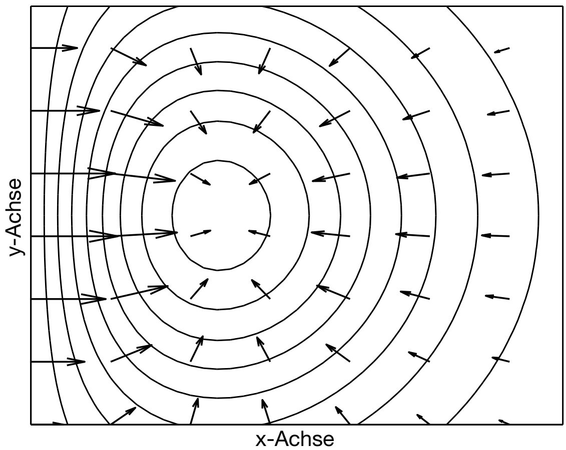

Abb. 17.2 Höhenlinien und Gradienten von $f(x, y) = x \cdot e^{x^2+y^2}$.

Differenziert man diese Identität nach t mittels Kettenregel, so folgt in $t = 0$:

$$\nabla f(\boldsymbol{x}^0)^{\mathrm{T}} \boldsymbol{c}'(0) = 0\,.$$

Dies bedeutet gerade geometrisch, dass der Gradient $\nabla f(\boldsymbol{x}^0)$ auf allen Tangentialvektoren $\boldsymbol{c}'(0)$ an die Niveaumenge $N_f(\boldsymbol{x}^0)$ in $\boldsymbol{x}^0$ senkrecht steht.

zu b): Die Steigung von f in Richtung eines Einheitsvektors $\boldsymbol{v}$, $\|\boldsymbol{v}\| = 1$, ist durch die Richtungsableitung

$$D_{\boldsymbol{v}} f(\boldsymbol{x}^0) = \nabla f(\boldsymbol{x}^0)^{\mathrm{T}} \boldsymbol{v}$$

gegeben. Die Cauchy-Schwarzsche Ungleichung für das obige Skalarprodukt ergibt die Abschätzung

$$|D_{\boldsymbol{v}} f(\boldsymbol{x}^0)| \le \|\nabla f(\boldsymbol{x}^0)\|_2\,.$$

Ist $\nabla f(\boldsymbol{x}^0) = \boldsymbol{0}$, so verschwinden folglich auch alle Steigungen $D_{\boldsymbol{v}} f(\boldsymbol{x}^0)$, d. h., $\boldsymbol{x}^0$ ist ein **stationärer Punkt** von f.

Im Fall $\nabla f(\boldsymbol{x}^0) \neq \boldsymbol{0}$ werden die Schranken der obigen Abschätzung $\pm\|\nabla f(\boldsymbol{x}^0)\|_2$ auch angenommen, und zwar wird für die Richtungsvektoren

$$\boldsymbol{v} = \pm \frac{\nabla f(\boldsymbol{x}^0)}{\|\nabla f(\boldsymbol{x}^0)\|_2}\,.$$

Im Fall des positiven Vorzeichens wird $D_{\boldsymbol{v}} f(\boldsymbol{x}^0)$ maximal, im Fall des negativen Vorzeichens minimal.

$\nabla f(\boldsymbol{x}^0)$ weist also in Richtung des stärksten Anstiegs von f und $-\nabla f(\boldsymbol{x}^0)$ in Richtung des stärksten Abstiegs von f. ■

Bemerkungen (17.36)
Die obige Eigenschaft des Gradienten 17.35b) lässt sich vorteilhaft für die numerische Berechnung der lokalen Minima einer Funktion f verwenden. Auf dieser Eigenschaft beruhen im Wesentlichen die verschiedenen Varianten des **Gradientenverfahrens**, die auch **Verfahren des steilsten Abstiegs (steepest descent methods)** genannt werden.

Krummlinige Koordinaten

Die Kettenregel ist ein wichtiges Hilfsmittel, um Ableitungsausdrücke einer Funktion f auf neue Koordinatensysteme umzurechnen. Dabei werden neben den linearen Koordinatentransformationen, die wir bereits in Kap. 5 des ersten Bandes kennengelernt haben, auch allgemeinere nichtlineare Transformationen zugelassen. Man spricht dann von krummlinigen Koordinatensystemen.

Dazu sei $\boldsymbol{x} = \boldsymbol{\Phi}(\boldsymbol{u})$ eine C^1-Abbildung $\boldsymbol{\Phi}\colon U \to V$ zweier offener Bereiche $U, V \subset \mathbb{R}^n$. Wir nehmen an, dass die Jacobi-Matrix $\boldsymbol{J\Phi}(\boldsymbol{u}^0)$ an jeder Stelle $\boldsymbol{u}^0 \in U$ regulär ist. Wie wir später sehen werden, ist $\boldsymbol{\Phi}$ dann *lokal* bei $\boldsymbol{u}^0$ eine bijektive Transformation.

Durch eventuelle Einschränkung von U und V lässt sich also erreichen, dass $\boldsymbol{\Phi}\colon U \to V$ bijektiv ist. Ferner sei auch die Umkehrabbildung $\boldsymbol{\Phi}^{-1}\colon V \to U$ eine C^1-Abbildung. Auch diese Eigenschaft lässt sich aus der Regularität der Jacobi-Matrix ableiten.

Wie in Abb. 17.3 angedeutet, lässt sich die Wirkung der Koordinatentransformation $\boldsymbol{x} = \boldsymbol{\Phi}(\boldsymbol{u})$ für $n = 2$ dadurch veranschaulichen, dass man die Bilder der Koordinatenlinien $u_i = \text{const.}$, $i = 1, 2$, betrachtet. Beispielsweise erhält man für Polarkoordinaten

$$x_1 = r\cos\varphi\,, \quad x_2 = r\sin\varphi\,, \quad r > 0\,, \quad -\pi < \varphi < \pi$$

als Bilder der Koordinatenlinien $r = \text{const.}$ konzentrische Kreise um den Ursprung und als Bilder der Koordinatenlinien $\varphi = \text{const.}$ Halbgeraden.

Wir interessieren uns nun dafür, wie sich Ableitungsausdrücke von Funktionen bei Koordinatentransformationen $\boldsymbol{x} = \boldsymbol{\Phi}(\boldsymbol{u})$ auf die neuen Koordinaten $\boldsymbol{u}$ umrechnen lassen.

