1.3 Differenzierbarkeit

Definition

Sei $B \subset \mathbb{R}^n$ offen, $\mathbf{a} \in B$, $f : B \to \mathbb{R}$ eine Funktion und $\mathbf{v} \neq \mathbf{0}$ ein beliebiger Vektor im $\mathbb{R}^n$. Wenn der Grenzwert

$$D_{\mathbf{v}} f(\mathbf{a}) := \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t}$$

existiert, so bezeichnet man ihn als die ***Richtungsableitung*** von f in $\mathbf{a}$ in Richtung $\mathbf{v}$. Man sagt dann auch, dass f im Punkte $\mathbf{a}$ ***in Richtung* v *differenzierbar*** ist.

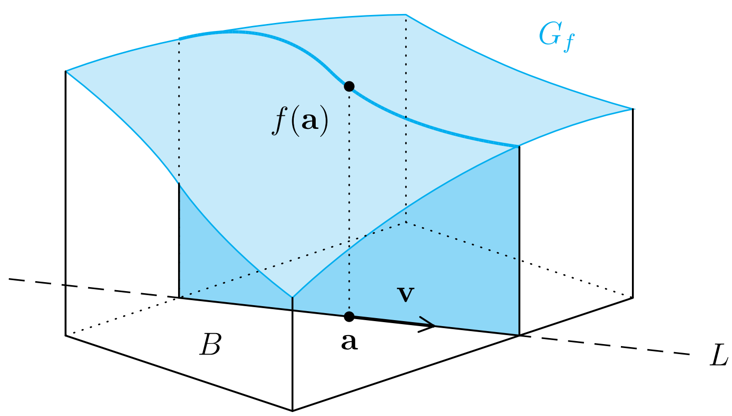
Was bedeutet das anschaulich?

Durch $\boldsymbol{\alpha}(t) := \mathbf{a} + t\mathbf{v}$ wird eine Gerade $L \subset \mathbb{R}^n$ parametrisiert, die bei $t = 0$ den Punkt $\mathbf{a}$ trifft und $\mathbf{v}$ als Richtungsvektor besitzt. Die Funktion

$$f_{\mathbf{a},\mathbf{v}}(t) := f \circ \boldsymbol{\alpha}(t) = f(\mathbf{a} + t\mathbf{v})$$

ist eine gewöhnliche Funktion einer Veränderlichen, und die Richtungsableitung $D_{\mathbf{v}} f(\mathbf{a})$ ist nichts anderes als die gewöhnliche Ableitung $(f_{\mathbf{a},\mathbf{v}})'(0)$.

Den Graphen von $f_{\mathbf{a},\mathbf{v}}$ erhält man, indem man den Graphen von f mit der über der Geraden L gelegenen Ebene $\{(\mathbf{x}, z) \in \mathbb{R}^n \times \mathbb{R} : \mathbf{x} \in L\}$ schneidet.



3.1. Beispiel

Sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch $f(x) := 1 - \mathbf{x} \bullet \mathbf{x}$, $\mathbf{a} \in \mathbb{R}^2$ ein beliebiger Punkt und $\mathbf{v} \neq \mathbf{0}$ eine beliebige Richtung. Dann ist

$$\begin{aligned} f_{\mathbf{a},\mathbf{v}}(t) &= f(\mathbf{a} + t\mathbf{v}) = 1 - (\mathbf{a} + t\mathbf{v}) \bullet (\mathbf{a} + t\mathbf{v}) \\ &= 1 - \mathbf{a} \bullet \mathbf{a} - (2\mathbf{a} \bullet \mathbf{v})t - (\mathbf{v} \bullet \mathbf{v})t^2, \end{aligned}$$

also $D_{\mathbf{v}} f(\mathbf{a}) = (f_{\mathbf{a},\mathbf{v}})'(0) = -2\mathbf{a} \bullet \mathbf{v}$.

Eine besondere Rolle spielen die Ableitungen in Richtung der Einheitsvektoren.

Definition

Die Funktion f besitze in $\mathbf{a}$ eine Richtungsableitung in Richtung des i–ten Einheits–Vektors $\mathbf{e}_i$. Dann sagt man, f ist in $\mathbf{a}$ nach x_i ***partiell differenzierbar***, und die Zahl

$$\boxed{D_i f(\mathbf{a}) := D_{\mathbf{e}_i} f(\mathbf{a}) = \lim_{t \to 0} \frac{1}{t}\big(f(\mathbf{a} + t\mathbf{e}_i) - f(\mathbf{a})\big)}$$

heißt die i–te ***partielle Ableitung*** von f in $\mathbf{a}$. Statt $D_i f(\mathbf{a})$ schreibt man auch

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) \quad \text{oder} \quad f_{x_i}(\mathbf{a}).$$

Wenn alle partiellen Ableitungen von f in $\mathbf{a}$ existieren, dann heißt f in $\mathbf{a}$ ***partiell differenzierbar***.

Ist $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}$ in allen Punkten von B partiell differenzierbar, so bilden die partiellen Ableitungen $D_i f$ wieder reellwertige Funktionen auf B. Sind sie alle in einem Punkt $\mathbf{a} \in B$ stetig, so nennt man f in $\mathbf{a}$ ***stetig partiell differenzierbar***.

