13. Differentiation

13.1 Das Differential einer Abbildung

Gegeben: $\mathbf{f} : \mathbb{R}^n \supset D \to \mathbb{R}^m$, also eine vektorwertige Funktion von n Variablen $\mathbf{x} = (x_1, \ldots, x_n)^\mathsf{T}$, wobei D wiederum offen sei.

Wir nennen die Funktion $\mathbf{f}$ im Punkt $\mathbf{x}^0$ **differenzierbar**, auch **vollständig** oder **total differenzierbar**, falls es eine lineare Abbildung $\boldsymbol{\ell} : \mathbb{R}^n \to \mathbb{R}^m$ gibt, für die die folgende Approximationseigenschaft gilt

$$\mathbf{f}(\mathbf{x}) \;=\; \mathbf{f}(\mathbf{x}^0) + \boldsymbol{\ell}(\mathbf{x} - \mathbf{x}^0) + \mathbf{o}(\|\mathbf{x} - \mathbf{x}^0\|), \quad \text{bzw.}$$

$$\lim_{\mathbf{x} \to \mathbf{x}^0} \frac{\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}^0) - \boldsymbol{\ell}(\mathbf{x} - \mathbf{x}^0)}{\|\mathbf{x} - \mathbf{x}^0\|} \;=\; \mathbf{0}\,. \tag{13.1.1}$$

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

Bemerkungen (13.1.2)

a) Eine differenzierbare Funktion lässt sich also bei $\mathbf{x}^0$ mit hinreichender Genauigkeit durch eine affin-lineare Funktion $\tilde{\mathbf{f}}(x) = \mathbf{f}(\mathbf{x}^0) + \mathbf{df}(\mathbf{x}^0)(\mathbf{x} - \mathbf{x}^0)$ approximieren.

b) Ist $\mathbf{f}$ in $\mathbf{x}^0$ differenzierbar, so ist das Differential $\boldsymbol{\ell} = \mathbf{df}(\mathbf{x}^0)$ eindeutig bestimmt. Ferner besitzt $\mathbf{df}(\mathbf{x}^0)$ die (kanonische) Matrixdarstellung:

$$\boldsymbol{\ell}(\mathbf{z}) = \mathbf{df}(\mathbf{x}^0)(\mathbf{z}) = \mathbf{Jf}(\mathbf{x}^0) \cdot \mathbf{z}, \quad \mathbf{z} \in \mathbb{R}^n,$$

wobei $\mathbf{Jf}(\mathbf{x}^0) \in \mathbb{R}^{(m,n)}$ die Jacobi–Matrix von $\mathbf{f}$ in $\mathbf{x}^0$ bezeichnet.

c) Im Fall einer skalaren Funktion, $m = 1$, ist $\mathbf{J}f(x^0) = \boldsymbol{\nabla} f(\mathbf{x}^0)^\mathsf{T}$ ein Zeilenvektor und $\mathbf{J}f(\mathbf{x}^0)(\mathbf{x} - \mathbf{x}^0)$ ein Skalarprodukt: $\left\langle \boldsymbol{\nabla} f(\mathbf{x}^0), \mathbf{x} - \mathbf{x}^0 \right\rangle = \boldsymbol{\nabla} f(\mathbf{x}^0)^\mathsf{T}(\mathbf{x} - \mathbf{x}^0)$.

Satz (13.1.3)

Sei $\mathbf{f} : \mathbb{R}^n \supset D \to \mathbb{R}^m$, D offen und $\mathbf{x}^0 \in D$.

a) Ist $\mathbf{f}$ in $\mathbf{x}^0$ differenzierbar, so ist $\mathbf{f}$ in $\mathbf{x}^0$ auch stetig.

b) Ist $\mathbf{f}$ eine C^1–Funktion auf D, vgl. 12.3, so ist $\mathbf{f}$ auf D auch (vollständig) differenzierbar.

Bemerkung (13.1.4)

Die partielle Differenzierbarkeit impliziert daher *nicht* die (vollständige) Differenzierbarkeit, vgl. das Beispiel (12.1.10).

