Kapitel 10

Differenzierbarkeit von Funktionen mehrerer Variablen

10.1 Definition und grundlegende Eigenschaften

Bemerkung 10.1 Motivation. Auch die Differenzierbarkeit von Funktionen mehrerer Veränderlicher lässt sich aus einer Eigenschaft differenzierbarer reellwertiger Funktionen einer reellen Veränderlichen motivieren. Wenn eine reellwertige Funktion einer reellen Veränderlichen im Punkt ξ ihres Definitionsbereiches differenzierbar ist, kann man in diesem Punkt eine Tangente anlegen. Die Tangente ist eine Gerade, also kann man die Funktion in ξ durch ein lineares Polynom (Taylor–Polynom 1. Grades, Satz 7.2) approximieren. Auch bei der Definition der Differentiation von vektorwertigen Funktionen mehrerer Veränderlicher besteht die Grundidee darin, dass man die Funktion in einer Umgebung eines Punktes durch ein lineares Polynom approximieren kann.

Aus der Schule ist bereits das Beispiel der Tangentialebene an einer Kugeloberfläche bekannt. □

Definition 10.2 Differenzierbarkeit vektorwertiger Funktionen. Seien $D \subset \mathbb{R}^n$ offen und $\mathbf{f} : D \to \mathbb{R}^m$. Die Funktion $\mathbf{f}$ heißt im Punkt $\boldsymbol{\xi} \in D$ differenzierbar, wenn es eine lineare Abbildung $M : \mathbb{R}^n \to \mathbb{R}^m$ und eine in $\boldsymbol{\xi}$ stetige Funktion $\mathbf{r} : D \to \mathbb{R}^m$ gibt, so dass

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\boldsymbol{\xi}) + M(\mathbf{x} - \boldsymbol{\xi}) + \mathbf{r}(\mathbf{x}) \left\| \mathbf{x} - \boldsymbol{\xi} \right\|_2, \quad \mathbf{x} \in D, \tag{10.1}$$

und $\mathbf{r}(\boldsymbol{\xi}) = \mathbf{0}$ gelten. Man nennt M die erste Ableitung von $\mathbf{f}(\mathbf{x})$ in $\boldsymbol{\xi}$, im Zeichen $\mathbf{f}'(\boldsymbol{\xi}) := M$. □

Bemerkung 10.3 Komponentenschreibweise. Aus der linearen Algebra ist bekannt, dass sich lineare Abbildungen $M : \mathbb{R}^n \to \mathbb{R}^m$ durch $m \times n$–Matrizen und die Wirkung von M auf den Vektor $\mathbf{x} - \boldsymbol{\xi}$ als Matrix–Vektor–Produkt schreiben lassen. In Komponentenschreibweise kann (10.1) daher mit $M = (m_{ij})_{i=1,\ldots,m;j=1,\ldots,n}$ in der Form

$$f_i(\mathbf{x}) = f_i(\boldsymbol{\xi}) + \sum_{j=1}^{n} m_{ij} (x_j - \xi_j) + r_i(\mathbf{x}) \left\| \mathbf{x} - \boldsymbol{\xi} \right\|_2, \ \mathbf{x} \in D, \ i = 1, \ldots, m, \tag{10.2}$$

geschrieben werden. □

Bemerkung 10.4 Eindeutigkeit von M und $\mathbf{r}(\mathbf{x})$. Es muss geklärt werden, ob M und $\mathbf{r}$ in (10.1) eindeutig bestimmt sind. Seien $\mathbf{e}_j$ der j–te kartesische Einheitsvektor und $t \in \mathbb{R}$ aus einer hinreichend kleinen Umgebung von Null, so dass $\mathbf{x} = \boldsymbol{\xi} + t\mathbf{e}_j$ in D liegt. Dann folgt aus (10.2)

$$m_{ij} = \frac{f_i(\mathbf{x}) - f_i(\boldsymbol{\xi}) - r_i(\mathbf{x}) \|\mathbf{x} - \boldsymbol{\xi}\|_2}{t}.$$

Das gilt für alle hinreichend kleinen Werte t. Betrachte jetzt $t \to 0$. Wegen

$$\|\mathbf{x} - \boldsymbol{\xi}\|_2 = t \|\mathbf{e}_j\|_2 = |t|,$$

der Stetigkeit von $\mathbf{r}(\mathbf{x})$ in $\boldsymbol{\xi}$ und $\mathbf{r}(\boldsymbol{\xi}) = \mathbf{0}$ gilt

$$\begin{aligned} m_{ij} &= \lim_{t\to 0} \frac{f_i(\mathbf{x}) - f_i(\boldsymbol{\xi}) - r_i(\mathbf{x}) \|\mathbf{x} - \boldsymbol{\xi}\|_2}{t} = \lim_{t\to 0} \frac{f_i(\mathbf{x}) - f_i(\boldsymbol{\xi})}{t} - \lim_{t\to 0} \frac{|t|\, r_i(\mathbf{x})}{t} \\ &= \lim_{t\to 0} \frac{f_i(\boldsymbol{\xi} + t\mathbf{e}_j) - f_i(\boldsymbol{\xi})}{t} =: \frac{\partial f_i}{\partial x_j}(\boldsymbol{\xi}). \end{aligned}$$

Die Abbildung M ist also eindeutig bestimmt. Sie wird Jacobi[1]–Matrix genannt, Schreibweise:

$$M = \mathbf{f}'(\boldsymbol{\xi}) = J\mathbf{f}(\boldsymbol{\xi}).$$

Aus (10.1) folgt somit auch, dass $\mathbf{r}(\mathbf{x})$ in eindeutiger Weise gegeben ist

$$\mathbf{r}(\mathbf{x}) = \begin{cases} \dfrac{\mathbf{f}(\mathbf{x}) - \mathbf{f}(\boldsymbol{\xi}) - M(\mathbf{x} - \boldsymbol{\xi})}{\|\mathbf{x} - \boldsymbol{\xi}\|_2} & \mathbf{x} \neq \boldsymbol{\xi}, \\ \mathbf{0} & \mathbf{x} = \boldsymbol{\xi}. \end{cases}$$

Die Differenzierbarkeitsbedingung (10.1) kann nun komponentenweise wie folgt geschrieben werden

$$\begin{aligned} \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix} &= \begin{pmatrix} f_1(\boldsymbol{\xi}) \\ \vdots \\ f_m(\boldsymbol{\xi}) \end{pmatrix} + \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(\boldsymbol{\xi}) & \cdots & \dfrac{\partial f_1}{\partial x_n}(\boldsymbol{\xi}) \\ \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(\boldsymbol{\xi}) & \cdots & \dfrac{\partial f_m}{\partial x_n}(\boldsymbol{\xi}) \end{pmatrix} \begin{pmatrix} x_1 - \xi_1 \\ \vdots \\ x_n - \xi_n \end{pmatrix} \\ &\quad + \begin{pmatrix} r_1(\mathbf{x}) \\ \vdots \\ r_m(\mathbf{x}) \end{pmatrix} \|\mathbf{x} - \boldsymbol{\xi}\|_2 . \end{aligned}$$

□

Satz 10.5 Aus Differenzierbarkeit folgt Stetigkeit. *Ist die Funktion $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ in $\boldsymbol{\xi} \in D$ differenzierbar, so ist $\mathbf{f}(\mathbf{x})$ in $\boldsymbol{\xi}$ stetig.*

Beweis: Aus der Differenzierbarkeit von $\mathbf{f}(\mathbf{x})$ in $\boldsymbol{\xi}$ folgt mit der Dreiecksungleichung

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\boldsymbol{\xi})\|_2 \le \|M(\mathbf{x} - \boldsymbol{\xi})\|_2 + \|\mathbf{r}(\mathbf{x})\|_2 \|\mathbf{x} - \boldsymbol{\xi}\|_2 .$$