In der Praxis funktioniert die Bildung partieller Ableitungen folgendermaßen:

Um $\dfrac{\partial f}{\partial x_i}(a_1, \ldots, a_n)$ zu bilden, so differenziere man die Funktion

$$t \mapsto f(a_1, a_2, \ldots, a_{i-1}, a_i + t, a_{i+1}, \ldots, a_n) \quad \text{bei } t = 0,$$

bzw. die Funktion

$$x \mapsto f(a_1, a_2, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_n) \quad \text{bei } x = a_i.$$

Soll etwa $f(x, y) = \sin(x^2 y)$ bei $(x, y) = (x_0, y_0)$ partiell differenziert werden, so erhält man:

$$\frac{\partial f}{\partial x}(x_0, y_0) = \frac{d}{dx}\Big|_{x = x_0} \sin(x^2 y_0) = \cos(x_0^2 y_0) \cdot 2x_0 y_0$$

und

$$\frac{\partial f}{\partial y}(x_0, y_0) = \frac{d}{dy}\Big|_{y = y_0} \sin(x_0^2 y) = \cos(x_0^2 y_0) \cdot x_0^2.$$

3.2. Beispiel

Sei $f(x, y, z) := x^n \cos y + e^{x^2 + y^2 + z^2} \sin y - xyze^{2x - 3y}$. Dann ist

$$\begin{aligned}
D_1 f(x,y,z) &= \frac{\partial f}{\partial x}(x,y,z) &&= nx^{n-1}\cos y + 2xe^{x^2+y^2+z^2}\sin y \\
&&& \quad -yze^{2x-3y} - 2xyze^{2x-3y}, \\
D_2 f(x,y,z) &= \frac{\partial f}{\partial y}(x,y,z) &&= -x^n \sin y + 2ye^{x^2+y^2+z^2}\sin y \\
&&& \quad +e^{x^2+y^2+z^2}\cos y - xze^{2x-3y} \\
&&& \quad +3xyze^{2x-3y} \\
\text{und} \quad D_3 f(x,y,z) &= \frac{\partial f}{\partial z}(x,y,z) &&= 2ze^{x^2+y^2+z^2}\sin y - xye^{2x-3y}.
\end{aligned}$$

Leider hat die partielle Differenzierbarkeit einer Funktion noch nicht einmal deren Stetigkeit zur Folge. Betrachten wir dazu die Funktion

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

Die Funktionen $x \mapsto f(x,0) \equiv 0$ und $y \mapsto f(0,y) \equiv 0$ sind natürlich im Nullpunkt differenzierbar. Also ist f in $\mathbf{0} = (0,0)$ partiell differenzierbar. Andererseits ist f dort nicht stetig. Ist nämlich (a_ν) eine Nullfolge, so konvergiert $\mathbf{x}_\nu := ((a_\nu)^2, a_\nu)$ gegen $(0,0)$, aber es ist

$$\lim_{\nu\to\infty} f(\mathbf{x}_\nu) = \lim_{\nu\to\infty} \frac{(a_\nu)^4}{2(a_\nu)^4} = \frac{1}{2} \neq f(0,0).$$

Eine weitere Schwäche der partiellen Differenzierbarkeit tritt auf, wenn man höhere Ableitungen betrachtet.

Definition

Sei $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}$ überall partiell differenzierbar. Sind alle partiellen Ableitungen $D_j f$ in einem Punkt $\mathbf{a} \in B$ wiederum partiell differenzierbar, so definiert man für $i, j = 1, \ldots, n$:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) := D_i(D_j f)(\mathbf{a}).$$

Man nennt diesen Ausdruck die ***2-te partielle Ableitung*** von f nach x_i und x_j an der Stelle $\mathbf{a}$ und schreibt dafür auch $f_{x_i x_j}(\mathbf{a})$.

Bei zweiten partiellen Ableitungen werden die einzelnen Ableitungen von rechts nach links abgearbeitet, d.h. es ist $f_{x_i x_j} = (f_{x_j})_{x_i}$.

Sei z.B. $f(x_1, x_2) := e^{kx_1} \cdot \cos x_2$. Dann gilt:

$$\frac{\partial f}{\partial x_1}(\mathbf{x}) = k \cdot e^{kx_1} \cdot \cos x_2 \quad \text{und} \quad \frac{\partial f}{\partial x_2}(\mathbf{x}) = -e^{kx_1} \cdot \sin x_2,$$

sowie $\quad \dfrac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) = \dfrac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) = -k e^{k a_1} \sin a_2.$

Im vorliegenden Beispiel konnten die Ableitungen miteinander vertauscht werden. Leider ist das nicht generell der Fall.

3.3. Beispiel

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

Dann gilt für $(x,y) \neq (0,0)$:

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= \frac{\partial}{\partial x}\left(\frac{x^3y - y^3x}{x^2+y^2}\right) \\ &= \frac{(3x^2y - y^3)(x^2+y^2) - (x^3y - y^3x)2x}{(x^2+y^2)^2} \\ &= \frac{x^4y + 4x^2y^3 - y^5}{(x^2+y^2)^2}, \end{aligned}$$

also $\dfrac{\partial f}{\partial x}(0,y) = -y$ für $y \neq 0$.