Beispiele (13.1.5)

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

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

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

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

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

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

$$\mathbf{Jf}(\mathbf{x}) = \mathbf{A}, \qquad \forall\ \mathbf{x} \in \mathbb{R}^n.$$

d) Für eine quadratisches Polynom

$$f(\mathbf{x}) := \frac{1}{2}\mathbf{x}^\top \mathbf{A}\,\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$$

mit $\mathbf{A} \in \mathbb{R}^{(n,n)}$, $\mathbf{b} \in \mathbb{R}^n$ und $c \in \mathbb{R}$ erhält man mit der verallgemeinerten Produktregel, vgl. (4.2.3)

$$\begin{aligned} \frac{\partial f}{\partial x_i} &= \tfrac{1}{2}\left(\mathbf{e}_i^\top \mathbf{A}\,\mathbf{x} + \mathbf{x}^\top \mathbf{A}\,\mathbf{e}_i\right) + b_i \\ &= \tfrac{1}{2}\,\mathbf{x}^\top(\mathbf{A} + \mathbf{A}^\top)\,\mathbf{e}_i + b_i. \end{aligned}$$

Damit ergibt sich für die Jacobi–Matrix

$$\mathbf{J}f(\mathbf{x}) = \boldsymbol{\nabla} f(\mathbf{x})^\top = \frac{1}{2}\,\mathbf{x}^\top(\mathbf{A} + \mathbf{A}^\top) + \mathbf{b}^\top.$$

Fehlerrechnung (13.1.6)

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

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

Vernachlässigt man für kleine Variationen $\|\mathbf{\Delta x}\|$ den Term $\mathbf{o}(\|\mathbf{\Delta x}\|)$, so ergibt sich „in erster Ordnung“ :

$$\mathbf{\Delta y} \;:=\; \mathbf{f}(\mathbf{x} + \Delta\mathbf{x}) \;-\; \mathbf{f}(\mathbf{x}) \;\approx\; \sum_{k=1}^{n} \frac{\partial \mathbf{f}}{\partial x_k}(\mathbf{x}) \;\Delta x_k\,. \tag{13.1.7}$$

Beispiel (13.1.8)

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

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 (13.1.7) in erster Näherung

$$\Delta S \;\approx\; 2\,(y+z)\,\Delta x \;+\; 2\,(x+z)\,\Delta y \;+\; 2\,(x+y)\,\Delta z$$

$$\Rightarrow \quad |\Delta S| \;\leq\; (64+60+44)\cdot 0.1\ \text{cm}^2 \;=\; 16.8\ \text{cm}^2\,.$$

Differentiationsregeln (13.1.9)

Linearität: $\mathbf{J}(\alpha\mathbf{f}+\beta\mathbf{g})\,(\mathbf{x}^0) \;=\; \alpha\,\mathbf{Jf}(\mathbf{x}^0)+\beta\,\mathbf{Jg}(\mathbf{x}^0)\,.$

Kettenregel: Ist $\mathbf{f} : D_f \to \mathbb{R}^m$ in $\mathbf{x}^0 \in D_f \subset \mathbb{R}^n$ (offen) diffb. und ist $\mathbf{g} : D_g \to \mathbb{R}^k$ diffb. in $\mathbf{y}^0 := \mathbf{f}(\mathbf{x}^0) \in D_g \subset \mathbb{R}^m$ (offen), so ist auch die Hintereinanderausführung $\mathbf{g}\circ\mathbf{f}$ in $\mathbf{x}^0$ diffb. und für die Jacobi–Matrizen gilt

$$\mathbf{J}(\mathbf{g}\circ\mathbf{f})(\mathbf{x}^0) \;=\; \mathbf{Jg}(\mathbf{f}(\mathbf{x}^0))\cdot\mathbf{Jf}(\mathbf{x}^0)\,.$$

Spezialfall der Kettenregel (13.1.10)