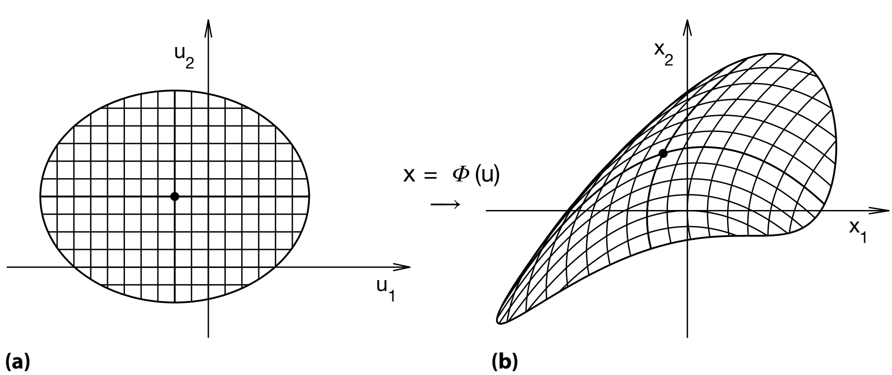

Abb. 17.3 Nichtlineare Koordinatentransformation.

Allgemein gilt nun für alle $\boldsymbol{u} \in U$ die Identität $\boldsymbol{\Phi}^{-1}(\boldsymbol{\Phi}(\boldsymbol{u})) = \boldsymbol{u}$. Diese Identität wird nun nach $\boldsymbol{u}$ differenziert. Mit der Kettenregel erhält man

$$\boldsymbol{J\Phi}^{-1}(\boldsymbol{x}) \cdot \boldsymbol{J\Phi}(\boldsymbol{u}) = \boldsymbol{I}_n , \quad \boldsymbol{x} = \boldsymbol{\Phi}(\boldsymbol{u})$$

und damit

$$\boldsymbol{J\Phi}^{-1}(\boldsymbol{x}) = (\boldsymbol{J\Phi}(\boldsymbol{u}))^{-1} , \quad \boldsymbol{x} = \boldsymbol{\Phi}(\boldsymbol{u}) . \tag{17.12}$$

Sei nun $\widetilde{f}$ eine Funktion in der Variablen $\boldsymbol{x}$, $\widetilde{f}\colon V \to \mathbb{R}$ und es bezeichne $f(\boldsymbol{u}) := \widetilde{f}(\boldsymbol{\Phi}(\boldsymbol{u}))$ die „gleiche" Funktion, allerdings ausgedrückt in den neuen Koordinaten $\boldsymbol{u}$.

Die Differentiation dieser Funktion mittels Kettenregel ergibt:

$$\frac{\partial f}{\partial u_i} = \sum_{j=1}^{n} \frac{\partial \widetilde{f}}{\partial x_j} \cdot \frac{\partial \Phi_j}{\partial u_i} = \sum_{j=1}^{n} g^{ij} \cdot \frac{\partial \widetilde{f}}{\partial x_j} .$$

Dabei wird die Matrix $\boldsymbol{G}(\boldsymbol{u}) := (g^{ij}) \in \mathbb{R}^{(n,n)}$ definiert durch

$$g^{ij} := \frac{\partial \Phi_j}{\partial u_i} , \quad \boldsymbol{G}(\boldsymbol{u}) := (g^{ij}) = (\boldsymbol{J\Phi}(\boldsymbol{u}))^{\mathrm{T}} . \tag{17.13}$$

Für die obige Beziehung schreiben wir auch abkürzend:

$$\frac{\partial}{\partial u_i} = \sum_{j=1}^{n} g^{ij} \frac{\partial}{\partial x_j} \quad \text{oder} \quad \boldsymbol{\nabla}_{\boldsymbol{u}} = \boldsymbol{G}\boldsymbol{\nabla}_{\boldsymbol{x}} . \tag{17.14}$$

Die Umkehrung dieser Relationen erhält man durch eine analoge Untersuchung für $\boldsymbol{\Phi}^{-1}$ anstelle von $\boldsymbol{\Phi}$:

$$\frac{\partial}{\partial x_i} = \sum_{j=1}^{n} g_{ij} \frac{\partial}{\partial u_j} \quad \text{oder} \quad \boldsymbol{\nabla}_{\boldsymbol{x}} = \boldsymbol{G}^{-1}\boldsymbol{\nabla}_{\boldsymbol{u}} . \tag{17.15}$$

wobei

$$(g_{ij}) := (g^{ij})^{-1} = (\boldsymbol{J\Phi})^{-\mathrm{T}} = (\boldsymbol{J\Phi}^{-1})^{\mathrm{T}} . \tag{17.16}$$

Beispiel (17.37)
Das Polarkoordinatensystem ist gegeben durch die Transformation

$$\boldsymbol{x} := \boldsymbol{\Phi}(\boldsymbol{u}) := \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix} , \qquad \begin{array}{l} \boldsymbol{u} = (r, \varphi) , \\ r > 0 , \ \ -\pi < \varphi < \pi . \end{array}$$

Damit erhält man für die Jacobi-Matrix von $\boldsymbol{\Phi}$:

$$\boldsymbol{J\Phi}(\boldsymbol{u}) = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix} ,$$

und somit

$$(g^{ij}) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -r\sin\varphi & r\cos\varphi \end{pmatrix}, \quad (g_{ij}) = \begin{pmatrix} \cos\varphi & -\frac{1}{r}\sin\varphi \\ \sin\varphi & \frac{1}{r}\cos\varphi \end{pmatrix}.$$

Nach (17.15) ergibt sich also für die Umrechnung der partiellen Ableitungen:

$$\begin{aligned} \frac{\partial}{\partial x_1} &= \cos\varphi\frac{\partial}{\partial r} - \frac{1}{r}\sin\varphi\frac{\partial}{\partial\varphi} \\ \frac{\partial}{\partial x_2} &= \sin\varphi\frac{\partial}{\partial r} + \frac{1}{r}\cos\varphi\frac{\partial}{\partial\varphi}\,. \end{aligned} \tag{17.17}$$

Beispiel (17.38)
Gegeben sei die Funktion $\widetilde{f}(x,y) := y\sqrt{x^2+y^2}$. Die Umrechnung in Polarkoordinaten ergibt $f(r,\varphi) = r^2\sin\varphi$ und damit die partiellen Ableitungen $f_r = 2r\sin\varphi$ und $f_\varphi = r^2\cos\varphi$. Die Anwendung von (17.17) ergibt nun

$$\begin{aligned} \widetilde{f}_x &= \cos(\varphi)f_r - \frac{1}{r}\sin(\varphi)f_\varphi = \frac{xy}{r} \\ \widetilde{f}_y &= \sin(\varphi)f_r + \frac{1}{r}\cos(\varphi)f_\varphi = r + \frac{y^2}{r}\,. \end{aligned}$$

Diese Ergebnisse stimmen offenbar mit den Resultaten, die man durch direkte partielle Differentiation von $\widetilde{f}$ erhält, überein.