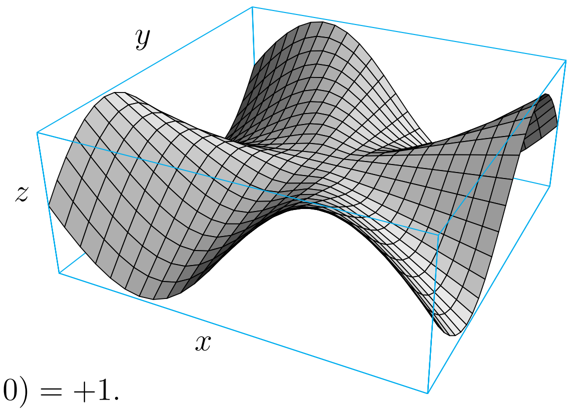
Weiter ist $\dfrac{\partial f}{\partial x}(0,0) = \lim\limits_{x \to 0} \dfrac{f(x,0) - f(0,0)}{x} = 0.$

Also ist sogar $\dfrac{\partial f}{\partial x}(0,y) \equiv -y$ für alle y und $\dfrac{\partial^2 f}{\partial y \partial x}(0,0) = -1.$

Entsprechend erhalten wir für $(x,y) \neq (0,0)$:

$$\begin{aligned} \frac{\partial f}{\partial y}(x,y) &= \frac{\partial}{\partial y}\left(\frac{x^3y - y^3x}{x^2+y^2}\right) \\ &= \frac{(x^3 - 3y^2x)(x^2+y^2) - (x^3y - y^3x)2y}{(x^2+y^2)^2} \\ &= \frac{x^5 - 4x^3y^2 - xy^4}{(x^2+y^2)^2}, \end{aligned}$$

also $\dfrac{\partial f}{\partial y}(x,0) \equiv x$ für $x \neq 0$, und $\dfrac{\partial f}{\partial y}(0,0) = \lim\limits_{y \to 0} \dfrac{f(0,y) - f(0,0)}{y} = 0.$



Somit ist $\dfrac{\partial^2 f}{\partial x \partial y}(0,0) = +1.$

Zum Glück gilt folgendes hinreichende Kriterium für die Gleichheit der gemischten zweiten Ableitungen:

3.4. Satz von Schwarz

Sei $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}$ auf ganz B nach allen Variablen partiell differenzierbar, $\mathbf{x}_0 \in B$.

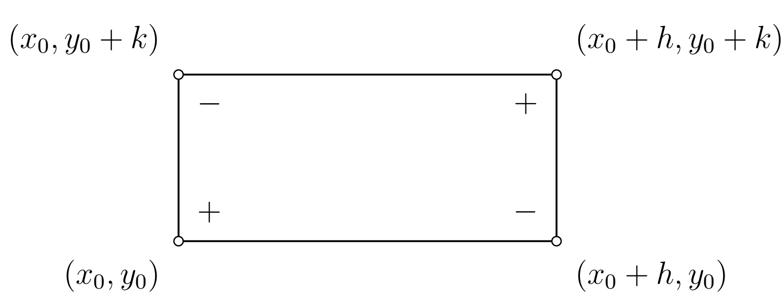
Wenn die gemischten zweiten Ableitungen $D_i(D_j f)$ und $D_j(D_i f)$ auf einer Umgebung von $\mathbf{x}_0$ in B existieren und in $\mathbf{x}_0$ stetig sind, so ist

$$D_i(D_j f)(\mathbf{x}_0) = D_j(D_i f)(\mathbf{x}_0).$$

Beweis: Es reicht, den Fall $n = 2$ zu betrachten. Wir bezeichnen die Variablen mit x und y und betrachten f in der Nähe eines Punktes (x_0, y_0), in dem f stetig partiell differenzierbar ist.

Für kleines $\varepsilon > 0$ ist f auf dem Rechteck $[x_0 - \varepsilon, x_0 + \varepsilon] \times [y_0 - \varepsilon, y_0 + \varepsilon]$ definiert und zweimal partiell differenzierbar. Für $0 < h < \varepsilon$ und $0 < k < \varepsilon$ betrachten wir die Größe

$$F(h,k) := f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) - f(x_0, y_0 + k) + f(x_0, y_0).$$



Wir halten h und k fest und setzen

$$\varphi(x) := f(x, y_0 + k) - f(x, y_0) \quad \text{und} \quad \psi(y) := f(x_0 + h, y) - f(x_0, y).$$