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

Dann ist auch die Hintereinanderausführung

$$f \circ \mathbf{c} : \; I \to \mathbb{R}, \qquad (f \circ \mathbf{c})\,(t) \;=\; f(c_1(t), \ldots, c_n(t))$$

in t_0 differenzierbar und für die Ableitung gilt

$$\begin{aligned} (f \circ \mathbf{c})'\,(t_0) \quad &= \; \mathbf{J}f(\mathbf{c}(t_0)) \cdot \mathbf{J}\mathbf{c}(t_0) \;=\; \boldsymbol{\nabla} f(\mathbf{c}(t_0))^{\mathsf{T}} \mathbf{c}'(t_0) \\ \Rightarrow \; (f(\mathbf{c}(t))'\,(t_0) \; &= \; \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\mathbf{c}(t_0)) \cdot c_k'(t_0)\,. \end{aligned}$$

Beispiel (13.1.11)

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

Anwendung der Kettenregel ergibt:

$$\begin{aligned}\frac{\mathrm{d}T}{\mathrm{d}t} &= \nabla T(m,v)^{\mathsf{T}} \cdot \begin{pmatrix} m'(t) \\ v'(t) \end{pmatrix} = \frac{1}{2}\, v^2 \cdot m' + m\, v \cdot v' \\ &= -\frac{1}{2}\alpha\, v^2 + 3\,\beta\, m\, v\, t^2.\end{aligned}$$

Setzt man hierin $m(t)$ und $v(t)$ ein, so folgt:

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

Richtungsableitungen

Für $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ offen, $\mathbf{x}^0 \in D$ und $\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ heißt

$$D_{\mathbf{v}} f(\mathbf{x}^0) := \lim_{t\to 0} \frac{f(\mathbf{x}^0 + t\mathbf{v}) - f(\mathbf{x}^0)}{t} \tag{13.1.12}$$

die **Richtungsableitung**, auch **Gâteaux–Ableitung**, von f in Richtung $\mathbf{v}$.

Bemerkungen (13.1.13)

a) Ist $\mathbf{v}$ ein Einheitsvektor, $\|\mathbf{v}\| = 1$, so gibt $D_{\mathbf{v}} f(\mathbf{x}^0)$ den Anstieg von f in $\mathbf{x}^0$ in Richtung $\mathbf{v}$ an.

b) $D_{\mathbf{e}_i} f(\mathbf{x}^0) \; = \; \dfrac{\partial f}{\partial x_i}(\mathbf{x}^0)\,, \; i = 1, \ldots, n\,.$

c) Ist f in $\mathbf{x}^0$ differenzierbar, so gilt $D_{\mathbf{v}} f(\mathbf{x}^0) \; = \; \dfrac{\mathrm{d}}{\mathrm{d}t}(f \circ \mathbf{c})(0)$, mit $\mathbf{c}(t) \; := \; \mathbf{x}^0 + t\,\mathbf{v}$. Mit der Kettenregel folgt

$$D_{\mathbf{v}} f(\mathbf{x}^0) \; = \; \nabla f(\mathbf{x}^0)^{\mathsf{T}}\, \mathbf{c}'(0) \; = \; \nabla f(\mathbf{x}^0)^{\mathsf{T}}\, \mathbf{v}\,. \qquad (13.1.14)$$

Satz (13.1.15) (Eigenschaften des Gradienten)

Sei $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ offen, in $\mathbf{x}^0 \in D$ differenzierbar.