Mit der Cauchy–Schwarz[2]schen Ungleichung für Summen erhält man

$$\begin{aligned} \|M(\mathbf{x} - \boldsymbol{\xi})\|_2^2 &= \sum_{i=1}^m \left(\sum_{j=1}^n m_{ij}(x_j - \xi_j) \right)^2 \\ &\le \sum_{i=1}^m \left[\left(\sum_{j=1}^n m_{ij}^2 \right)^{1/2} \left(\sum_{j=1}^n (x_j - \xi_j)^2 \right)^{1/2} \right]^2 \\ &= \left(\sum_{i=1}^m \sum_{j=1}^n m_{ij}^2 \right) \left(\sum_{j=1}^n (x_j - \xi_j)^2 \right) = \|M\|_F^2 \|\mathbf{x} - \boldsymbol{\xi}\|_2^2 , \end{aligned}$$

[1] Carl Gustav Jacob Jacobi (1804 – 1851)
[2] Hermann Amandus Schwarz (1843 – 1921)

wobei $\|M\|_F$ die Frobenius[3]–Norm der Matrix M ist. Somit gibt es eine Konstante $K > 0$, so dass

$$\|M(\mathbf{x}-\boldsymbol{\xi})\|_2 \le K \|\mathbf{x}-\boldsymbol{\xi}\|_2$$

gilt. Nun ist $\mathbf{r}(\mathbf{x})$ stetig in $\boldsymbol{\xi}$ mit $\mathbf{r}(\boldsymbol{\xi}) = \mathbf{0}$. Man findet also zu jedem $\varepsilon > 0$ ein $\delta(\varepsilon) > 0$ derart, dass aus $\|\mathbf{x}-\boldsymbol{\xi}\|_2 < \delta(\varepsilon)$ die Ungleichung $\|\mathbf{r}(\mathbf{x})\|_2 < \varepsilon$ folgt. Setzt man

$$\delta^*(\varepsilon) := \min\left\{\delta(\varepsilon), \frac{\varepsilon}{K+\varepsilon}\right\},$$

so gilt für alle $\mathbf{x}$ mit $\|\mathbf{x}-\boldsymbol{\xi}\|_2 < \delta^*(\varepsilon)$ die Beziehung

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\boldsymbol{\xi})\|_2 < (K+\varepsilon)\|\mathbf{x}-\boldsymbol{\xi}\|_2 < (K+\varepsilon)\,\delta^*(\varepsilon) \le \varepsilon.$$

Das heißt, $\mathbf{f}(\mathbf{x})$ ist im Punkt $\boldsymbol{\xi}$ stetig. ■

Man hat in Satz 9.5 gesehen, dass die Stetigkeit vektorwertiger Funktionen $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ im Punkte $\boldsymbol{\xi} \in D$ äquivalent zur Stetigkeit aller Komponenten $f_i : D \subset \mathbb{R}^n \to \mathbb{R}$ im Punkte $\boldsymbol{\xi} \in D$ ist. Eine ähnliche Aussage gilt auch für die Differenzierbarkeit.

Satz 10.6 Komponentenweise Differenzierbarkeit. *Die Funktion $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ ist im Punkte $\boldsymbol{\xi} \in D$ genau dann differenzierbar, wenn jede ihrer Komponenten $f_i : D \subset \mathbb{R}^n \to \mathbb{R}$ im Punkte $\boldsymbol{\xi} \in D$ differenzierbar ist.*

Beweis: Der Beweis erfolgt direkt mit Hilfe der Definition der Differenzierbarkeit, aus Zeitgründen siehe Literatur. ■

Beispiel 10.7 Die Funktion

$$\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1^2 + x_2^2 \\ x_1 \\ x_2 \end{pmatrix}$$

ist für alle $\mathbf{x} \in \mathbb{R}^2$ differenzierbar. Es gilt

$$J\mathbf{f}(\mathbf{x}) = \begin{pmatrix} 2x_1 & 2x_2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Zum Beweis dieser Behauptung betrachtet man die Definitionsgleichung (10.1) mit beliebigem $\mathbf{x}, \boldsymbol{\xi} \in \mathbb{R}^2$. Ziel ist es, eine Funktion $\mathbf{r}(\mathbf{x})$ zu finden, welche die Bedingungen von Definition 10.2 erfüllt. Es ist

$$\begin{aligned}
\mathbf{r}(\mathbf{x})\,\|\mathbf{x}-\boldsymbol{\xi}\|_2 &= \begin{pmatrix} x_1^2 + x_2^2 \\ x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} \xi_1^2 + \xi_2^2 \\ \xi_1 \\ \xi_2 \end{pmatrix} - \begin{pmatrix} 2\xi_1 & 2\xi_2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 - \xi_1 \\ x_2 - \xi_2 \end{pmatrix} \\
&= \begin{pmatrix} x_1^2 + x_2^2 - \xi_1^2 - \xi_2^2 - 2\xi_1 x_1 + 2\xi_1^2 - 2\xi_2 x_2 + 2\xi_2^2 \\ 0 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} (x_1 - \xi_1)^2 + (x_2 - \xi_2)^2 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \|\mathbf{x}-\boldsymbol{\xi}\|_2^2 \\ 0 \\ 0 \end{pmatrix}.
\end{aligned}$$

Daraus folgt

$$\mathbf{r}(\mathbf{x}) = \begin{pmatrix} \|\mathbf{x}-\boldsymbol{\xi}\|_2 \\ 0 \\ 0 \end{pmatrix}.$$

Diese Funktion ist für alle $\mathbf{x} \in \mathbb{R}^2$ definiert, sie ist im $\mathbb{R}^2$ stetig und für $\mathbf{x} \to \boldsymbol{\xi}$ folgt $\mathbf{r}(\boldsymbol{\xi}) = \mathbf{0}$. Damit sind alle Bedingungen von Definition 10.2 erfüllt. □

[3] Ferdinand Georg Frobenius (1849 – 1917)

10.2 Richtungsableitung und Gradient

Bemerkung 10.8 Motivation. In Analysis I wurde die Ableitung einer skalaren Funktion einer Variablen in einem inneren Punkt ξ ihres Definitionsbereiches eingeführt

$$f'(\xi) := \lim_{h\to 0} \frac{f(\xi+h)-f(\xi)}{h}.$$

Dieser Ableitungsbegriff ist auch anwendbar bei vektorwertigen Funktionen einer Variablen. Für

$$\mathbf{f}(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

definiert man die Ableitung in einem inneren Punkt ξ der Definitionsbereiche aller Komponenten

$$\mathbf{f}'(\xi) := \begin{pmatrix} f_1'(\xi) \\ \vdots \\ f_m'(\xi) \end{pmatrix}.$$

In diesem Abschnitt wird das Konzept der Ableitung in eine Dimension (eine Richtung) auf skalare Funktionen mehrerer Variablen $f\ :\ D \subset \mathbb{R}^n \to \mathbb{R}$ verallgemeinert. $\square$

Definition 10.9 Richtungsableitung, partielle Ableitungen. Seien $D \subset \mathbb{R}^n$ offen, $f\ :\ D \to \mathbb{R}$ und $\boldsymbol{\xi} \in D$. Für einen Vektor $\mathbf{v} \in \mathbb{R}^n$ mit $\|\mathbf{v}\|_2 = 1$ heißt

$$\frac{\partial f}{\partial \mathbf{v}}(\boldsymbol{\xi}) := \lim_{h\to 0} \frac{f(\boldsymbol{\xi}+h\mathbf{v})-f(\boldsymbol{\xi})}{h}$$

die Richtungsableitung (Gateaux[4]–Ableitung) von $f(\mathbf{x})$ in $\boldsymbol{\xi}$ in Richtung $\mathbf{v}$.