Die Funktionen φ und ψ sind für $x_0 \le x \le x_0 + h$ bzw. $y_0 \le y \le y_0 + k$ differenzierbar. Eine zweimalige Anwendung des Mittelwertsatzes liefert Zahlen $c, \widetilde{c}$ zwischen x_0 und $x_0 + h$ und Zahlen $d, \widetilde{d}$ zwischen y_0 und $y_0 + k$, so dass gilt:

$$\begin{aligned} F(h,k) &= \varphi(x_0 + h) - \varphi(x_0) \;=\; \varphi'(c) \cdot h \\ &= \big(D_1 f(c, y_0 + k) - D_1 f(c, y_0)\big) \cdot h \\ &= D_2 D_1 f(c, d) \cdot hk \end{aligned}$$

und

$$\begin{aligned} F(h,k) &= \psi(y_0 + k) - \psi(y_0) \;=\; \psi'(\widetilde{d}) \cdot k \\ &= \big(D_2 f(x_0 + h, \widetilde{d}) - D_2 f(x_0, \widetilde{d})\big) \cdot k \\ &= D_1 D_2 f(\widetilde{c}, \widetilde{d}) \cdot hk. \end{aligned}$$

Also ist $D_2 D_1 f(c, d) = D_1 D_2 f(\widetilde{c}, \widetilde{d})$, und wegen der Stetigkeit der zweiten Ableitungen in (x_0, y_0) erhält man beim Grenzübergang $h \to 0$ und $k \to 0$ die Gleichung $D_2 D_1 f(x_0, y_0) = D_1 D_2 f(x_0, y_0)$. ∎

Es genügt übrigens schon, dass **eine** der beiden gemischten Ableitungen in der Nähe von $\mathbf{x}_0$ existiert und in $\mathbf{x}_0$ stetig ist. Dann kann man die Existenz der anderen Ableitung und die Gleichheit beweisen.

Wir wollen jetzt den Differenzierbarkeitsbegriff noch einmal überdenken. Dazu erinnern wir uns an die Situation in einer Veränderlichen.

Zur Erinnerung Sei $I \subset \mathbb{R}$ ein offenes Intervall, $t_0 \in I$ und $f : I \to \mathbb{R}$ eine Funktion. Ist f in t_0 differenzierbar, so gibt es eine Darstellung

$$f(t) = f(t_0) + f'(t_0) \cdot (t - t_0) + \delta(t) \cdot (t - t_0),$$

mit $\lim\limits_{t \to t_0} \delta(t) = 0$.

Dann ist $L : \mathbb{R} \to \mathbb{R}$ mit $L(h) := f'(t_0) \cdot h$ eine lineare Abbildung. Setzt man $r(h) := \delta(t_0 + h) \cdot h$, so ist r in der Nähe von $h = 0$ definiert,

$$f(t_0 + h) = f(t_0) + L(h) + r(h)$$

und

$$\lim_{h \to 0} \frac{r(h)}{|h|} = \lim_{h \to 0} \delta(t_0 + h) \cdot \frac{h}{|h|} = 0.$$

Dies wollen wir auf den Fall von mehreren Veränderlichen verallgemeinern.

Definition

Sei $B \subset \mathbb{R}^n$ offen, $f : B \to \mathbb{R}$ eine Funktion und $\mathbf{x}_0 \in B$ ein Punkt. f heißt in $\mathbf{x}_0$ ***(total) differenzierbar***, wenn es eine Linearform $L : \mathbb{R}^n \to \mathbb{R}$ und eine auf einer Umgebung $U = U(\mathbf{0})$ definierte Funktion r gibt, so dass in der Nähe von $\mathbf{x}_0$ gilt:

1. $f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + L(\mathbf{h}) + r(\mathbf{h})$.

2. $\displaystyle\lim_{\mathbf{h}\to\mathbf{0}} \frac{r(\mathbf{h})}{\|\mathbf{h}\|} = 0$.

Die Linearform $Df(\mathbf{x}_0) := L$ (die man manchmal auch mit $(df)_{\mathbf{x}_0}$ bezeichnet) nennt man die ***(totale) Ableitung*** (oder das ***(totale) Differential***) von f in $\mathbf{x}_0$.

3.5. Satz

Sei $B \subset \mathbb{R}^n$ offen, $\mathbf{x}_0 \in B$ und $f : B \to \mathbb{R}$ eine Funktion. Dann sind die folgenden Aussagen äquivalent:

1. *f ist in $\mathbf{x}_0$ total differenzierbar.*

2. *Es gibt eine Linearform $L : \mathbb{R}^n \to \mathbb{R}$, so dass gilt:*

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = 0\,.$$

3. *Es gibt einen Vektor $\mathbf{a} \in \mathbb{R}^n$ und eine Funktion $\boldsymbol{\delta} : B \to \mathbb{R}^n$, so dass gilt:*

   (a) $f(\mathbf{x}) = f(\mathbf{x}_0) + \mathbf{a} \bullet (\mathbf{x} - \mathbf{x}_0) + \boldsymbol{\delta}(\mathbf{x}) \bullet (\mathbf{x} - \mathbf{x}_0)$.

   (b) $\displaystyle\lim_{\mathbf{x}\to\mathbf{x}_0} \boldsymbol{\delta}(\mathbf{x}) = \mathbf{0}$.