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

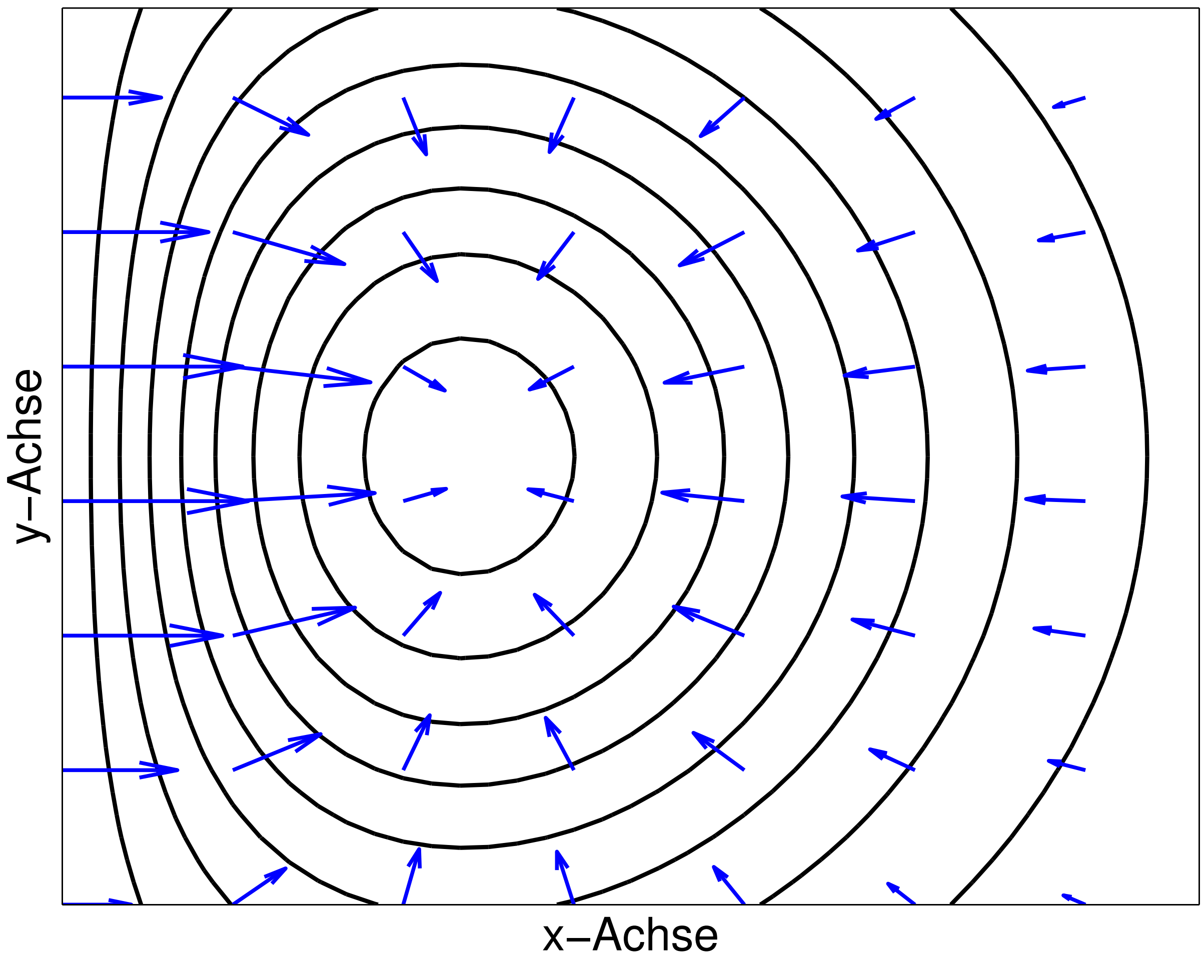
$$N_f(\mathbf{x}^0) \; := \; \{\mathbf{x} \in D : \; f(\mathbf{x}) = f(\mathbf{x}^0)\}\,.$$

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

Im Fall $n = 2$ nennt man die Niveaumengen $N_f(\mathbf{x}^0)$ auch **Höhenlinien**, im Fall $n = 3$ **Äquipotentialflächen**.

Bemerkung (13.1.16)

Die obige Eigenschaft des Gradienten (13.1.15) b) lässt sich für die numerische Berechnung der lokalen Minima einer Funktion f verwenden. Auf dieser Eigenschaft beruhen im die verschiedenen Varianten des **Gradientenverfahrens**, die auch **Verfahren des steilsten Abstiegs (steepest descent methods)** genannt werden.