Die Ableitungen in Richtung der kartesischen Einheitsvektoren $\mathbf{e}_1, \ldots, \mathbf{e}_n$ nennt man partielle Ableitungen von $f(\mathbf{x})$ nach x_i im Punkt $\boldsymbol{\xi}$. Man schreibt

$$f_{x_i}(\boldsymbol{\xi}) := \frac{\partial f}{\partial x_i}(\boldsymbol{\xi}) := \lim_{h\to 0} \frac{f(\boldsymbol{\xi}+h\mathbf{e}_i)-f(\boldsymbol{\xi})}{h}.$$

$\square$

Beispiel 10.10 Partielle Ableitung. Betrachte

$$f\ :\ \mathbb{R}^2 \to \mathbb{R} \quad \text{mit} \quad f(x_1,x_2) = x_1^3 \cos x_2.$$

Es lässt sich mit Hilfe der Definition der partiellen Ableitungen leicht nachrechnen, dass

$$\frac{\partial f}{\partial x_1}(x_1,x_2) = 3x_1^2 \cos x_2, \quad \frac{\partial f}{\partial x_2}(x_1,x_2) = -x_1^3 \sin x_2$$

gelten. Das bedeutet, man erhält die partielle Ableitung nach x_i indem man wie gewohnt nach x_i differenziert und alle anderen Variablen x_j, $j \neq i$, dabei als Konstante betrachtet. $\square$

Bemerkung 10.11 Zur Jacobi–Matrix. Die i–te Zeile der Jacobi–Matrix $J\mathbf{f}(\boldsymbol{\xi})$ enthält gerade die partiellen Ableitungen der i–ten Komponente $f_i(\mathbf{x})$ im Punkt $\boldsymbol{\xi}$. $\square$

[4] René Eugène Gateaux (1889 – 1914)

Satz 10.12 Existenz und Darstellung der Richtungsableitungen. *Sei $f : D \subset \mathbb{R}^n \to \mathbb{R}$ in $\boldsymbol{\xi} \in D$ differenzierbar. Dann existiert in $\boldsymbol{\xi}$ die Richtungsableitung in jede Richtung $\mathbf{v}$ und es gilt*

$$\frac{\partial f}{\partial \mathbf{v}}(\boldsymbol{\xi}) = Jf(\boldsymbol{\xi}) \cdot \mathbf{v} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\boldsymbol{\xi}) v_i. \tag{10.3}$$

Beweis: Man setzt in die Definition der Differenzierbarkeit $\mathbf{x} = \boldsymbol{\xi} + h\mathbf{v}$. Durch Umstellen erhält man unter Nutzung der Linearität von $Jf(\boldsymbol{\xi})$

$$\frac{f(\boldsymbol{\xi} + h\mathbf{v}) - f(\boldsymbol{\xi})}{h} = Jf(\boldsymbol{\xi}) \cdot \mathbf{v} + r(\boldsymbol{\xi} + h\mathbf{v}) \underbrace{\|\mathbf{v}\|_2}_{=1}.$$

Für $h \to 0$ verschwindet der letzte Term, da $r(\mathbf{x})$ stetig ist und $r(\boldsymbol{\xi}) = 0$ ist. ■

Beispiel 10.13 Richtungsableitung. Gesucht ist die Richtungsableitung von

$$f(x, y) = e^{-x} \sin y$$

in Richtung

$$\mathbf{v} = \begin{pmatrix} 3/5 \\ 4/5 \end{pmatrix}$$

im Punkt $(0, \pi/6)$.

Es gelten $\|\mathbf{v}\|_2 = 1$,

$$\begin{aligned} f_x(x, y) &= -e^{-x} \sin y &\Longrightarrow\quad& f_x\left(0, \frac{\pi}{6}\right) = (-1)\frac{1}{2} = -\frac{1}{2}, \\ f_y(x, y) &= e^{-x} \cos y &\Longrightarrow\quad& f_y\left(0, \frac{\pi}{6}\right) = (1)\frac{1}{2}\sqrt{3} = \frac{1}{2}\sqrt{3}. \end{aligned}$$

Daraus folgt mit (10.3)

$$\frac{\partial f}{\partial \mathbf{v}}\left(0, \frac{\pi}{6}\right) = \mathbf{v}^T \nabla f\left(0, \frac{\pi}{6}\right) = \frac{3}{5}\left(-\frac{1}{2}\right) + \frac{4}{5}\frac{1}{2}\sqrt{3} = -\frac{3}{10} + \frac{2}{5}\sqrt{3}.$$

□

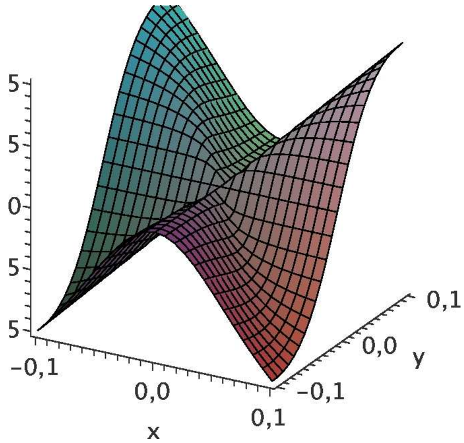
Beispiel 10.14 Existenz aller Richtungsableitungen reicht nicht für Differenzierbarkeit. Die Umkehrung des eben bewiesenen Satzes 10.12 gilt nicht. Betrachte dafür die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

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

Betrachtet man den Koordinatenursprung, dann hat ein beliebiger Einheitsvektor dort die Gestalt $\mathbf{v} = (v_1, v_2)$ mit $v_1^2 + v_2^2 = 1$. Man erhält

$$\frac{\partial f}{\partial \mathbf{v}}(0, 0) = \lim_{h \to 0} \frac{h^3 v_1^2 v_2 - 0}{h(h^2(v_1^2 + v_2^2))} = \lim_{h \to 0} \frac{h^3 v_1^2 v_2}{h^3} = v_1^2 v_2.$$

Diese Richtungsableitung ist keine lineare Funktion von $\mathbf{v}$, wie es in der Definition der Differenzierbarkeit verlangt wird. Daher kann $f(x, y)$ in $(x, y) = (0, 0)$ nicht differenzierbar sein.



Man sieht an diesem Beispiel, dass die Näherung auf jeder beliebigen Gerade einen Wert für die Richtungsableitung in der betreffenden Richtung liefert. Für Differenzierbarkeit benötigt man jedoch die Untersuchung beliebiger Annäherungen, nicht nur die auf Geraden. □

Definition 10.15 Gradient, Nabla–Operator. Seien $D \subset \mathbb{R}^n$ offen und $f : D \to \mathbb{R}$ in $\boldsymbol{\xi} \in D$ nach allen x_i partiell differenzierbar. Dann nennt man den Vektor der ersten partiellen Ableitungen

$$\mathrm{grad} f(\boldsymbol{\xi}) := \nabla f(\boldsymbol{\xi}) := \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(\boldsymbol{\xi}) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(\boldsymbol{\xi}) \end{pmatrix}$$

den Gradienten von $f(\mathbf{x})$ in $\boldsymbol{\xi}$. Manche Bücher definieren $\mathrm{grad} f(\mathbf{x})$ als Zeilenvektor und $\nabla f(\mathbf{x})$ als Spaltenvektor.

Den vektorwertigen Operator

$$\nabla := \begin{pmatrix} \dfrac{\partial}{\partial x_1} \\ \vdots \\ \dfrac{\partial}{\partial x_n} \end{pmatrix}$$

nennt man Nabla–Operator (nach der Form eines ägyptischen Musikinstruments). □

Folgerung 10.16 Existenz des Gradienten und Darstellung der Richtungsableitungen für skalare Funktionen. *Sei $f : D \subset \mathbb{R}^n \to \mathbb{R}$ in $\boldsymbol{\xi} \in D$ differenzierbar. Dann existiert der Gradient von $f(\mathbf{x})$ in $\boldsymbol{\xi}$ und für die Richtungsableitung von $f(\mathbf{x})$ in $\boldsymbol{\xi}$ in Richtung $\mathbf{v}$ gilt*

$$\frac{\partial f}{\partial \mathbf{v}}(\boldsymbol{\xi}) = \nabla f(\boldsymbol{\xi}) \cdot \mathbf{v}.$$

Beispiel 10.17 Aus Existenz des Gradienten folgt nicht Existenz aller Richtungsableitungen. Aus der Existenz des Gradienten folgt nicht notwendig die Existenz der Richtungsableitung in einer von den Koordinatenrichtungen verschiedenen Richtung. Betrachte dazu die Funktion

$$f(x, y) = \begin{cases} 0 & \text{für } xy = 0 \\ 1 & \text{sonst.} \end{cases}$$

Für diese Funktion existiert der Gradient im Punkt $(0,0)$. In allen anderen Richtungen als den Koordinatenrichtungen ist die Funktion im Punkt $(0,0)$ jedoch unstetig, also existiert in diese Richtungen keine Richtungsableitung. $\square$

Bemerkung 10.18 Bedeutung des Gradienten für skalare Funktionen. Sei $f : D \subset \mathbb{R}^n \to \mathbb{R}$ eine skalare Funktion.