4. *Es gibt eine in $\mathbf{x}_0$ stetige Abbildung $\boldsymbol{\Delta} : B \to \mathbb{R}^n$ mit*

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\Delta}(\mathbf{x})^\top\,.$$

Gilt eine der äquivalenten Aussagen, so ist

$$L(\mathbf{v}) = \mathbf{a} \bullet \mathbf{v} = Df(\mathbf{x}_0)(\mathbf{v}) = \mathbf{v} \cdot \boldsymbol{\Delta}(\mathbf{x}_0)^\top.$$

BEWEIS: (1) $\Longrightarrow$ (4): Ist f in $\mathbf{x}_0$ differenzierbar, so gibt es eine Linearform L und eine Funktion r mit $\lim_{\mathbf{h}\to\mathbf{0}} r(\mathbf{h})/\|\mathbf{h}\| = 0$, so dass gilt:

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + L(\mathbf{h}) + r(\mathbf{h}).$$

Dann gibt es einen Vektor $\mathbf{a}$, so dass $L(\mathbf{h}) = \mathbf{a} \bullet \mathbf{h}$ ist, und wir setzen

$$\boldsymbol{\Delta}(\mathbf{x}_0 + \mathbf{h}) := \mathbf{a} + \frac{\mathbf{h}}{\|\mathbf{h}\|^2} \cdot r(\mathbf{h}) \text{ (zunächst für } \mathbf{h} \neq \mathbf{0}).$$

Offensichtlich gilt

$$\lim_{\mathbf{h} \to \mathbf{0}} \boldsymbol{\Delta}(\mathbf{x}_0 + \mathbf{h}) = \mathbf{a} + \lim_{\mathbf{h} \to \mathbf{0}} \frac{\mathbf{h}}{\|\mathbf{h}\|} \cdot \frac{r(\mathbf{h})}{\|\mathbf{h}\|} = \mathbf{a}.$$

Damit kann $\boldsymbol{\Delta}$ in $\mathbf{x}_0$ durch den Wert $\boldsymbol{\Delta}(\mathbf{x}_0) := \mathbf{a}$ stetig fortgesetzt werden, und es ist $f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) = L(\mathbf{h}) + r(\mathbf{h}) = \boldsymbol{\Delta}(\mathbf{x}_0 + \mathbf{h}) \bullet \mathbf{h} = \mathbf{h} \cdot \boldsymbol{\Delta}(\mathbf{x}_0 + \mathbf{h})^\top$.

(4) $\Longrightarrow$ (3): Trifft Aussage (4) zu, so setzen wir $\mathbf{a} := \boldsymbol{\Delta}(\mathbf{x}_0)$ und $\boldsymbol{\delta}(\mathbf{x}) := \boldsymbol{\Delta}(\mathbf{x}) - \mathbf{a}$. Dann ist $\lim_{\mathbf{x} \to \mathbf{x}_0} \boldsymbol{\delta}(\mathbf{x}) = \boldsymbol{\Delta}(\mathbf{x}_0) - \mathbf{a} = \mathbf{0}$ und $f(\mathbf{x}) - f(\mathbf{x}_0) = \boldsymbol{\Delta}(\mathbf{x}) \bullet (\mathbf{x} - \mathbf{x}_0) = \mathbf{a} \bullet (\mathbf{x} - \mathbf{x}_0) + \boldsymbol{\delta}(\mathbf{x}) \bullet (\mathbf{x} - \mathbf{x}_0)$.

(3) $\Longrightarrow$ (2): Sind $\mathbf{a}$ und $\boldsymbol{\delta}$ wie in (3) gegeben, so setzen wir $L(\mathbf{h}) := \mathbf{a} \bullet \mathbf{h}$. Dann ist

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = \lim_{\mathbf{x} \to \mathbf{x}_0} \boldsymbol{\delta}(\mathbf{x}) \cdot \frac{\mathbf{x} - \mathbf{x}_0}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$$

(2) $\Longrightarrow$ (1): Ist die Linearform L mit der Eigenschaft (2) gegeben, so setzen wir $r(\mathbf{h}) := f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - L(\mathbf{h})$. Damit ist die gewünschte Gleichung gegeben, und aus der Voraussetzung folgt auch, dass $\lim_{\mathbf{h} \to \mathbf{0}} r(\mathbf{h})/\|\mathbf{h}\| = 0$ ist. Also ist f in $\mathbf{x}_0$ total differenzierbar.

Die Zusatz-Aussage ist nun offensichtlich. ■

Partiell differenzierbare Funktionen brauchen nicht stetig zu sein. Dagegen gilt:

3.6. Total differenzierbare Funktionen sind stetig

Ist f in $\mathbf{x}_0$ total differenzierbar, so ist f dort auch stetig.

Beweis: Wir haben $f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\Delta}(\mathbf{x})^\top$, mit einer in $\mathbf{x}_0$ stetigen Funktion $\boldsymbol{\Delta}$. Also ist f in $\mathbf{x}_0$ stetig. ■

Die Bildung der partiellen Ableitung $D_i f$ kann man auch als Anwendung eines „linearen Operators“ D_i auf die Funktion f verstehen. Man fasst nun gerne die n partiellen Ableitungs–Operatoren zu einem vektoriellen Operator zusammen:

$$\nabla := (D_1, \ldots, D_n) \quad \text{(gesprochen: „Nabla“).}$$

Ist $f : B \to \mathbb{R}$ eine partiell differenzierbare Funktion, so heißt der Vektor

$$\mathbf{grad}\, f(\mathbf{a}) = \nabla f(\mathbf{a}) = \Big(\frac{\partial f}{\partial x_1}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{a}) \Big)$$

der ***Gradient*** von f im Punkt $\mathbf{a}$.