Die Beziehung (17.17) lässt sich verwenden, um auch andere Differentialoperatoren des $\mathbb{R}^2$ auf Polarkoordinaten umzurechnen. So erhält man beispielsweise für die zweiten partiellen Ableitungen:

$$\begin{aligned} \frac{\partial^2}{\partial x_1^2} &= \cos^2(\varphi)\frac{\partial^2}{\partial r^2} - \frac{\sin(2\varphi)}{r}\frac{\partial^2}{\partial r\partial\varphi} + \frac{\sin^2\varphi}{r^2}\frac{\partial^2}{\partial\varphi^2} + \frac{\sin(2\varphi)}{r^2}\frac{\partial}{\partial\varphi} + \frac{\sin^2\varphi}{r}\frac{\partial}{\partial r} \\ \frac{\partial^2}{\partial x_2^2} &= \sin^2\varphi\frac{\partial^2}{\partial r^2} + \frac{\sin(2\varphi)}{r}\frac{\partial^2}{\partial r\partial\varphi} + \frac{\cos^2\varphi}{r^2}\frac{\partial^2}{\partial\varphi^2} - \frac{\sin(2\varphi)}{r^2}\frac{\partial}{\partial\varphi} + \frac{\cos^2\varphi}{r}\frac{\partial}{\partial r} \end{aligned}$$

Damit ergibt sich die folgende **Umrechnungsformel für den Laplace-Operator in Polarkoordinaten**:

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} = \frac{\partial^2}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2} + \frac{1}{r}\frac{\partial}{\partial r}\,. \tag{17.18}$$

Man vergleiche hierzu auch das Beispiel 17.14.

Beispiel (17.39)
Die Kugelkoordinaten des $\mathbb{R}^3$ sind gegeben durch die Koordinatentransformation

$$\boldsymbol{x} = \boldsymbol{\Phi}(\boldsymbol{u}) = \begin{pmatrix} r\cos\varphi\cos\theta \\ r\sin\varphi\cos\theta \\ r\sin\theta \end{pmatrix}, \qquad \begin{aligned} &\boldsymbol{u} = (r,\ \varphi,\ \theta)\,, \\ &r > 0\,, \quad -\pi < \varphi < \pi\,, \\ &-\pi/2 < \theta < \pi/2\,. \end{aligned}$$

Für die Jacobi-Matrix und die Transformationsmatrizen $(\boldsymbol{J\Phi}(u))^{\mathrm{T}}$ und $(\boldsymbol{J\Phi}(u))^{-\mathrm{T}}$ ergibt sich

$$
\begin{aligned}
\boldsymbol{J\Phi}(\boldsymbol{u}) &= \begin{pmatrix} \cos\varphi\cos\theta & -r\sin\varphi\cos\theta & -r\cos\varphi\sin\theta \\ \sin\varphi\cos\theta & r\cos\varphi\cos\theta & -r\sin\varphi\sin\theta \\ \sin\theta & 0 & r\cos\theta \end{pmatrix}, \\
(g^{ij}) &= (\boldsymbol{J\Phi})(\boldsymbol{u}))^{\mathrm{T}}, \\
(g_{ij}) &= \begin{pmatrix} \cos\varphi\cos\theta & -\dfrac{\sin\varphi}{r\cos\theta} & -\dfrac{1}{r}\cos\varphi\sin\theta \\ \sin\varphi\cos\theta & \dfrac{\cos\varphi}{r\cos\theta} & -\dfrac{1}{r}\sin\varphi\sin\theta \\ \sin\theta & 0 & \dfrac{1}{r}\cos\theta \end{pmatrix}.
\end{aligned}
$$

Mit (17.17) folgt dann für die partiellen Ableitungen erster Ordnung:

$$
\begin{aligned}
\frac{\partial}{\partial x_1} &= \cos\varphi\cos\theta\frac{\partial}{\partial r} - \frac{\sin\varphi}{r\cos\theta}\frac{\partial}{\partial\varphi} - \frac{1}{r}\cos\varphi\sin\theta\frac{\partial}{\partial\theta} \\
\frac{\partial}{\partial x_2} &= \sin\varphi\cos\theta\frac{\partial}{\partial r} + \frac{\cos\varphi}{r\cos\theta}\frac{\partial}{\partial\varphi} - \frac{1}{r}\sin\varphi\sin\theta\frac{\partial}{\partial\theta} \\
\frac{\partial}{\partial x_3} &= \sin\theta\frac{\partial}{\partial r} + \frac{1}{r}\cos\theta\frac{\partial}{\partial\theta}\,.
\end{aligned}
\tag{17.19}
$$

Mit einer etwas aufwendigeren Rechnung findet man mit Hilfe der obigen Beziehungen die Umrechnung des dreidimensionalen Laplace-Operators in Kugelkoordinaten:

$$
\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r^2\cos^2\theta}\frac{\partial^2}{\partial\varphi^2} + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2} + \frac{2}{r}\frac{\partial}{\partial r} - \frac{\tan\theta}{r^2}\frac{\partial}{\partial\theta}\,. \tag{17.20}
$$

17.3 Mittelwertsätze und Taylorscher Satz

Wir betrachten zunächst die Verallgemeinerung des ersten Mittelwertsatzes, vgl. Satz 10.8, auf skalare Funktionen von mehreren unabhängigen Variablen.

Satz (17.40): Mittelwertsatz
Sei $f : D \to \mathbb{R}$ eine auf einer offenen Menge $D \subset \mathbb{R}^n$ differenzierbare, skalare Funktion. Ferner seien $\boldsymbol{a}, \boldsymbol{b} \in D$ zwei Punkte in D, für die die Verbindungsstrecke

$$[\boldsymbol{a}, \boldsymbol{b}] := \{\boldsymbol{a} + t(\boldsymbol{b} - \boldsymbol{a}) : t \in [0, 1]\}$$

ganz in D liegt. Dann gibt es eine Zahl $\theta \in\,]0, 1[$ mit

$$f(\boldsymbol{b}) - f(\boldsymbol{a}) = \boldsymbol{\nabla} f(\boldsymbol{a} + \theta(\boldsymbol{b} - \boldsymbol{a}))^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{a})\,.$$

Beweis.
Mit einer geeigneten Parametrisierung der Verbindungsstrecke $[\boldsymbol{a}, \boldsymbol{b}]$ lässt sich die Behauptung sofort auf den Fall einer unabhängigen Veränderlichen zurückführen. Dazu definieren wir die Verbindungsstrecke

$$\boldsymbol{c}(t) := \boldsymbol{a} + t(\boldsymbol{b} - \boldsymbol{a}) , \quad 0 \leq t \leq 1 .$$