y–Achse
x–Achse

Krummlinige Koordinaten

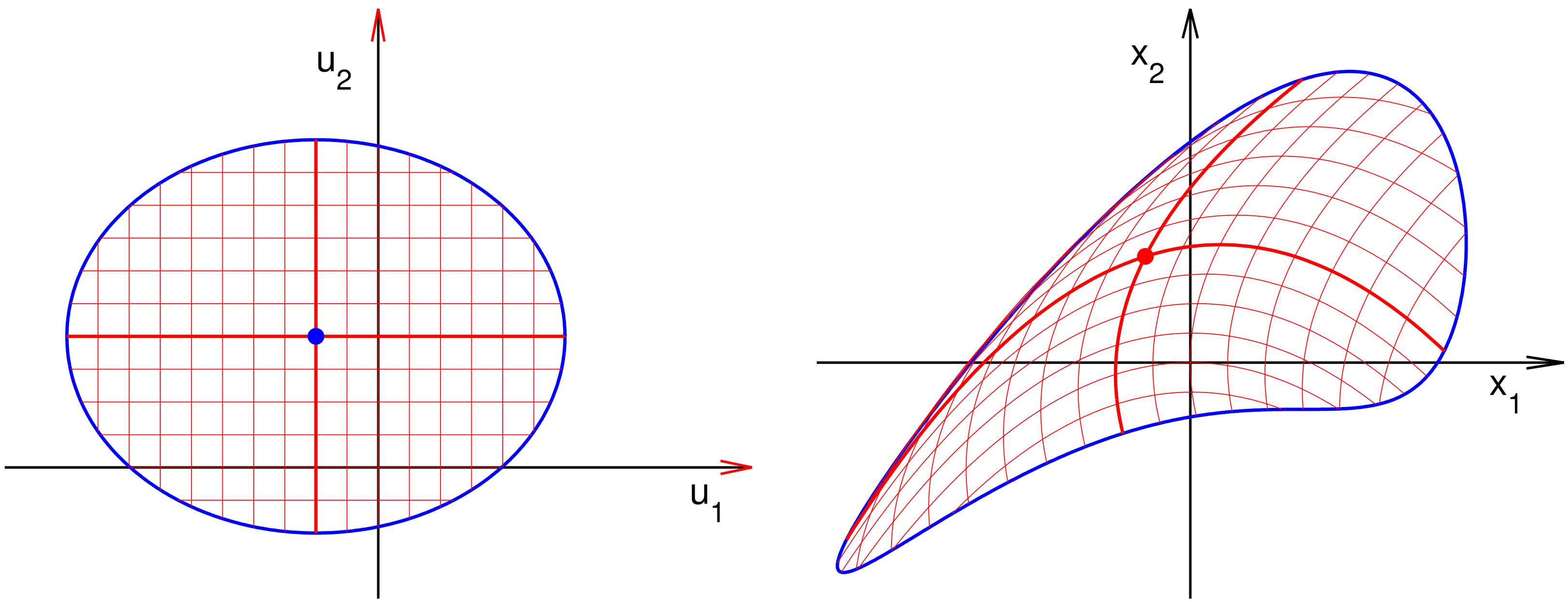
Wir betrachten allgemeine Koordinatentransformationen, die wir auch als nichtlineare Transformationen zulassen wollen. Man spricht dann von krummlinigen Koordinatensystemen.

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

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

Abbildungen mit diesen Eigenschaften heißen **(lokale) C^1–Diffeomorphismen**.

$$\mathbf{u} \;\mapsto\; \mathbf{x} = \mathbf{\Phi}(\mathbf{u})$$



Wir fragen, wie sich die Ableitungen einer Funktionen $y = \tilde{f}(\mathbf{x})$ bei einer Koordinatentransformation $\mathbf{x} = \mathbf{\Phi}(\mathbf{u})$ auf die neuen Koordinaten $\mathbf{u}$ umrechnen lassen. Es bezeichne $f(\mathbf{u}) := \tilde{f}(\mathbf{\Phi}(\mathbf{u}))$ nun die „gleiche“ Funktion $\tilde{f}$, ausgedrückt in den neuen Koordinaten $\mathbf{u}$.