3.7. Berechnung der totalen Ableitung

Sei $f : B \to \mathbb{R}$ in $\mathbf{x}_0 \in B$ differenzierbar. Dann existieren in $\mathbf{x}_0$ sämtliche Richtungsableitungen von f, die Ableitung $Df(\mathbf{x}_0)$ ist eindeutig bestimmt und für jeden Richtungsvektor $\mathbf{v} \neq \mathbf{0}$ ist

$$Df(\mathbf{x}_0)(\mathbf{v}) = D_{\mathbf{v}}f(\mathbf{x}_0).$$

Insbesondere ist f in $\mathbf{x}_0$ nach allen Variablen partiell differenzierbar und für alle $\mathbf{v} \in \mathbb{R}^n$ ist

$$Df(\mathbf{x}_0)(\mathbf{v}) = \nabla f(\mathbf{x}_0) \bullet \mathbf{v} = \sum_{\nu=1}^{n} v_\nu \frac{\partial f}{\partial x_\nu}(\mathbf{x}_0).$$

Beweis: Ist f in $\mathbf{x}_0$ differenzierbar, so haben wir eine Darstellung

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \mathbf{a} \bullet (\mathbf{x} - \mathbf{x}_0) + \boldsymbol{\delta}(\mathbf{x}) \bullet (\mathbf{x} - \mathbf{x}_0) \text{ mit } \lim_{\mathbf{x} \to \mathbf{x}_0} \boldsymbol{\delta}(\mathbf{x}) = \mathbf{0}.$$

Ist $\mathbf{v}$ ein beliebiger Richtungsvektor, so strebt

$$\frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0)}{t} = \mathbf{a} \bullet \mathbf{v} + \boldsymbol{\delta}(\mathbf{x}_0 + t\mathbf{v}) \bullet \mathbf{v}$$

für $t \to 0$ gegen $\mathbf{a} \bullet \mathbf{v}$. Das bedeutet: Es existiert die Richtungsableitung $D_{\mathbf{v}}f(\mathbf{x}_0) = \mathbf{a} \bullet \mathbf{v} = Df(\mathbf{x}_0)(\mathbf{v})$.

Ist $\mathbf{a} = (a_1, \ldots, a_n)$, so ist $a_i = \mathbf{a} \bullet \mathbf{e}_i = Df(\mathbf{x}_0)(\mathbf{e}_i) = D_i f(\mathbf{x}_0)$, also $\mathbf{a} = \nabla f(\mathbf{x}_0)$. ■

Wir stehen nun vor folgendem Dilemma: Wir wissen, wie man partielle Ableitungen berechnet, und mit Hilfe dieser Ableitungen erhalten wir auch die totale Ableitung. Bevor wir aber die Ableitung einer Funktion f mit Hilfe der partiellen Ableitungen ausrechnen können, müssen wir die totale Differenzierbarkeit beweisen. Zum Glück gibt es dafür ein hinreichendes Kriterium. Um das zu beweisen, brauchen wir einen Hilfssatz:

3.8. Lemma (schwacher Mittelwertsatz im $\mathbb{R}^n$)

Sei $f : U_\varepsilon(\mathbf{x}_0) \to \mathbb{R}$ partiell differenzierbar und $\mathbf{x} \in U_\varepsilon(\mathbf{x}_0)$ beliebig. Die Punkte $\mathbf{z}_0, \ldots, \mathbf{z}_n$ seien definiert durch $\mathbf{z}_0 := \mathbf{x}_0$ und $\mathbf{z}_i := \mathbf{z}_{i-1} + (x_i - x_i^{(0)}) \cdot \mathbf{e}_i$ für $i = 1, \ldots, n$.

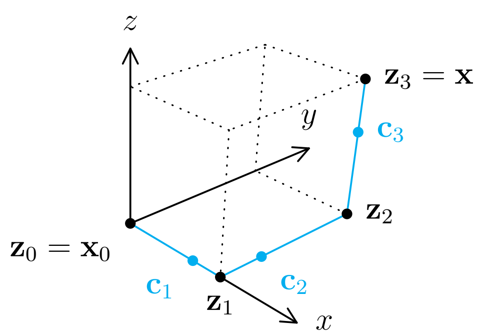
Dann liegen alle $\mathbf{z}_i$ und die Verbindungsstrecken von $\mathbf{z}_{i-1}$ nach $\mathbf{z}_i$ in $U_\varepsilon(\mathbf{x}_0)$, und auf jeder dieser Verbindungsstrecken gibt es einen Punkt $\mathbf{c}_i$, so dass gilt:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{c}_i) \cdot (x_i - x_i^{(0)}) \,.$$

Beweis: Es ist $\mathbf{z}_i = (x_1, \dots, x_i, x_{i+1}^{(0)}, \dots, x_n^{(0)})$, also

$$\|\mathbf{z}_i - \mathbf{x}_0\| = \|(x_1 - x_1^{(0)}, \dots, x_i - x_i^{(0)}, 0, \dots, 0)\| \leq \|\mathbf{x} - \mathbf{x}_0\| < \varepsilon.$$

Das bedeutet, dass alle $\mathbf{z}_i$ in $U_\varepsilon(\mathbf{x}_0)$ liegen. Es ist klar, dass dann auch die Verbindungsstrecken in $U_\varepsilon(\mathbf{x}_0)$ liegen.