Die Komposition $h(t) := f(\boldsymbol{c}(t))$ ist stetig auf $[0, 1]$ und nach der Kettenregel 17.30 auch differenzierbar auf dem offenen Intervall $]0, 1[$. Damit sind die Voraussetzungen des ersten Mittelwertsatzes 10.8 für skalare Funktionen einer Variablen erfüllt, und wir erhalten

$$\begin{aligned} f(\boldsymbol{b}) - f(\boldsymbol{a}) &= h(1) - h(0) \\ &= h'(\theta) \cdot (1 - 0) , \quad 0 < \theta < 1 \\ &= \boldsymbol{\nabla} f(\boldsymbol{a} + \theta(\boldsymbol{b} - \boldsymbol{a}))^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{a}) . \end{aligned}$$

■

Bemerkungen (17.41)
Gilt die Voraussetzung des Mittelwertsatzes $[\boldsymbol{a}, \boldsymbol{b}] \subset D$ für *alle* Punktepaare $\boldsymbol{a},\ \boldsymbol{b} \in D$, so heißt die Menge D **konvex**, vgl. Abb. 17.4.

Konvexe Mengen und konvexe Funktionen, vgl. auch Abschn. 10.1, spielen innerhalb der angewandten Analysis, insbesondere in der Optimierungstheorie, eine wichtige Rolle.

Bei konvexen Definitionsbereichen gilt der Mittelwertsatz also nach Definition für beliebige Punkte $\boldsymbol{a},\ \boldsymbol{b} \in D$.

Beispiel (17.42)
Gegeben ist die Funktion $f(x_1, x_2) := \cos x_1 + \sin x_2$. Es gilt $f(0, 0) = f(\pi/2, \pi/2) = 1$.

Nach dem Mittelwertsatz muss es daher ein $\theta \in\ ]0, 1[$ geben mit der Eigenschaft

$$\begin{aligned} 0 &= \boldsymbol{\nabla} f\left(\theta \begin{pmatrix} \pi/2 \\ \pi/2 \end{pmatrix}\right)^{\mathrm{T}} \begin{pmatrix} \pi/2 \\ \pi/2 \end{pmatrix} \\ &= \left(-\sin\left(\theta\frac{\pi}{2}\right),\ \cos\left(\theta\frac{\pi}{2}\right)\right) \begin{pmatrix} \pi/2 \\ \pi/2 \end{pmatrix} \\ &= \frac{\pi}{2}\left(\cos\left(\theta\frac{\pi}{2}\right) - \sin\left(\theta\frac{\pi}{2}\right)\right) . \end{aligned}$$

In der Tat ist diese Gleichung für $\theta := \frac{1}{2}$ erfüllt.

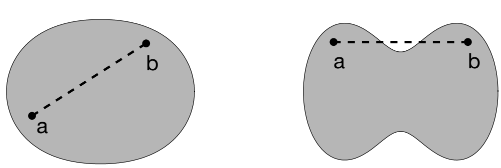

Abb. 17.4 Konvexe und nichtkonvexe Menge.

Leider lässt sich der Mittelwertsatz in obiger Gleichungsform *nicht* auf den Fall vektorwertiger Funktionen übertragen. Dies liegt im Wesentlichen daran, dass der Mittelwertsatz für jede Komponente einer vektorwertigen Funktion die Existenz eines entsprechenden Zwischenwertes θ garantiert. Man kann jedoch nicht erwarten, für alle Komponenten ein universelles θ zu finden. Das folgende Beispiel verdeutlicht dies.

Beispiel (17.43)
Gegeben ist die Funktion $\boldsymbol{f}(t) := (\cos t,\ \sin t)^{\mathrm{T}}$ auf dem Intervall $0 \le t \le \pi/2$. Es gilt

$$\boldsymbol{f}\left(\frac{\pi}{2}\right) - \boldsymbol{f}(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} .$$

Würde der Mittelwertsatz für vektorwertige Funktionen gelten, müsste es demnach ein $\theta \in\ ]0,1[$ geben mit

$$\boldsymbol{f}'\left(\theta\frac{\pi}{2}\right) \cdot \left(\frac{\pi}{2} - 0\right) = \frac{\pi}{2}\begin{pmatrix} -\sin(\theta\pi/2) \\ \cos(\theta\pi/2) \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} .$$

Die Gleichheit dieser beiden Vektoren ist jedoch für kein θ erfüllt, da der erste Vektor die Länge $\pi/2$, der zweite aber die Länge $\sqrt{2}$ besitzt.

Anstelle des Mittelwertsatzes gibt es für vektorwertige Funktionen jedoch eine etwas schwächere Variante, den sogenannten Mittelwert-Abschätzungssatz.

Satz (17.44): Mittelwert-Abschätzungssatz
Die Funktion $\boldsymbol{f} : D \to \mathbb{R}^m$ sei differenzierbar auf der offenen Menge $D \subset \mathbb{R}^n$. Ferner seien $\boldsymbol{a}, \boldsymbol{b}$ Punkte in D mit $[\boldsymbol{a}, \boldsymbol{b}] \subset D$.

Dann existiert ein $\theta \in\ ]0,1[$ mit

$$\|\boldsymbol{f}(\boldsymbol{b}) - \boldsymbol{f}(\boldsymbol{a})\|_2 \le \|\boldsymbol{J}\boldsymbol{f}(\boldsymbol{a} + \theta(\boldsymbol{b} - \boldsymbol{a})) \cdot (\boldsymbol{b} - \boldsymbol{a})\|_2 .$$

Beweis.
Zu einem festen Vektor $\boldsymbol{v} \in \mathbb{R}^m \setminus \{\boldsymbol{0}\}$ definieren wir die skalare Hilfsfunktion $g(\boldsymbol{x}) := \boldsymbol{v}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x})$. Die Funktion g ist dann auf D differenzierbar mit

$$\boldsymbol{\nabla} g(\boldsymbol{x}) = \boldsymbol{J}\boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{v} .$$

Die Anwendung des Mittelwertsatzes 17.40 auf g liefert:

$$\exists\, \theta \in\ ]0,1[:\quad g(\boldsymbol{b}) - g(\boldsymbol{a}) = \boldsymbol{\nabla} g(\boldsymbol{a} + \theta(\boldsymbol{b} - \boldsymbol{a}))^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{a}) .$$

Hierin wird nun g und $\boldsymbol{\nabla} g$ eingesetzt:

$$\boldsymbol{v}^{\mathrm{T}}(\boldsymbol{f}(\boldsymbol{b}) - \boldsymbol{f}(\boldsymbol{a})) = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{J}\boldsymbol{f}(\boldsymbol{a} + \theta(\boldsymbol{b} - \boldsymbol{a})) \cdot (\boldsymbol{b} - \boldsymbol{a}) .$$