Differentiation mittels Kettenregel ergibt:

$$\frac{\partial f}{\partial u_i} = \sum_{j=1}^{n} \frac{\partial \tilde{f}}{\partial x_j} \cdot \frac{\partial \mathbf{\Phi}_j}{\partial u_i} = \sum_{j=1}^{n} g^{ij} \cdot \frac{\partial \tilde{f}}{\partial x_j} .$$

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

$$g^{ij} := \frac{\partial \mathbf{\Phi}_j}{\partial u_i}, \quad \mathbf{G}(\mathbf{u}) := (g^{ij}) = (\mathbf{J}\mathbf{\Phi}(\mathbf{u}))^{\mathsf{T}} . \tag{13.1.17}$$

Für die obige Beziehung schreiben wir auch abkürzend:

$$\frac{\partial}{\partial u_i} = \sum_{j=1}^{n} g^{ij} \frac{\partial}{\partial x_j} \quad \text{oder} \quad \mathbf{\nabla_u} = \mathbf{G}\, \mathbf{\nabla_x}. \tag{13.1.18}$$

Umkehrung:

$$\frac{\partial}{\partial x_i} = \sum_{j=1}^{n} g_{ij} \frac{\partial}{\partial u_j} \quad \text{oder} \quad \boldsymbol{\nabla}_{\mathbf{x}} = \mathbf{G}^{-1} \boldsymbol{\nabla}_{\mathbf{u}}. \tag{13.1.19}$$

wobei

$$(g_{ij}) := (g^{ij})^{-1} = (\mathbf{J}\boldsymbol{\Phi})^{-\mathsf{T}} = (\mathbf{J}\boldsymbol{\Phi}^{-1})^{\mathsf{T}}. \tag{13.1.20}$$

Anwendung auf Polarkoordinaten

Das Polarkoordinatensystem ist gegeben durch

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

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

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

und somit

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

Nach (13.1.19) ergibt sich also für die Umrechnung der partiellen Ableitungen:

$$\begin{aligned} \frac{\partial}{\partial x_1} &= \cos\varphi\,\frac{\partial}{\partial r} - \frac{1}{r}\sin\varphi\,\frac{\partial}{\partial\varphi} \\ \frac{\partial}{\partial x_2} &= \sin\varphi\,\frac{\partial}{\partial r} + \frac{1}{r}\cos\varphi\,\frac{\partial}{\partial\varphi}. \end{aligned} \tag{13.1.21}$$

Beispiel (13.1.22)

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

Anwendung von (13.1.21) ergibt

$$\tilde{f}_x = \cos(\varphi)\, f_r - \frac{1}{r}\sin(\varphi)\, f_\varphi = \frac{x\,y}{r}$$

$$\tilde{f}_y = \sin(\varphi)\, f_r + \frac{1}{r}\cos(\varphi)\, f_\varphi = r + \frac{y^2}{r}\,.$$

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

Umrechnung von Δf auf Polarkoordinaten (im $\mathbb{R}^2$)

$$\frac{\partial^2}{\partial x_1^2} = \cos^2(\varphi)\frac{\partial^2}{\partial r^2} - \frac{\sin(2\varphi)}{r}\frac{\partial^2}{\partial r\partial\varphi} + \frac{\sin^2\varphi}{r^2}\frac{\partial^2}{\partial\varphi^2} + \frac{\sin(2\,\varphi)}{r^2}\frac{\partial}{\partial\varphi} + \frac{\sin^2\varphi}{r}\frac{\partial}{\partial r}$$

$$\frac{\partial^2}{\partial x_2^2} = \sin^2\varphi\,\frac{\partial^2}{\partial r^2} + \frac{\sin(2\varphi)}{r}\frac{\partial^2}{\partial r\partial\varphi} + \frac{\cos^2\varphi}{r^2}\frac{\partial^2}{\partial\varphi^2} - \frac{\sin(2\varphi)}{r^2}\frac{\partial}{\partial\varphi} + \frac{\cos^2\varphi}{r}\frac{\partial}{\partial r}$$

Damit folgt:

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} = \frac{\partial^2}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2} + \frac{1}{r}\frac{\partial}{\partial r}. \tag{13.1.23}$$