Sei $\sigma_i : [0,1] \to \mathbb{R}$ definiert durch

$$\sigma_i(t) := x_i^{(0)} + t(x_i - x_i^{(0)})$$

und $F_i : (x_i^{(0)} - \varepsilon, x_i^{(0)} + \varepsilon) \to \mathbb{R}$ definiert durch

$$F_i(s) := f(x_1, \dots, x_{i-1}, s, x_{i+1}^{(0)}, \dots, x_n^{(0)}) .$$

Beide Funktionen sind differenzierbar, also auch

$$f_i(t) := F_i \circ \sigma_i(t) = f(x_1, \dots, x_{i-1}, \sigma_i(t), x_{i+1}^{(0)}, \dots, x_n^{(0)}) = f(\mathbf{z}_{i-1} + t(\mathbf{z}_i - \mathbf{z}_{i-1})) .$$

Nach dem Mittelwertsatz gibt es ein $\xi_i \in (0,1)$ mit $f_i(1) - f_i(0) = f_i'(\xi_i)$. Dabei ist $f_i(0) = f(\mathbf{z}_{i-1})$, $f_i(1) = f(\mathbf{z}_i)$ und

$$f_i'(\xi_i) = F_i'(\sigma_i(\xi_i)) \cdot \sigma_i'(\xi_i) = f_{x_i}(\mathbf{z}_{i-1} + \xi_i(\mathbf{z}_i - \mathbf{z}_{i-1})) \cdot (x_i - x_i^{(0)}).$$

Setzen wir $\mathbf{c}_i := \mathbf{z}_{i-1} + \xi_i(\mathbf{z}_i - \mathbf{z}_{i-1})$, so ist

$$\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{c}_i) \cdot (x_i - x_i^{(0)}) = \sum_{i=1}^{n} (f(\mathbf{z}_i) - f(\mathbf{z}_{i-1})) = f(\mathbf{x}) - f(\mathbf{x}_0) .$$

∎

3.9. Hinreichendes Kriterium für die Differenzierbarkeit

Sei $B \subset \mathbb{R}^n$ offen, $f : B \to \mathbb{R}$ eine Funktion und $\mathbf{x}_0 \in B$ ein Punkt. Wenn es eine offene Umgebung U von $\mathbf{x}_0$ in B gibt, so dass alle partiellen Ableitungen von f auf U existieren und in $\mathbf{x}_0$ stetig sind, dann ist f in $\mathbf{x}_0$ total differenzierbar.

Beweis: Wir benutzen den schwachen Mittelwertsatz. Existieren die partiellen Ableitungen von f auf $U = U_\varepsilon(\mathbf{x}_0)$, so definieren wir zu jedem $\mathbf{x} \in U$ die Punkte $\mathbf{z}_0 := \mathbf{x}_0$ und $\mathbf{z}_i := \mathbf{z}_{i-1} + (x_i - x_i^{(0)}) \cdot \mathbf{e}_i$ für $i = 1, \ldots, n$. Dann gibt es Punkte $\mathbf{c}_i$ zwischen $\mathbf{z}_{i-1}$ und $\mathbf{z}_i$ mit

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{c}_i) \cdot (x_i - x_i^{(0)}) .$$

Wir setzen

$$\boldsymbol{\Delta}(\mathbf{x}) := \Big(\frac{\partial f}{\partial x_1}(\mathbf{c}_1), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{c}_n) \Big).$$

Strebt $\mathbf{x}$ gegen $\mathbf{x}_0$, so streben auch alle Punkte $\mathbf{c}_i$ gegen $\mathbf{x}_0$. Deshalb ist $\boldsymbol{\Delta}$ stetig in $\mathbf{x}_0$. Weil außerdem $f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\Delta}(\mathbf{x})^\top$ ist, folgt die Differenzierbarkeit von f in $\mathbf{x}_0$. ∎

Man beachte, dass die Stetigkeit der partiellen Ableitungen für die totale Differenzierbarkeit nicht notwendig ist!

3.10. Beispiele

A. Sei $f(\mathbf{x}) \equiv c$ konstant. Dann verschwinden alle partiellen Ableitungen, und da die Nullfunktion stetig ist, ist f total differenzierbar und $Df(\mathbf{x}) = 0$ (die „Null-Form") in jedem Punkt $\mathbf{x}$ des $\mathbb{R}^n$.