Mit Hilfe der Cauchy-Schwarzschen Ungleichung, vgl. Satz 3.8, folgt hieraus:

$$|\ \boldsymbol{v}^{\mathrm{T}}(\boldsymbol{f}(\boldsymbol{b}) - \boldsymbol{f}(\boldsymbol{a}))\ | \le \|\boldsymbol{v}\|_2\ \|\ \boldsymbol{J}\boldsymbol{f}(\boldsymbol{a} + \theta(\boldsymbol{b} - \boldsymbol{a}))\,(\boldsymbol{b} - \boldsymbol{a})\ \|_2 .$$

Speziell für $\boldsymbol{v} := \boldsymbol{f}(\boldsymbol{b}) - \boldsymbol{f}(\boldsymbol{a})$ ergibt sich hieraus die Behauptung. ∎

Bemerkungen (17.45)

a) Die Abschätzung in Satz 17.44 lässt sich abschwächen zu den folgenden Abschätzungen

$$\text{(i)} \quad \|\boldsymbol{f}(\boldsymbol{b}) - \boldsymbol{f}(\boldsymbol{a})\|_2 \leq \|\boldsymbol{J}\boldsymbol{f}(\boldsymbol{a} + \theta(\boldsymbol{b} - \boldsymbol{a}))\|_2 \cdot \|\boldsymbol{b} - \boldsymbol{a}\|_2 \,,$$

$$\text{(ii)} \quad \|\boldsymbol{f}(\boldsymbol{b}) - \boldsymbol{f}(\boldsymbol{a})\|_2 \leq \sup_{\boldsymbol{x} \in [\boldsymbol{a},\boldsymbol{b}]} \|\boldsymbol{J}\boldsymbol{f}(\boldsymbol{x})\|_2 \cdot \|\boldsymbol{b} - \boldsymbol{a}\|_2 \,,$$

wobei die Matrixnormen analog zum quadratischen Fall, vgl. Definition 7.41, definiert werden.

Zumeist wird der Mittelwert-Abschätzungssatz in einer dieser Formen zitiert.

b) Ist $D \subset \mathbb{R}^n$ konvex und offen und ist die Funktion $\boldsymbol{f}\colon D \to \mathbb{R}^m$ differenzierbar mit

$$L := \sup_{\boldsymbol{x} \in [\boldsymbol{a},\boldsymbol{b}]} \|\boldsymbol{J}\boldsymbol{f}(\boldsymbol{x})\|_2 < \infty \,,$$

so ist $\boldsymbol{f}$ Lipschitz-stetig auf D mit der Lipschitz-Konstanten L (bezogen auf $\|\cdot\|_2$), vgl. Definition 10.42.

Bemerkungen (17.46)

Der Mittelwert-Abschätzungssatz gilt in der abgeschwächten Form Bemerkung 17.45 (ii) auch für beliebige Vektornormen und der jeweils zugehörigen Matrixnorm, vgl. Definition 7.41

$$\|\boldsymbol{f}(\boldsymbol{b}) - \boldsymbol{f}(\boldsymbol{a})\| \leq \sup_{\boldsymbol{x} \in [\boldsymbol{a},\boldsymbol{b}]} \|\boldsymbol{J}\boldsymbol{f}(\boldsymbol{x})\| \cdot \|\boldsymbol{b} - \boldsymbol{a}\| \,.$$

Man beweist dies etwa mit Hilfe der Integraldarstellung

$$\boldsymbol{f}(\boldsymbol{b}) - \boldsymbol{f}(\boldsymbol{a}) = \left(\int_0^1 \boldsymbol{J}\boldsymbol{f}(\boldsymbol{a} + t(\boldsymbol{b} - \boldsymbol{a}))\,\mathrm{d}t \right) \cdot (\boldsymbol{b} - \boldsymbol{a})$$

sowie der Abschätzung $\| \int_0^1 \boldsymbol{A}(t)\,\mathrm{d}t \| \leq \int_0^1 \|\boldsymbol{A}(t)\|\,\mathrm{d}t$.

Satz (17.47): Satz von Taylor[5]

Die Funktion $f\colon D \to \mathbb{R}$ sei eine (skalare) C^{m+1}-Funktion auf einer offenen und konvexen Menge $D \subset \mathbb{R}^n$. Ferner sei $\boldsymbol{x}^0 \in D$ (**Entwicklungspunkt**).

Dann gilt für alle $\boldsymbol{x} \in D$ die folgende **Taylor-Entwicklung**

$$f(\boldsymbol{x}) = T_m(\boldsymbol{x}; \boldsymbol{x}^0) + R_m(\boldsymbol{x}; \boldsymbol{x}^0)$$

$$T_m(\boldsymbol{x}; \boldsymbol{x}^0) = \sum_{j=0}^{m} \frac{1}{j\,!} \, [(\boldsymbol{x} - \boldsymbol{x}^0)^{\mathrm{T}} \boldsymbol{\nabla}]^j f(\boldsymbol{x}^0)$$

(Taylor-Polynom m-ten Grades)

$$R_m(\boldsymbol{x}, \boldsymbol{x}^0) = \frac{1}{(m+1)\,!} \, [(\boldsymbol{x} - \boldsymbol{x}^0)^{\mathrm{T}} \boldsymbol{\nabla}]^{m+1} f(\boldsymbol{x}^0 + \theta(\boldsymbol{x} - \boldsymbol{x}^0))$$

(Restgliedformel nach Lagrange[6]**)**

mit einem geeignetem $\theta = \theta(\boldsymbol{x}, \boldsymbol{x}^0) \in \,]0, 1[$.

5) Brook Taylor (1685–1731); Cambridge.

6) Joseph Louis Lagrange (1736–1813); Turin, Paris.

Erläuterung (17.48)
Die Differentialoperatoren $[(\boldsymbol{x}-\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\nabla}]^j$ mit $\boldsymbol{\nabla} = (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n})^{\mathrm{T}}$ sind hierbei folgendermaßen zu bilden:

$$[(\boldsymbol{x}-\boldsymbol{x}^0)^{\mathrm{T}}\boldsymbol{\nabla}]^j = \left(\sum_{i=1}^{n}\left(x_i - x_i^0\right)\frac{\partial}{\partial x_i}\right)^j .$$

Die rechte Seite ist hierbei formal auszumultiplizieren. Dabei sind die entstehenden Ausdrücke nach den partiellen Ableitungen (Differentialoperatoren) zu sortieren. Die resultierende Linearkombination der partiellen Ableitungen ist anschließend auf die Funktion f anzuwenden, und die partiellen Ableitungen sind dann in $\boldsymbol{x}^0$ bzw. in $\boldsymbol{x}^0 + \theta(\boldsymbol{x}-\boldsymbol{x}^0)$ auszuwerten. Man beachte also: Erst wird differenziert, dann werden die Ableitungen an den entsprechenden Argumenten ausgewertet.