1. Für $\nabla f(\boldsymbol{\xi}) \neq \mathbf{0}$ nimmt die Richtungsableitung ihren größten Wert für

$$\mathbf{v} = \frac{\nabla f(\boldsymbol{\xi})}{\|\nabla f(\boldsymbol{\xi})\|_2}$$

an. Das folgt aus der Beziehung

$$\mathbf{v} \cdot \nabla f(\boldsymbol{\xi}) = \cos(\mathbf{v}, \nabla f(\boldsymbol{\xi})) \|\mathbf{v}\|_2 \|\nabla f(\boldsymbol{\xi})\|_2 = \cos(\mathbf{v}, \nabla f(\boldsymbol{\xi})) \|\nabla f(\boldsymbol{\xi})\|_2,$$

da der Kosinus am größten wird, wenn $\mathbf{v}$ in Richtung $\nabla f(\boldsymbol{\xi})$ zeigt. Dann gilt

$$\frac{\partial f}{\partial \mathbf{v}}(\boldsymbol{\xi}) = \mathbf{v} \cdot \nabla f(\boldsymbol{\xi}) = \frac{(\nabla f(\boldsymbol{\xi}))}{\|\nabla f(\boldsymbol{\xi})\|_2} \cdot \nabla f(\boldsymbol{\xi}) = \frac{\|\nabla f(\boldsymbol{\xi})\|_2^2}{\|\nabla f(\boldsymbol{\xi})\|_2} = \|\nabla f(\boldsymbol{\xi})\|_2.$$

Der Gradient zeigt in Richtung des steilsten Anstiegs.

2. In der Gegenrichtung $\mathbf{v} = -\dfrac{\nabla f(\boldsymbol{\xi})}{\|\nabla f(\boldsymbol{\xi})\|_2}$ gilt

$$\frac{\partial f}{\partial \mathbf{v}}(\boldsymbol{\xi}) = -\|\nabla f(\boldsymbol{\xi})\|_2.$$

3. Wählt man eine Richtung $\mathbf{v}$ die orthogonal zu $\nabla f(\boldsymbol{\xi})$ ist, erhält man

$$\frac{\partial f}{\partial \mathbf{v}}(\boldsymbol{\xi}) = \frac{(\nabla f^{\perp}(\boldsymbol{\xi}))}{\|\nabla f^{\perp}(\boldsymbol{\xi})\|_2} \cdot \nabla f(\boldsymbol{\xi}) = 0.$$

Somit ist die Steigung Null, wenn man eine Richtung orthogonal zum Gradienten betrachtet. Damit steht der Gradient $\nabla f(\boldsymbol{\xi})$ in $\boldsymbol{\xi}$ senkrecht zu den sogenannten Höhenlinien $\{\mathbf{x} \in D : f(\mathbf{x}) = f(\boldsymbol{\xi}) = const\}$.

$\square$

Satz 10.19 Hinreichendes Kriterium für die Differenzierbarkeit. *Existieren für $f : D \subset \mathbb{R}^n \to \mathbb{R}$ in einer Umgebung $B(\boldsymbol{\xi}, \rho)$ (Kugel mit Mittelpunkt $\boldsymbol{\xi}$ und Radius ρ) von $\boldsymbol{\xi}$ alle partiellen Ableitungen*

$$\frac{\partial f}{\partial x_i} : B(\boldsymbol{\xi}, \rho) \subset \mathbb{R}^n \to \mathbb{R},$$

und seien diese in $B(\boldsymbol{\xi}, \rho)$ stetig. Dann ist $f(\mathbf{x})$ in $\boldsymbol{\xi}$ differenzierbar.

Beweis: Ohne Beschränkung der Allgemeinheit beschränken wir uns auf den Fall $n = 2$. Das Ziel besteht darin, die Definition der Differenzierbarkeit zu überprüfen. Mit Hilfe des ersten Mittelwertsatzes der Differentialrechnung gilt

$$\begin{aligned} f(x_1, x_2) - f(\xi_1, \xi_2) &= f(x_1, x_2) - f(\xi_1, x_2) + f(\xi_1, x_2) - f(\xi_1, \xi_2) \\ &= \frac{\partial f}{\partial x_1}(\xi_1 + \theta_1(x_1 - \xi_1), x_2)(x_1 - \xi_1) \\ &\quad + \frac{\partial f}{\partial x_2}(\xi_1, \xi_2 + \theta_2(x_2 - \xi_2))(x_2 - \xi_2) \\ &= \sum_{i=1}^{2} \frac{\partial f}{\partial x_i}(\xi_1, \xi_2)(x_i - \xi_i) + R, \end{aligned}$$

mit $\theta_1, \theta_2 \in (0,1)$ und

$$\begin{aligned} R &= \left(\frac{\partial f}{\partial x_1}(\xi_1 + \theta_1(x_1-\xi_1), x_2) - \frac{\partial f}{\partial x_1}(\xi_1,\xi_2)\right)(x_1-\xi_1) \\ &\quad + \left(\frac{\partial f}{\partial x_2}(\xi_1, \xi_2 + \theta_2(x_2-\xi_2)) - \frac{\partial f}{\partial x_2}(\xi_1,\xi_2)\right)(x_2-\xi_2). \end{aligned}$$

Da die partiellen Ableitungen in (ξ_1, ξ_2) stetig sind, gibt es zu jedem $\varepsilon > 0$ ein $\delta(\varepsilon)$, so dass für $\|\mathbf{x} - \boldsymbol{\xi}\|_2 < \delta(\varepsilon)$

$$|R| < \varepsilon \|\mathbf{x} - \boldsymbol{\xi}\|_2$$

gilt. Dies zeigt, dass R in der Form $R = r(\mathbf{x}) \|\mathbf{x} - \boldsymbol{\xi}\|_2$ geschrieben werden kann und $r(\mathbf{x})$ in $B(\boldsymbol{\xi}, \rho)$ stetig ist mit $r(\boldsymbol{\xi}) = 0$. ■

Bemerkung 10.20

- Die Funktion aus Beispiel 10.17 besitzt keine stetigen partiellen Ableitungen auf den Koordinatenachsen, mit Ausnahme des Koordinatenursprungs.
- Das Kriterium von Satz 10.19 ist nicht notwendig. Es gibt Funktionen in mehreren Dimensionen, die an einer Stelle differenzierbar sind, deren partielle Ableitungen aber in der Umgebung dieser Stelle nicht stetig sind, siehe Heuser, Analysis II, S. 266, Aufg. 7.

□

10.3 Eigenschaften differenzierbarer Funktionen

Die vertrauten Differentiationsregeln übertragen sich ohne Änderung, wenn auch äußerlich etwas umständlicher, auf vektorwertige Funktionen von mehreren Veränderlichen.