Anwendung auf Kugelkoordinaten im $\mathbb{R}^3$

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

Für die Jacobi-Matrix und die Transformationsmatrizen (g^{ij}), (g_{ij}) ergibt sich

$$\mathbf{J\Phi}(\mathbf{u}) = \begin{pmatrix} \cos\varphi \cos\theta & -r \sin\varphi \cos\theta & -r \cos\varphi \sin\theta \\ \sin\varphi \cos\theta & r \cos\varphi \cos\theta & -r \sin\varphi \sin\theta \\ \sin\theta & 0 & r \cos\theta \end{pmatrix},$$

$$(g^{ij}) = (\mathbf{J\Phi})(\mathbf{u}))^{\top},$$

$$
(g_{ij}) = \begin{pmatrix} \cos\varphi\cos\theta & -\dfrac{\sin\varphi}{r\cos\theta} & -\dfrac{1}{r}\cos\varphi\sin\theta \\ \sin\varphi\cos\theta & \dfrac{\cos\varphi}{r\cos\theta} & -\dfrac{1}{r}\sin\varphi\sin\theta \\ \sin\theta & 0 & \dfrac{1}{r}\cos\theta \end{pmatrix} .
$$

Für die partiellen Ableitungen erster Ordnung folgt somit:

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

Umrechnung von Δf auf Kugelkoordinaten (im $\mathbb{R}^3$)

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r^2 \cos^2\theta}\frac{\partial^2}{\partial \varphi^2} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} + \frac{2}{r}\frac{\partial}{\partial r} - \frac{\tan\theta}{r^2}\frac{\partial}{\partial \theta}. \tag{13.1.24}$$

13.2 Mittelwertsätze und Taylorscher Satz

Wir betrachten im Folgenden Verallgemeinerungen des ersten Mittelwertsatzes, vgl. Analysis 1, (5.1.5), auf Funktionen von mehreren unabhängigen Variablen.

Dazu sei $D \subset \mathbb{R}^n$ eine offene Menge. Zu zwei Punkten $\mathbf{a}, \mathbf{b} \in D$ wird die Verbindungsstrecke definiert durch:

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

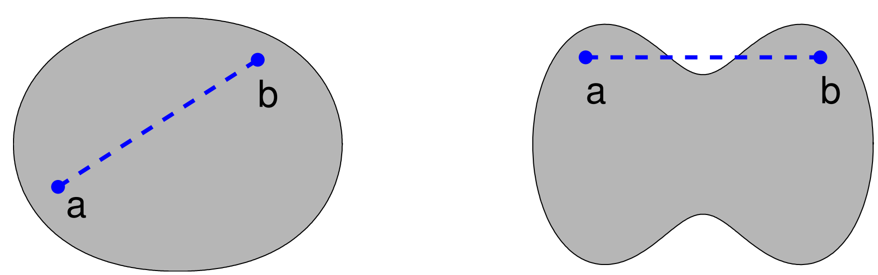
Satz (13.2.1) (Mittelwertsatz)

Sei $f : D \to \mathbb{R}$ eine differenzierbare, skalare Funktion. Ferner seien $\mathbf{a}, \mathbf{b} \in D$ zwei Punkte in D, für die die Verbindungsstrecke $[\mathbf{a}, \mathbf{b}]$ ganz in D liegt. Dann gibt es eine Zahl $\theta \in]0,1[$ mit

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

Definition (13.2.2)

Gilt die Voraussetzung $[\mathbf{a}, \mathbf{b}] \subset D$ für *alle* Punktepaare $\mathbf{a}$, $\mathbf{b} \in D$, so heißt die Menge D **konvex**.



Konvexe und nichtkonvexe Menge

Beispiel (13.2.3)

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

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

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

Bemerkung (13.2.4)

Der Mittelwertsatz (in Gleichungsform) lässt sich *nicht* auf den Fall vektorwertiger Funktionen übertragen!

Beispiel (13.2.5)

Sei $\mathbf{f}(t) := (\cos t, \sin t)^\mathsf{T}$ definiert auf $0 \le t \le \pi/2$. Es gilt

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

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

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

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.