Für $n = 2$ ergeben sich beispielsweise so die folgenden Ausdrücke

$$\begin{aligned}
[(\boldsymbol{x}-\boldsymbol{x}^0)^{\mathrm{T}}\boldsymbol{\nabla}]^0 f(\boldsymbol{x}^0) &= f(\boldsymbol{x}^0) \\
[(\boldsymbol{x}-\boldsymbol{x}^0)^{\mathrm{T}}\boldsymbol{\nabla}]^1 f(\boldsymbol{x}^0) &= \left(x_1 - x_1^0\right)\frac{\partial f}{\partial x_1}(\boldsymbol{x}^0) + \left(x_2 - x_2^0\right)\frac{\partial f}{\partial x_2}(\boldsymbol{x}^0) \\
[(\boldsymbol{x}-\boldsymbol{x}^0)^{\mathrm{T}}\boldsymbol{\nabla}]^2 f(\boldsymbol{x}^0) &= \left((x_1 - x_1^0)\frac{\partial}{\partial x_1} + \left(x_2 - x_2^0\right)\frac{\partial}{\partial x_2}\right)^2 f(\boldsymbol{x}^0) \\
&= \left(x_1 - x_1^0\right)^2 \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}^0) \\
&\quad + 2\left(x_1 - x_1^0\right)\left(x_2 - x_2^0\right)\frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}^0) \\
&\quad + \left(x_2 - x_2^0\right)^2 \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}^0) .
\end{aligned}$$

Beweis (zu Satz 17.47).
Mit $\boldsymbol{x} \in D$ gehört aufgrund der vorausgesetzten Konvexität auch die gesamte Verbindungsstrecke $[\boldsymbol{x}^0, \boldsymbol{x}]$ zum Definitionsbereich D. Da D ferner offen ist, gibt es also ein $\varepsilon_1 > 0$ und ein $\varepsilon_2 > 1$ mit

$$\boldsymbol{x}^0 + t(\boldsymbol{x}-\boldsymbol{x}^0) \in D , \quad \forall\, t\colon\ -\varepsilon_1 < t < \varepsilon_2 .$$

Die Taylor-Entwicklung der Funktion $h(t) := f(\boldsymbol{x}^0 + t(\boldsymbol{x}-\boldsymbol{x}^0))$, $-\varepsilon_1 < t < \varepsilon_2$, zum Entwicklungspunkt $t_0 = 0$ liefert somit

$$h(1) = \sum_{j=0}^{m}\frac{1}{j!}h^{(j)}(0) + \frac{1}{(m+1)!}h^{(m+1)}(\theta) , \quad 0 < \theta < 1 .$$

Wir drücken hierin nun die Ableitungen von h durch die partiellen Ableitungen von f aus:

$$\begin{aligned}
h(1) &= f(\boldsymbol{x}) \\
h(t) &= f(\boldsymbol{x}^0 + t(\boldsymbol{x} - \boldsymbol{x}^0)) = [(\boldsymbol{x} - \boldsymbol{x}^0)^{\mathrm{T}}\boldsymbol{\nabla}]^0 f(\boldsymbol{x}^0 + t(\boldsymbol{x} - \boldsymbol{x}^0)) \\
h'(t) &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\boldsymbol{x}^0 + t(\boldsymbol{x} - \boldsymbol{x}^0)) \left(x_i - x_i^0\right) \\
&= \sum_{i_1=1}^{n} \left(x_{i_1} - x_{i_1}^0\right) \frac{\partial f}{\partial x_{i_1}}(\boldsymbol{x}^0 + t(\boldsymbol{x} - \boldsymbol{x}^0)) \\
&= [(\boldsymbol{x} - \boldsymbol{x}^0)^{\mathrm{T}}\boldsymbol{\nabla}]^1 f(\boldsymbol{x}^0 + t(\boldsymbol{x} - \boldsymbol{x}^0)) \\
h''(t) &= \sum_{i_1=1}^{n} \left[(x_{i_1} - x_{i_1}^0) \sum_{i_2=1}^{n} \left(x_{i_2} - x_{i_2}^0\right) \frac{\partial^2 f}{\partial x_{i_1}\partial x_{i_2}}(\boldsymbol{x}^0 + t(\boldsymbol{x} - \boldsymbol{x}^0))\right] \\
&= \left(\sum_{i_1=1}^{n} \left(x_{i_1} - x_{i_1}^0\right) \frac{\partial}{\partial x_{i_1}}\right)\left(\sum_{i_2=1}^{n} \left(x_{i_2} - x_{i_2^0}\right) \frac{\partial}{\partial x_{i_2}}\right) f(\boldsymbol{x}^0 + t(\boldsymbol{x} - \boldsymbol{x}^0)) \\
&= [(\boldsymbol{x} - \boldsymbol{x}^0)^{\mathrm{T}}\boldsymbol{\nabla}]^2 f(\boldsymbol{x}^0 + t(\boldsymbol{x} - \boldsymbol{x}^0))
\end{aligned}$$

Man erkennt hieran, dass auch allgemein gilt:

$$h^{(j)}(t) = [(\boldsymbol{x} - \boldsymbol{x}^0)^{\mathrm{T}}\boldsymbol{\nabla}]^j f(\boldsymbol{x}^0 + t(\boldsymbol{x} - \boldsymbol{x}^0)), \quad j = 1, 2, \ldots, m+1 .$$

Setzt man diese Ausdrücke nun in die obige Taylor-Entwicklung von h ein, so folgt gerade die Behauptung des Taylorschen Satzes. ■

Beispiel (17.49)
Wir bestimmen das Taylor-Polynom $T_2(\boldsymbol{x}, \boldsymbol{x}^0)$ zweiten Grades der Funktion

$$f(x, y, z) = xy^2 \sin z$$

zum Entwicklungspunkt $\boldsymbol{x}^0 = (1, 2, 0)^{\mathrm{T}}$.