Satz 10.21 Linearität der Differentiation. *Seien $\mathbf{f}, \mathbf{g} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ zwei vektorwertige Abbildungen, die im Punkt $\boldsymbol{\xi} \in D$ differenzierbar sind. Dann ist mit beliebigen $\alpha, \beta \in \mathbb{R}$ die Linearkombination $(\alpha\mathbf{f} + \beta\mathbf{g})(\mathbf{x})$ in $\boldsymbol{\xi}$ differenzierbar und es gilt*

$$(\alpha\mathbf{f} + \beta\mathbf{g})'(\boldsymbol{\xi}) = \alpha\mathbf{f}'(\boldsymbol{\xi}) + \beta\mathbf{g}'(\boldsymbol{\xi}).$$

Beweis: Der Beweis erfolg direkt mit der Definition der Differenzierbarkeit, Übungsaufgabe. ■

Bemerkung 10.22 Linearität der partiellen Ableitungen. Da die Jacobi–Matrix $\mathbf{f}'(\boldsymbol{\xi})$ die Matrix aller partiellen Ableitungen erster Ordnung ist, sind die partiellen Ableitungen auch linear

$$\frac{\partial}{\partial x_i}(\alpha f_j(\mathbf{x}) + \beta g_j(\mathbf{x})) = \alpha \frac{\partial f_j}{\partial x_i}(\mathbf{x}) + \beta \frac{\partial g_j}{\partial x_i}(\mathbf{x}) \quad \forall\ \alpha, \beta \in \mathbb{R}.$$

□

Satz 10.23 Kettenregel. *Seien $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^p$ und $\boldsymbol{\phi}(\mathbf{x}) = (\mathbf{f} \circ \mathbf{g})(\mathbf{x}) = \mathbf{f}(\mathbf{g}(\mathbf{x}))$ die Verkettung. Sind $\mathbf{g}(\mathbf{x})$ in $\boldsymbol{\xi} \in \mathbb{R}^n$ und $\mathbf{f}(\mathbf{y})$ in $\boldsymbol{\eta} = \mathbf{g}(\boldsymbol{\xi})$ differenzierbar, so ist auch die Verkettung $\boldsymbol{\phi}(\mathbf{x})$ in $\boldsymbol{\xi}$ differenzierbar und es gilt*

$$\boldsymbol{\phi}'(\boldsymbol{\xi}) = \mathbf{f}'(\mathbf{g}(\boldsymbol{\xi})) \circ \mathbf{g}'(\boldsymbol{\xi})$$

beziehungsweise

$$\frac{\partial \phi_i}{\partial x_j}(\boldsymbol{\xi}) = \frac{\partial f_i\left(g_1(\xi_1,\ldots,\xi_n),\ldots,g_m(\xi_1,\ldots,\xi_n)\right)}{\partial x_j}(\boldsymbol{\xi}) = \sum_{k=1}^{m} \frac{\partial f_i}{\partial y_k}(\mathbf{g}(\boldsymbol{\xi}))\frac{\partial g_k}{\partial x_j}(\boldsymbol{\xi}),$$

$i = 1,\ldots,p,\ j = 1,\ldots,n.$

Beweis: Der Beweis erfolgt analog zum Fall $m = n = p = 1$, siehe Analysis I. ■

Bemerkung 10.24 Kettenregel in Matrixschreibweise. Die Verkettung linearer Abbildungen kann als Matrizenprodukt geschrieben werden, siehe Lineare Algebra. Interpretiert man die Ableitungen als entsprechende rechteckige Matrizen, so entspricht der Verkettung von $\mathbf{f}'(\mathbf{g}(\boldsymbol{\xi}))$ mit $\mathbf{g}'(\boldsymbol{\xi})$ also dem Produkt dieser Matrizen. In Komponentenschreibweise gilt daher

$$\begin{aligned}
&\begin{pmatrix} \frac{\partial \phi_1}{\partial x_1}(\boldsymbol{\xi}) & \cdots & \frac{\partial \phi_1}{\partial x_n}(\boldsymbol{\xi}) \\ \vdots & & \vdots \\ \frac{\partial \phi_p}{\partial x_1}(\boldsymbol{\xi}) & \cdots & \frac{\partial \phi_p}{\partial x_n}(\boldsymbol{\xi}) \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial f_1}{\partial y_1}(\mathbf{g}(\boldsymbol{\xi})) & \cdots & \frac{\partial f_1}{\partial y_m}(\mathbf{g}(\boldsymbol{\xi})) \\ \vdots & & \vdots \\ \frac{\partial f_p}{\partial y_1}(\mathbf{g}(\boldsymbol{\xi})) & \cdots & \frac{\partial f_p}{\partial y_m}(\mathbf{g}(\boldsymbol{\xi})) \end{pmatrix} \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\boldsymbol{\xi}) & \cdots & \frac{\partial g_1}{\partial x_n}(\boldsymbol{\xi}) \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_1}(\boldsymbol{\xi}) & \cdots & \frac{\partial g_m}{\partial x_n}(\boldsymbol{\xi}) \end{pmatrix}.
\end{aligned}$$

Hat man insbesondere eine Funktion $\phi(s,t) = f(x(s,t), y(s,t)) \; : \; D \subset \mathbb{R}^2 \to \mathbb{R}$ gegeben, so erhält man

$$\begin{aligned}
&\begin{pmatrix} \frac{\partial \phi}{\partial s}(s_0,t_0) & \frac{\partial \phi}{\partial t}(s_0,t_0) \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial f}{\partial x}(x_0,y_0) & \frac{\partial f}{\partial y}(x_0,y_0) \end{pmatrix} \begin{pmatrix} \frac{\partial x}{\partial s}(s_0,t_0) & \frac{\partial x}{\partial t}(s_0,t_0) \\ \frac{\partial y}{\partial s}(s_0,t_0) & \frac{\partial y}{\partial t}(s_0,t_0) \end{pmatrix}
\end{aligned}$$

oder ausgeschrieben

$$\begin{aligned}
\frac{\partial \phi}{\partial s}(s_0,t_0) &= \frac{\partial f}{\partial x}(x_0,y_0)\frac{\partial x}{\partial s}(s_0,t_0) + \frac{\partial f}{\partial y}(x_0,y_0)\frac{\partial y}{\partial s}(s_0,t_0), \\
\frac{\partial \phi}{\partial t}(s_0,t_0) &= \frac{\partial f}{\partial x}(x_0,y_0)\frac{\partial x}{\partial t}(s_0,t_0) + \frac{\partial f}{\partial y}(x_0,y_0)\frac{\partial y}{\partial t}(s_0,t_0) \qquad (10.4)
\end{aligned}$$

mit $(x_0, y_0) = (x(s_0,t_0), y(s_0,t_0))$. □

Beispiel 10.25 Kettenregel. Betrachte die Abbildungen

$$\begin{aligned}
\mathbf{g} &: \mathbb{R}^2 \to \mathbb{R}^2 \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1^5 - 3x_1^2x_2^5 + x_2^2 - 7 \\ x_1 + x_2^2 \end{pmatrix}, \\
f &: \mathbb{R}^2 \to \mathbb{R} \quad \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \mapsto g_1^{1700} + g_2^{28},
\end{aligned}$$

sowie ihre Komposition

$$\phi(x_1,x_2) = f(\mathbf{g}(x_1,x_2)) = f\left(g_1(x_1,x_2), g_2(x_1,x_2)\right).$$

Für die partielle Ableitung nach x_1 erhält man nach Kettenregel

$$\begin{aligned}
\frac{\partial \phi}{\partial x_1}(x_1,x_2) &= \left(\frac{\partial f}{\partial g_1}\frac{\partial g_1}{\partial x_1} + \frac{\partial f}{\partial g_2}\frac{\partial g_2}{\partial x_1}\right)(x_1,x_2) \\
&= \left(1700 g_1^{1699}\frac{\partial g_1}{\partial x_1} + 28 g_2^{27}\frac{\partial g_2}{\partial x_1}\right)(x_1,x_2) \\
&= 1700 g_1^{1699}\left(5x_1^4 - 6x_1x_2^5\right) + 28 g_2^{27} \cdot 1 \\
&= 1700\left(x_1^5 - 3x_1^2x_2^5 + x_2^2 - 7\right)^{1699}\left(5x_1^4 - 6x_1x_2^5\right) + 28\left(x_1 + x_2^2\right)^{27}.
\end{aligned}$$