Satz (13.2.6) (Mittelwert–Abschätzungssatz)

Sei $\mathbf{f} : D \to \mathbb{R}^m$ differenzierbar auf der offenen Menge $D \subset \mathbb{R}^n$.
Ferner seien $\mathbf{a}, \mathbf{b}$ Punkte in D mit $[\mathbf{a}, \mathbf{b}] \subset D$.
Dann existiert ein $\theta \in]0, 1[$ mit

$$||\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})||_2 \leq ||\mathbf{Jf}(\mathbf{a} + \theta\,(\mathbf{b} - \mathbf{a})) \cdot (\mathbf{b} - \mathbf{a})||_2 \,.$$

Bemerkung (13.2.7)

a) Die obige Abschätzung lässt sich abschwächen zu

(i) $$||\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})||_2 \leq ||\mathbf{Jf}(\mathbf{a} + \theta\,(\mathbf{b} - \mathbf{a})||_2 \cdot ||\mathbf{b} - \mathbf{a}||_2 \,,$$

(ii) $$||\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})||_2 \leq \sup_{\mathbf{x} \in [\mathbf{a}, \mathbf{b}]} ||\mathbf{Jf}(\mathbf{x})||_2 \cdot ||\mathbf{b} - \mathbf{a}||_2 \,,$$

wobei die Matrix–Normen analog zum quadratischen Fall definiert werden.

b) Ist $D \subset \mathbb{R}^n$ konvex und offen und ist $\mathbf{f} : D \to \mathbb{R}^m$ differenzierbar mit

$$L := \sup_{\mathbf{x} \in D} ||\mathbf{Jf}(\mathbf{x})||_2 < \infty,$$

so ist $\mathbf{f}$ Lipschitz–stetig auf D mit der Lipschitz-Konstanten L (bezogen auf $\|\cdot\|_2$).

c) Der Mittelwert–Abschätzungssatz gilt in abgeschwächter Form auch für beliebige Vektornormen und der jeweils zugehörigen Matrixnorm

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

Satz (13.2.8) (Satz von Taylor)

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

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

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

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

(Taylor–Polynom m–ten Grades)

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

(Restgliedformel nach Lagrange)

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

Erläuterung

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

Speziell für $n=2$ ergibt sich

$$[\,(\mathbf{x}-\mathbf{x}^0)^{\mathsf{T}}\boldsymbol{\nabla}\,]^j f \;=\; \sum_{i=0}^{j}\binom{j}{i}(x_1-x_1^0)^i\,(x_2-x_2^0)^{j-i}\frac{\partial^j f}{\partial x_1^i\;\partial x_2^{j-i}}\,(\mathbf{x}^0)\,.$$

Beispiel (13.2.9)

Man bestimme das Taylor–Polynom $T_2(\mathbf{x},\mathbf{x}^0)$ zweiten Grades der Funktion

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

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

Die Berechnung und Auswertung der partiellen Ableitungen von f bis zur zweiten Ordung liefert

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

Alternative: Man schreibe $x = (x-1)+1$, $y^2 = (y-2)^2 + 4(y-2)+4$, $\sin z = z - z^3/6 + \ldots$ und multipliziere aus!

Folgerung (13.2.10)

Für die **Approximationsgüte des Taylor-Polynoms** einer C^{m+1}–Funktion ergibt sich mit dem Taylorschen Satz:

$$|R_m(\mathbf{x};\mathbf{x}^0)| \;\le\; \frac{n^{m+1}}{(m+1)!}\cdot C\cdot\|\mathbf{x}-\mathbf{x}^0\|_\infty^{m+1}\,.$$

Hierbei ist C eine gemeinsame Schranke der partiellen Ableitungen $(m+1)$-ter Ordnung von f. - Alternativ:

$$f(\mathbf{x}) \;=\; T_m(\mathbf{x};\mathbf{x}^0) \;+\; O\left(\|\mathbf{x}-\mathbf{x}^0\|^{m+1}\right).$$