Zunächst stellen wir die benötigten partiellen Ableitungen in einer Tabelle zusammen:

Ableitungen	**Auswertung in $\boldsymbol{x}^0$**
$f = xy^2 \sin z$	0
$f_x = y^2 \sin z$	0
$f_y = 2xy \sin z$	0
$f_z = xy^2 \cos z$	4
$f_{xx} = 0$	0
$f_{xy} = 2y \sin z$	0
$f_{xz} = y^2 \cos z$	4
$f_{yy} = 2x \sin z$	0
$f_{yz} = 2xy \cos z$	4
$f_{zz} = -xy^2 \sin z$	0

Weiter gilt mit den Abkürzungen $\partial_x := \frac{\partial}{\partial x}$, $\partial_y := \frac{\partial}{\partial y}$ und analog für die höheren Ableitungen

$$\begin{aligned}
[(\boldsymbol{x} - \boldsymbol{x}^0)^{\mathrm{T}}\boldsymbol{\nabla}] &= (x-1)\partial_x + (y-2)\partial_y + z\,\partial_z \\
[(\boldsymbol{x} - \boldsymbol{x}^0)^{\mathrm{T}}\boldsymbol{\nabla}]^2 &= (x-1)^2\,\partial_{xx} + (y-2)^2\,\partial_{yy} + z^2\,\partial_{zz} \\
&\quad + 2(x-1)(y-2)\partial_{xy} + 2(x-1)z\,\partial_{xz} + 2(y-2)z\,\partial_{yz}\,.
\end{aligned}$$

Damit folgt:

$$\begin{aligned}
T_2(\boldsymbol{x};\boldsymbol{x}^0) &= 0 + \frac{1}{1!}\,[(x-1)\cdot 0 + (y-2)\cdot 0 + z\cdot 4] \\
&\quad + \frac{1}{2!}\,[(x-1)^2\cdot 0 + (y-2)^2\cdot 0 + z^2\cdot 0 \\
&\quad + 2(x-1)(y-2)\cdot 0 + 2(x-1)z\cdot 4 + 2(y-2)z\cdot 4] \\
&= 4z + \frac{1}{2}\,(8(x-1)z + 8(y-2)z) \\
&= 4z(x+y-2)\,.
\end{aligned}$$

Beispiel (17.50)
Gesucht ist das Taylor-Polynom $T_3(\boldsymbol{x};\boldsymbol{x}^0)$ dritten Grades für die Funktion $f(x,y) := \cos(x)\mathrm{e}^y$ zum Entwicklungspunkt $\boldsymbol{x}^0 = (0,0)^{\mathrm{T}}$.

Bei gleichem Vorgehen wie in Beispiel 17.49, also Berechnung der partiellen Ableitung bis zur dritten Ordnung, Auswertung im Entwicklungspunkt und Einsetzen in die allgemeine Formel aus Satz 17.47 erhält man:

$$T_3(\boldsymbol{x};\boldsymbol{x}^0) = 1 + y + \frac{1}{2\,!}\,(-x^2 + y^2) + \frac{1}{3\,!}\,(-3x^2y + y^3)\,.$$

Alternativ hätte man für dieses Beispiel auch die bekannten eindimensionalen Reihenentwicklungen für $\cos x$ und für e^y verwenden können. Beim Ausmultiplizieren braucht man dann nur alle Polynomterme maximal dritten Grades zu berücksichtigen:

$$\begin{aligned}
f(x,y) &= \left(1 - \frac{x^2}{2\,!} + \frac{x^4}{4\,!} - \ldots\right)\left(1 + y + \frac{y^2}{2\,!} + \frac{y^3}{3\,!} + \ldots\right) \\
&= 1 + y + \frac{1}{2}y^2 + \frac{1}{6}y^3 - \frac{1}{2}x^2 - \frac{1}{2}x^2y + O(\|(x,y)^{\mathrm{T}}\|^4)
\end{aligned}$$

Bemerkungen (17.51)

a) Setzt man für den Abstand zum Entwicklungspunkt $\boldsymbol{v} := \boldsymbol{x} - \boldsymbol{x}^0$, so lässt sich der Differentialoperator

$$[(\boldsymbol{x} - \boldsymbol{x}^0)^{\mathrm{T}}\boldsymbol{\nabla}] = \sum_{i=1}^{n} v_i \frac{\partial}{\partial x_i}$$

als Richtungsableitung $D_{\boldsymbol{v}}$ in Richtung $\boldsymbol{v}$ interpretieren, vgl. (17.11). Die in der Taylor-Entwicklung auftretende Potenz $D_{\boldsymbol{v}}^j$ entspricht dann der j-fachen Anwendung dieses Differentialoperators $D_{\boldsymbol{v}}$.

Damit lässt sich die Taylor-Entwicklung auch folgendermaßen schreiben:

$$f(\boldsymbol{x}^0+\boldsymbol{v}) = \sum_{j=0}^{m} \frac{1}{j\,!} D_{\boldsymbol{v}}^{j} f(\boldsymbol{x}^0) + \frac{1}{(m+1)\,!} D_{\boldsymbol{v}}^{m+1} f(\boldsymbol{x}^0+\theta\boldsymbol{v}) \tag{17.21}$$

mit $0 < \theta < 1$. Man erkennt die formale Ähnlichkeit zur eindimensionalen Taylor-Entwicklung Satz 10.13.

b) Eine weitere wichtige Darstellung der Taylor-Entwicklung erhält man mit Hilfe von Multiindices, die später in Definition 25.22 wieder auftauchen. Ein **Multiindex** ist ein n-Tupel

$$\alpha := (\alpha_1, \alpha_2, \dots, \alpha_n) \in \mathbb{N}_0^n \,,$$

dessen **Länge** wir als

$$|\alpha| := \sum_{i=1}^{n} \alpha_1$$

definieren. Aus Gründen der Bezeichnungskonsistenz müssten wir eigentlich $\boldsymbol{\alpha}$ und nicht α schreiben, denn α ist ja ein Vektor, aber das ist bei Multiindices ganz unüblich. Ist $\boldsymbol{x} \in \mathbb{R}^n$, dann sei

$$\boldsymbol{x}^\alpha := x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdots x_n^{\alpha_n} \,.$$

Als Fakultät eines Multiindex setzen wir

$$\alpha! := \alpha_1!\alpha_2! \cdots \alpha_n! \,.$$

Mit Hilfe der Multiindices können wir nun die α-te Ableitung einer Funktion $f: D \to \mathbb{R}$ definieren:

$$\partial^\alpha f(\boldsymbol{x}) := \frac{\partial^{\alpha_1}\partial^{\alpha_2} \cdots \partial^{\alpha_n}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}} f(\boldsymbol{x}) = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}} f(\boldsymbol{x}) \,.$$