□

10.4 Der Mittelwertsatz

Beispiel 10.26 Einfache Übertragung des Mittelwertsatzes aus einer Dimension gilt nicht. Der Mittelwertsatz in einer Dimension besagt, dass man für eine in einem Intervall $[a,b]$ differenzierbare Funktion $f \; : \; [a,b] \to \mathbb{R}$ einen Wert $\xi \in (a,b)$ findet, so dass

$$f(b) - f(a) = f'(\xi)(b-a)$$

gilt. Solche eine Aussage gilt für vektorwertige Funktionen nicht. Wir betrachten dazu die Funktion

$$\mathbf{f} \; : \; [0,2\pi] \to \mathbb{R}^2 \quad x \mapsto \begin{pmatrix} \cos x \\ \sin x \end{pmatrix}.$$

Dann gilt für alle $\xi \in (0,2\pi)$

$$\mathbf{f}(b) - \mathbf{f}(a) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \neq 2\pi \begin{pmatrix} -\sin\xi \\ \cos\xi \end{pmatrix} = \mathbf{f}'(\xi)(b-a),$$

da Sinus und Cosinus nicht gleichzeitig Null sind, Satz 2.18. □

Man kann aber für skalare Funktionen mehrerer Veränderlicher einen Mittelwertsatz formulieren.

Definition 10.27 Konvexe Menge. Eine Menge $D \subset \mathbb{R}^n$ heißt konvex, wenn mit zwei beliebigen Punkten $\mathbf{a}, \mathbf{b} \in D$ immer die Verbindungsstrecke zwischen diesen Punkten, das heißt die Menge

$$\{\boldsymbol{\xi} \; : \; \boldsymbol{\xi} = \mathbf{a} + \theta(\mathbf{b} - \mathbf{a}), \theta \in (0,1)\}$$

in D liegt. □

Satz 10.28 Mittelwertsatz für skalare Funktionen mehrerer Veränderlicher. *Seien $\mathbf{a}, \mathbf{b} \in D$ zwei Punkte einer konvexen, offenen Menge D und $f \; : \; D \subset \mathbb{R}^n \to \mathbb{R}$ auf D differenzierbar. Dann gibt es auf dem $\mathbf{a}$ und $\mathbf{b}$ verbindenden Segment einen Punkt $\boldsymbol{\xi} = \mathbf{a} + \theta(\mathbf{b} - \mathbf{a})$, $\theta \in (0,1)$, so dass*

$$f(\mathbf{b}) - f(\mathbf{a}) = f'(\boldsymbol{\xi}) \cdot (\mathbf{b} - \mathbf{a})$$

gilt.

Beweis: Der Beweis erfolgt durch Anwendung der Kettenregel. Man betrachtet $f(\mathbf{x})$ auf dem Segment von $\mathbf{a}$ nach $\mathbf{b}$. Die Kurve über diesem Segment kann mit einer skalarwertigen Funktion einer reellen Veränderlichen identifiziert werden

$$\phi \; : \; [0,1] \to \mathbb{R} \quad t \mapsto f(\mathbf{a} + t(\mathbf{b} - \mathbf{a})).$$

Da $f(\mathbf{x})$ differenzierbar ist, existieren nach Satz 10.12 alle Richtungsableitungen in allen Punkten von D. Insbesonder existieren die Richtungsableitungen in allen Punkten des Segments zwischen $\mathbf{a}$ und $\mathbf{b}$ in Richtung $\mathbf{b} - \mathbf{a}$. Damit ist die Funktion $\phi(t)$ auf diesem Segment differenzierbar und sie genügt auf $[0,1]$ den Voraussetzungen des Mittelwertsatzes in einer Dimension. Damit gibt es ein $\theta \in (0,1)$ mit

$$f(\mathbf{b}) - f(\mathbf{a}) = \phi(1) - \phi(0) = \phi'(\theta) \cdot 1. \tag{10.5}$$

Andererseits kann man auf $\phi(t)$ die Kettenregel (10.4) anwenden, womit man

$$\phi'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))\frac{\partial x_i}{\partial t} = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))(b_i - a_i) = f'(\mathbf{a} + t(\mathbf{b} - \mathbf{a})) \cdot (\mathbf{b} - \mathbf{a}).$$

erhält. Für $t = \theta$ erhält man in (10.5) gerade die Behauptung des Satzes. ■

Beispiel 10.29 Mittelwertsatz. Sei

$$f(x,y) = \cos x + \sin y, \quad \mathbf{a} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \mathbf{b} = \frac{1}{2}\begin{pmatrix} \pi \\ \pi \end{pmatrix}.$$

Es gilt

$$f(0,0) = 1 = f\left(\frac{\pi}{2}, \frac{\pi}{2}\right).$$

Nach dem Mittelwertsatz existiert also ein $\theta \in (0,1)$ mit

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

In der Tat ist diese Gleichung für $\theta = 1/2$ erfüllt. □

10.5 Höhere partielle Ableitungen

Nun wird der Begriff der partiellen Ableitung auf partielle Ableitungen höherer Ordnung verallgemeinert.

Definition 10.30 Partielle Ableitungen höherer Ordnung. Seien $D \subset \mathbb{R}^n$ offen und $f : D \to \mathbb{R}$. Ist $f(\mathbf{x})$ partiell differenzierbar auf D und sind die partiellen Ableitungen selbst wieder differenzierbar, erhält man als partielle Ableitungen zweiter Ordnung

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

Induktiv definiert man die partiellen Ableitungen k–ter Ordnung durch

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

$i_1, \ldots, i_k \in \{1, \ldots n\}$. □

Man kann zeigen, dass unter gewissen Voraussetzungen die Reihenfolge, in welcher man partielle Ableitungen höherer Ordnung berechnet, keine Rolle spielt.

Satz 10.31 Vertauschbarkeitssatz von Schwarz (für partielle Ableitungen zweiter Ordnung). *Seien $D \subset \mathbb{R}^n$ offen, $\boldsymbol{\xi} \in D$ und $f : D \to \mathbb{R}$. Falls die partiellen Ableitungen von $f(\mathbf{x})$ bis zur zweiten Ordnung in einer Umgebung $B(\boldsymbol{\xi}, \rho) \subset D$ stetig sind, so sind sie vertauschbar*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{\xi}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\boldsymbol{\xi}) \quad \forall i,j \in \{1, \ldots, n\}.$$

Beweis: Der Beweis erfolgt durch Anwendung des Mittelwertsatzes, für Details siehe Literatur. ■

Folgerung 10.32 Vertauschbarkeitssatz von Schwarz für partielle Ableitungen k–ter Ordnung. *Sind die partiellen Ableitungen von $f(\mathbf{x})$ bis zur Ordnung k stetig, so kann man die Reihenfolge der partiellen Ableitungen von $f(\mathbf{x})$ bis zur Ordnung k beliebig vertauschen.*

Bemerkung 10.33 Multiindexschreibweise. Hat man eine Funktion, bei welcher die Reihenfolge der partiellen Ableitungen beliebig gewählt werden kann, verwendet man häufig eine abkürzende Schreibweise unter Verwendung von Multiindizes. Ein Multiindex ist ein n–Tupel $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ nichtnegativer ganzer Zahlen $\alpha_i \geq 0$ mit $|\boldsymbol{\alpha}| = \sum_{i=1}^{n} \alpha_i$. Man schreibt dann

$$D^{\boldsymbol{\alpha}} f(\mathbf{x}) := \frac{\partial^{|\boldsymbol{\alpha}|} f}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}(\mathbf{x}) = f_{\alpha_1 \alpha_2, \ldots, \alpha_n}(\mathbf{x}).$$

□

Beispiel 10.34 Höhere partielle Ableitungen. Für die Funktion

$$f(x, y, z) = z^2 \sin(x^3) + (\cos y \sin x - e^{-x}) z^2$$

soll $f_{111}(x, y, z) =: f_{xyz}(x, y, z)$ berechnet werden. Es ist zunächst klar, dass von dieser Funktion die partiellen Ableitungen beliebiger Ordnung existieren und stetig sind. Hier bietet es sich an, die Funktion zuerst nach y partiell zu integrieren, weil dann einige Terme sofort verschwinden

$$\begin{aligned} f_{xyz}(x, y, z) &= \frac{\partial^2}{\partial x \partial z}(f_y) = \frac{\partial^2}{\partial x \partial z}\left(-(\sin y \sin x) z^2\right) = \frac{\partial}{\partial x}(-2z \sin y \sin x) \\ &= -2z \sin y \cos x. \end{aligned}$$

□

Definition 10.35 Hesse[5]–Matrix. Seien $D \subset \mathbb{R}^n$ offen und $f : D \to \mathbb{R}$ sei im Punkt $\boldsymbol{\xi} \in D$ zweimal partiell differenzierbar. Dann nennt man die Matrix der partiellen zweiten Ableitungen

$$Hf(\boldsymbol{\xi}) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{\xi}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{\xi}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\boldsymbol{\xi}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{\xi}) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{\xi}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\boldsymbol{\xi}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\boldsymbol{\xi}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\boldsymbol{\xi}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\boldsymbol{\xi}) \end{pmatrix}$$

die Hesse–Matrix von $f(\mathbf{x})$ in $\boldsymbol{\xi}$. □

Bemerkung 10.36 Zur Hesse–Matrix.

- Für eine Funktion mit stetigen partiellen Ableitungen zweiter Ordnung ist die Hesse–Matrix nach dem Vertauschbarkeitssatz von Schwarz symmetrisch.
- Die Hesse–Matrix spielt eine große Rolle bei der Klassifikation von Extrema einer Funktion mehrerer Variablen, siehe Kapitel 11.
- Die Euklidische Norm eines Vektors ist invariant unter Rotationen. Ähnlich wie bei $\nabla f(\mathbf{x})$ die wichtige rotationsinvariante Größe $\|\nabla f(\mathbf{x})\|_2$ abgeleitet werden kann, kann man auch aus $Hf(\mathbf{x})$ eine rotationsinvariante skalare Größe gewinnen. Hierzu betrachtet man die Summe der Diagonalelemente, die sogenannte Spur, von $Hf(\mathbf{x})$.

□

[5] Ludwig Otto Hesse (1811 – 1874)

Definition 10.37 Laplace[6]–Operator. Seien $D \subset \mathbb{R}^n$ offen und $f : D \to \mathbb{R}$ in $\boldsymbol{\xi} \in D$ zweimal partiell differenzierbar. Dann nennt man

$$\Delta f(\boldsymbol{\xi}) := \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}(\boldsymbol{\xi}) = (f_{x_1x_1} + \ldots + f_{x_nx_n})(\boldsymbol{\xi})$$

den Laplace–Operator von $f(\mathbf{x})$ in $\boldsymbol{\xi}$. □

Bemerkung 10.38 Anwendungen des Laplace–Operators, partielle Differentialgleichungen. Viele Prozesse in der Physik und den Ingenieurswissenschaften lassen sich durch Gleichungen beschreiben, welche die Beziehungen zwischen einer gesuchten Funktion, zum Beispiel Druck, Temperatur, Konzentration, ..., und ihren partiellen Ableitungen der Ordnung kleiner oder gleich Zwei beinhalten. Dabei spielt der Laplace–Operator oft eine große Rolle. Man nennt diese Gleichungen partielle Differentialgleichungen 2. Ordnung, wegen der Ableitungen zweiter Ordnung im Laplace–Operator. Im folgenden werden einige Beispiele angegeben.

1. Die Laplace–Gleichung oder Potentialgleichung hat die Gestalt
$$\Delta u = 0 \quad \text{in } D.$$
Sie spielt zum Beispiel bei statischen Problemen eine Rolle, beispielsweise in der Elastizitätstheorie. Funktionen, die die Laplace–Gleichung erfüllen, nennt man harmonische Funktionen.
2. Die Wellengleichung besitzt die Form
$$u_{tt} = \Delta u \quad \text{in } (0,T) \times D.$$
Diese Gleichung beschreibt Schwingungsprobleme, zum Beispiel die Ausbreitung von Schallwellen. Hierbei ist t die Zeit und $\Delta u = u_{xx} + u_{yy} + u_{zz}$ misst die örtliche Veränderung.
3. Die Diffusionsgleichung oder Wärmeleitungsgleichung hat die Form
$$u_t = \Delta u \quad \text{in } (0,T) \times D.$$
Sie beschreibt die Dynamik von Ausgleichsprozessen wie Diffusion und Wärmeleitung. Hierbei ist $u(t,\mathbf{x})$ die Konzentration oder die Temperatur, und t ist die Zeit.
In der Bildverarbeitung verwendet man Diffusionsprozesse zum Entrauschen (Diffusionsfilter). Dabei beschreibt $u(t,\mathbf{x})$ den Grauwert, und die Diffusionszeit t ist ein Maß für die Glättung.

□

10.6 Die Taylorsche Formel

Bemerkung 10.39 Motivation. Ziel ist die Verallgemeinerung der Taylorschen Formel aus Satz 7.2 auf den Fall skalarer Funktionen $f : D \subset \mathbb{R}^n \to \mathbb{R}$. □

Satz 10.40 Taylorsche Formel mit dem Restglied von Lagrange. *Seien $D \subset \mathbb{R}^n$ offen und konvex und sei $f(\mathbf{x})$ $(m+1)$–mal stetig differenzierbar in D. Dann gilt für $\boldsymbol{\xi}, \boldsymbol{\xi} + \mathbf{h} \in D$*

$$\begin{aligned} f(\boldsymbol{\xi} + \mathbf{h}) &= \sum_{i=0}^{m} \frac{1}{i!} \left((\mathbf{h} \cdot \nabla)^i f \right)(\boldsymbol{\xi}) + \frac{1}{(m+1)!} \left((\mathbf{h} \cdot \nabla)^{m+1} f \right)(\boldsymbol{\xi} + \theta \mathbf{h}) \\ &=: T_m(\mathbf{h}, \boldsymbol{\xi}) + R_m(\mathbf{h}, \boldsymbol{\xi}), \end{aligned}$$

[6]Pierre-Simon (Marquis de) Laplace (1749 – 1827)

mit $\theta \in (0,1)$. Dabei werden $T_m(\mathbf{h},\boldsymbol{\xi})$ Taylor–Polynom vom Grad m und $R_m(\mathbf{h},\boldsymbol{\xi})$ Restglied nach Lagrange genannt.

Beweis: Die formale Schreibweise der Taylorschen Formel wird innerhalb des Beweises erklärt.