Ist $n = 2$ und $\boldsymbol{x} = (x, y)$, dann lauten die möglichen Multiindices bis zur Ableitungsordnung 2

$$\begin{aligned} &\text{Länge } |\alpha| = 0: \quad && (0,0)\,, \\ &\text{Länge } |\alpha| = 1: && (1,0), (0,1)\,, \\ &\text{Länge } |\alpha| = 2: && (2,0), (1,1), (0,2) \end{aligned}$$

und die zugehörigen partiellen Ableitungen sind

$$\begin{aligned} &\partial^{(0,0)} f(x,y) = f(x,y)\,, \\ &\partial^{(1,0)} f(x,y) = \frac{\partial f}{\partial x}(x,y)\,, \qquad \partial^{(0,1)} f(x,y) = \frac{\partial f}{\partial y}(x,y)\,, \\ &\partial^{(2,0)} f(x,y) = \frac{\partial^2 f}{\partial x^2}(x,y)\,, \quad \partial^{(1,1)} f(x,y) = \frac{\partial^2 f}{\partial x\, \partial y}(x,y)\,, \\ &\partial^{(0,2)} f(x,y) = \frac{\partial^2 f}{\partial y^2}(x,y)\,, \end{aligned}$$

aber $\partial^{(1,1)} f(x, y)$ kann auch $\frac{\partial^2 f}{\partial y\, \partial x}(x, y)$ sein, weshalb diese Ableitung doppelt zu zählen ist. Das Taylor-Polynom $T_2(\boldsymbol{x}, \boldsymbol{x}^0)$ lautet dann

$$T_2(\boldsymbol{x}, \boldsymbol{x}^0) = \sum_{|\alpha|=0}^{2} \frac{\partial^\alpha f(\boldsymbol{x}^0)}{\alpha!}(\boldsymbol{x} - \boldsymbol{x}^0)^\alpha ,$$

also

$$\begin{aligned}
&T_2(\boldsymbol{x}, \boldsymbol{x}^0)\\
&\quad = \partial^{(0,0)} f(x^0, y^0) + \partial^{(1,0)} f(x^0, y^0)(x - x^0) + \partial^{(0,1)} f(x^0, y^0)(y - y^0)\\
&\qquad + \frac{1}{2}\partial^{(2,0)} f(x^0, y^0)(x - x^0)^2 + \frac{1}{2} \cdot 2 \cdot \partial^{(1,1)} f(x^0, y^0)(x - x^0) \cdot (y - y^0)\\
&\qquad + \frac{1}{2}\partial^{(0,2)} f(x^0, y^0)(y - y^0)^2\\
&\quad = f(x^0, y^0) + \frac{\partial f}{\partial x}(x^0, y^0)(x - x^0) + \frac{\partial f}{\partial y}(x^0, y^0)(y - y^0)\\
&\qquad + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x - x^0)^2 + \frac{\partial^2 f}{\partial x\, \partial y}(x - x^0)(y - y^0) + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(y - y^0)^2 .
\end{aligned}$$

Die Taylor-Reihe einer Funktion $f : D \to \mathbb{R}$ in Multiindex-Schreibweise lautet dann

$$f(\boldsymbol{x}) = \sum_{|\alpha|\geq 0} \frac{\partial^\alpha f(\boldsymbol{x}^0)}{\alpha!}(\boldsymbol{x} - \boldsymbol{x}^0)^\alpha .$$

c) Das Restglied der Taylor-Formel lautet explizit geschrieben:

$$R_m(\boldsymbol{x}, \boldsymbol{x}^0) = \frac{1}{(m+1)!} \sum_{i_1=1}^{n} \cdots \sum_{i_{m+1}=1}^{n} f_{x_{i_1} \ldots x_{i_{m+1}}}(\boldsymbol{x}^0 + \theta \boldsymbol{v}) v_{i_1} \ldots v_{i_{m+1}} . \tag{17.22}$$

Ist daher C eine gemeinsame Schranke für alle partiellen Ableitungen der Ordnung $(m + 1)$ in einer Umgebung von $\boldsymbol{x}^0$, so gilt dort:

$$|R_m(\boldsymbol{x}; \boldsymbol{x}^0)| \leq \frac{n^{m+1}}{(m+1)!} \cdot C \cdot \|\boldsymbol{x} - \boldsymbol{x}^0\|_\infty^{m+1} . \tag{17.23}$$

Insbesondere folgt für die **Approximationsgüte des Taylor-Polynoms** einer C^{m+1}-Funktion:

$$f(\boldsymbol{x}) = T_m(\boldsymbol{x}; \boldsymbol{x}^0) + O(\|\boldsymbol{x} - \boldsymbol{x}^0\|^{m+1}) . \tag{17.24}$$

Für viele Untersuchungen von Funktionen mehrerer Veränderlicher ist es insbesondere von Bedeutung, das Verhalten der Terme zweiter Ordnung in der Taylor-Entwicklung beurteilen zu können. Hierzu ist es sinnvoll, die zweiten partiellen Ableitungen $f_{x_i x_j}(\boldsymbol{x}^0)$ einer Funktion f zu einer Matrix zusammenzufassen.

Definition (17.52)
Die Matrix der zweiten partiellen Ableitungen

$$\nabla^2 f(\boldsymbol{x}^0) := \begin{pmatrix} f_{x_1 x_1}(\boldsymbol{x}^0) & \dots & f_{x_1 x_n}(\boldsymbol{x}^0) \\ \vdots & & \vdots \\ f_{x_n x_1}(\boldsymbol{x}^0) & \dots & f_{x_n x_n}(\boldsymbol{x}^0) \end{pmatrix} \in \mathbb{R}^{(n,n)}$$

heißt **Hesse-Matrix**[7] der Funktion f im Punkt $\boldsymbol{x}^0$. Sie ist die Jacobi-Matrix des Gradienten $\nabla f(\boldsymbol{x})$ von f. Mitunter wird sie auch mit $\boldsymbol{H} f(\boldsymbol{x}^0)$ oder mit $D^2 f(\boldsymbol{x}^0)$ bezeichnet.

Aufgrund des Schwarzschen Vertauschungssatzes 17.11 ist die Hesse-Matrix einer C^2-Funktion stets **symmetrisch**.

Die ersten drei Summanden der Taylor-Entwicklung einer C^3-Funktion f lauten hiermit:

$$\begin{aligned} f(\boldsymbol{x}) = {} & f(\boldsymbol{x}^0) + \nabla f(\boldsymbol{x}^0)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^0) \\ & + \frac{1}{2\,!}(\boldsymbol{x} - \boldsymbol{x}^0)^{\mathrm{T}} \nabla^2 f(\boldsymbol{x}^0)(\boldsymbol{x} - \boldsymbol{x}^0) + O(\|\boldsymbol{x} - \boldsymbol{x}^0\|^3)\,. \end{aligned} \tag{17.25}$$

7) Ludwig Otto Hesse (1811–1874); Königsberg, Halle, Heidelberg, München.