Wie im Beweis des Mittelwertsatzes wird eine Funktion einer reellen Veränderlichen betrachtet, nämlich

$$\phi(t) = f(\boldsymbol{\xi} + t\mathbf{h}), \quad t \in \mathbb{R},\ \boldsymbol{\xi} + t\mathbf{h} \in D.$$

Da $f(\mathbf{x})$ $(m+1)$–mal stetig differenzierbar ist, existieren alle Ableitungen in Richtung $\boldsymbol{\xi}+t\mathbf{h}$ bis zur Ordnung $m+1$, Satz 10.12. Damit erfüllt die Funkion $\phi(t)$ die Voraussetzungen des Satzes von Taylor, Satz 7.2. Um diesen Satz anwenden zu können, muss man $\phi(t)$ differenzieren. Mit Hilfe der Kettenregel, Bemerkung 10.24, folgt

$$\begin{aligned}
\phi'(t) &= \left(\frac{\partial f}{\partial x_1}(\boldsymbol{\xi} + t\mathbf{h}), \ldots, \frac{\partial f}{\partial x_n}(\boldsymbol{\xi} + t\mathbf{h})\right) \cdot (h_1, \ldots, h_n) \\
&= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\boldsymbol{\xi} + t\mathbf{h}) h_i =: (\mathbf{h} \cdot \nabla f)\,(\boldsymbol{\xi} + t\mathbf{h})
\end{aligned}$$

mit der formalen Schreibweise

$$(\mathbf{h} \cdot \nabla) := h_1 \frac{\partial}{\partial x_i} + \ldots + h_n \frac{\partial}{\partial x_n}.$$

Für die zweite Ableitung erhält man analog (betrachte am einfachsten die Summendarstellung der ersten Ableitung)

$$\phi''(t) = \sum_{i_1,i_2=1}^{n} \frac{\partial^2 f}{\partial x_{i_1} \partial x_{i_2}}(\boldsymbol{\xi} + t\mathbf{h}) h_{i_1} h_{i_2} =: \left((\mathbf{h} \cdot \nabla)^2 f\right)(\boldsymbol{\xi} + t\mathbf{h})$$

mit

$$\begin{aligned}
(\mathbf{h} \cdot \nabla)^2 &:= \left(h_1 \frac{\partial}{\partial x_1} + \ldots + h_n \frac{\partial}{\partial x_n}\right) \cdot \left(h_1 \frac{\partial}{\partial x_1} + \ldots + h_n \frac{\partial}{\partial x_n}\right) \\
&= h_1^2 \frac{\partial^2}{\partial x_1^2} + h_1 h_2 \frac{\partial^2}{\partial x_1 \partial x_2} + \ldots + h_2 h_1 \frac{\partial^2}{\partial x_2 \partial x_1} + \ldots + h_n^2 \frac{\partial^2}{\partial x_n^2} \\
&= \sum_{i_1,i_2=1}^{n} \frac{\partial^2}{\partial x_{i_1} \partial x_{i_2}} h_{i_1} h_{i_2}.
\end{aligned}$$

Für die k–te Ableitung erhält man schließlich

$$\phi^{(k)}(t) = \sum_{i_1,\ldots,i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_1} \ldots \partial x_{i_k}}(\boldsymbol{\xi} + t\mathbf{h}) h_{i_1} \ldots h_{i_k} =: \left((\mathbf{h} \cdot \nabla)^k f\right)(\boldsymbol{\xi} + t\mathbf{h}).$$

Der Satz von Taylor für die Funktion $\phi(t)$ ergibt die Darstellung

$$\phi(1) = \sum_{i=0}^{m} \frac{\phi^{(i)}(0)}{i!} + \frac{\phi^{(m+1)}(\theta)}{(m+1)!}$$

mit $\theta \in (0,1)$. Setzt man die berechneten Ableitungen von $\phi(t)$ ein, ergibt sich gerade die Aussage des Satzes. ■

Bemerkung 10.41 Zur Schreibweise der Taylor–Formel. In der Literatur findet man die äquivalente Darstellung

$$(\mathbf{h} \cdot \nabla)^i = \left(\mathbf{h}^T \nabla\right)^i.$$

Wie bereits im Beweis angedeutet, wird dieser Ausdruck formal ausmultipliziert und dann nach Ableitungen sortiert. Die ersten drei Summanden lassen sich übersichtlich wie folgt schreiben:

$$f(\boldsymbol{\xi} + \mathbf{h}) = f(\boldsymbol{\xi}) + \mathbf{h}^T \nabla f(\boldsymbol{\xi}) + \frac{1}{2} \mathbf{h}^T H f(\boldsymbol{\xi}) \mathbf{h} + \ldots$$

mit der Hesse–Matrix $Hf(\boldsymbol{\xi})$.

Oft wird auch $\mathbf{x} = \boldsymbol{\xi} + \mathbf{h}$ gesetzt. Dann hat die Taylor–Formel die Gestalt

$$f(\mathbf{x}) = \sum_{i=0}^{m} \frac{1}{i!} \left(((\mathbf{x}-\boldsymbol{\xi}) \cdot \nabla)^i f \right)(\boldsymbol{\xi}) + \frac{1}{(m+1)!} \left(((\mathbf{x}-\boldsymbol{\xi}) \cdot \nabla)^{m+1} f \right)(\boldsymbol{\xi} + \theta(\mathbf{x}-\boldsymbol{\xi}))$$

mit $\theta \in (0,1)$. Dann wird das Taylor–Polynom mit $T_m(\mathbf{x}, \boldsymbol{\xi})$ bezeichnet und das Restglied mit $R_m(\mathbf{x}, \boldsymbol{\xi})$. □

Beispiel 10.42 Taylor–Polynom. Gesucht ist das Taylor–Polynom zweiten Grades von

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

im Entwicklungspunkt $\boldsymbol{\xi} = (1,2,0)^T$. Zur Lösung dieser Aufgabe benötigt man die Funktionswerte aller partiellen Ableitungen bis zur zweiten Ordnung im Entwicklungspunkt:

Ableitungen	Auswertung in $(1,2,0)^T$
$f = xy^2 \sin z$	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
$f_{yy} = 2x \sin z$	0
$f_{zz} = -xy^2 \sin z$	0
$f_{xy} = 2y \sin z$	0
$f_{xz} = y^2 \cos z$	4
$f_{yz} = 2xy \cos z$	4

Damit erhält man

$$f(\boldsymbol{\xi}) = 0, \quad \nabla f(\boldsymbol{\xi}) = \begin{pmatrix} 0 \\ 0 \\ 4 \end{pmatrix}, \quad Hf(\boldsymbol{\xi}) = \begin{pmatrix} 0 & 0 & 4 \\ 0 & 0 & 4 \\ 4 & 4 & 0 \end{pmatrix}$$

und es folgt für das Taylor–Polynom 2. Grades mit $\mathbf{h} = \mathbf{x} - \boldsymbol{\xi}$

$$\begin{aligned} & T_2(\mathbf{x}, \boldsymbol{\xi}) \\ = \; & 0 + \begin{pmatrix} x_1 - 1 \\ x_2 - 2 \\ x_3 \end{pmatrix}^T \begin{pmatrix} 0 \\ 0 \\ 4 \end{pmatrix} + \frac{1}{2} \left(\begin{pmatrix} x_1 - 1 \\ x_2 - 2 \\ x_3 \end{pmatrix}^T \begin{pmatrix} 0 & 0 & 4 \\ 0 & 0 & 4 \\ 4 & 4 & 0 \end{pmatrix} \begin{pmatrix} x_1 - 1 \\ x_2 - 2 \\ x_3 \end{pmatrix} \right) \\ = \; & 4x_3 + \frac{1}{2} \left(8(x_1 - 1)x_3 + 8(x_2 - 2)x_3 \right) = -8x_3 + 4x_1x_3 + 4x_2x_3. \end{aligned}$$

□

Bemerkung 10.43 Zum Satz von Taylor.

- Für $m = 0$ liefert der Satz von Taylor

$$f(\mathbf{x}) \quad = \quad f(\boldsymbol{\xi}) + (\mathbf{x}-\boldsymbol{\xi})^T \nabla f(\boldsymbol{\xi} + \theta(\mathbf{x}-\boldsymbol{\xi})), \quad \theta \in (0,1).$$

  Dies ist äquivalent zum Mittelwertsatz 10.28 angewendet auf $\mathbf{a} = \boldsymbol{\xi}$, $\mathbf{b} = \mathbf{x}$. Der Mittelwertsatz ist also ein Spezialfall des Satzes von Taylor.
- Für beschränkte partielle Ableitungen der Ordnung $m+1$ hat das Restglied die Fehlerordnung $\mathcal{O}(\|\mathbf{x}-\boldsymbol{\xi}\|_2^{m+1})$. Somit folgt für die Approximationsgüte des Taylorpolynoms

$$f(\mathbf{x}) = T_m(\mathbf{x}, \boldsymbol{\xi}) + \mathcal{O}\left(\|\mathbf{x}-\boldsymbol{\xi}\|_2^{m+1} \right).$$

□