B. Sei $f(\mathbf{x}) := \mathbf{u} \bullet \mathbf{x} = u_1 x_1 + \cdots + u_n x_n$ selbst schon eine Linearform. Dann ist $f_{x_i}(\mathbf{x}) \equiv u_i$ konstant (und damit stetig) für alle i. Also ist f überall total differenzierbar, und offensichtlich ist $Df(\mathbf{x})(\mathbf{v}) = \mathbf{u} \bullet \mathbf{v} = f(\mathbf{v})$ für jedes $\mathbf{x}$. Die Ableitung einer Linearform f stimmt in jedem Punkt $\mathbf{x}$ des $\mathbb{R}^n$ mit genau dieser Linearform überein: $Df(\mathbf{x}) = f$ für alle $\mathbf{x}$.

Ein Spezialfall ist die Linearform $x_i : \mathbf{v} \mapsto \mathbf{e}_i \bullet \mathbf{v} = v_i$. In jedem Punkt $\mathbf{x}$ ist das ***Differential*** $(dx_i)_\mathbf{x} = Dx_i(\mathbf{x})$ die Projektion auf die i-te Komponente.

C. Nun sei $A = (a_{ij}) \in M_{n,n}(\mathbb{R})$ eine symmetrische Matrix, d.h. $A^\top = A$, und

$$\begin{aligned} f(\mathbf{x}) &:= \mathbf{x} \cdot A \cdot \mathbf{x}^\top \\ &= (x_1, \ldots, x_n) \cdot \Big(\sum_{j=1}^{n} a_{1j} x_j, \ldots, \sum_{j=1}^{n} a_{nj} x_j \Big)^\top = \sum_{i,j=1}^{n} a_{ij} x_i x_j \end{aligned}$$

die durch A bestimmte „quadratische Form". Um die Ableitung in einem Punkt $\mathbf{x}_0$ zu bestimmen, bleiben wir bei der vektoriellen Schreibweise. Es ist

$$\begin{aligned} f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) &= (\mathbf{x}_0 + \mathbf{h}) \cdot A \cdot (\mathbf{x}_0 + \mathbf{h})^\top - \mathbf{x}_0 \cdot A \cdot \mathbf{x}_0^\top \\ &= \mathbf{x}_0 \cdot A \cdot \mathbf{x}_0^\top + \mathbf{h} \cdot A \cdot \mathbf{x}_0^\top + \mathbf{x}_0 \cdot A \cdot \mathbf{h}^\top + \mathbf{h} \cdot A \cdot \mathbf{h}^\top \\ &\quad - \mathbf{x}_0 \cdot A \cdot \mathbf{x}_0^\top \\ &= 2\mathbf{x}_0 \cdot A \cdot \mathbf{h}^\top + \mathbf{h} \cdot A \cdot \mathbf{h}^\top = \mathbf{a} \bullet \mathbf{h} + \boldsymbol{\delta}(\mathbf{x}_0 + \mathbf{h}) \bullet \mathbf{h}, \end{aligned}$$

mit $\mathbf{a} := 2\mathbf{x}_0 \cdot A$ und $\boldsymbol{\delta}(\mathbf{x}) := (\mathbf{x} - \mathbf{x}_0) \cdot A$. Offensichtlich ist $\lim\limits_{\mathbf{x}\to\mathbf{x}_0} \boldsymbol{\delta}(\mathbf{x}) = \mathbf{0}$ und daher f total differenzierbar und $Df(\mathbf{x}_0)(\mathbf{h}) = 2\mathbf{x}_0 \cdot A \cdot \mathbf{h}^\top$.

D. Sei $f(x,y) := e^{x^2} \cdot \cos(y)$ und $\mathbf{x}_0 := (0, \pi/4)$.

Dann ist $f_x = 2xe^{x^2} \cdot \cos(y)$ und $f_y = -e^{x^2} \cdot \sin(y)$, also

$$Df(\mathbf{x}_0)(v_1, v_2) = f_x(\mathbf{x}_0)\, v_1 + f_y(\mathbf{x}_0)\, v_2 = -\frac{1}{2}\sqrt{2}\, v_2.$$

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

Wir zeigen zunächst, dass f im Nullpunkt stetig ist: Sei $(\mathbf{x}_\nu)$ eine Nullfolge. Dann können wir schreiben:

$$\mathbf{x}_\nu = (r_\nu \cos\varphi_\nu, r_\nu \sin\varphi_\nu), \text{ für } \nu \in \mathbb{N}.$$

Dabei konvergiert $r_\nu = \|\mathbf{x}_\nu\|$ gegen Null, und unabhängig von φ_ν ist

$$(\cos\varphi_\nu)^2 + (\sin\varphi_\nu)^2 = 1 \text{ und } 0 \leq |\cos\varphi_\nu|, |\sin\varphi_\nu| \leq 1.$$

Also konvergiert

$$|f(\mathbf{x}_\nu)| = |\frac{r_\nu^3 \cos\varphi_\nu (\sin\varphi_\nu)^2}{r_\nu^2}| \leq r_\nu$$

gegen Null.

Weiter ist $f(x,0) \equiv 0$ und $f(0,y) \equiv 0$. Also ist f im Nullpunkt auch partiell differenzierbar, und es gilt:

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

Es existieren sogar beliebige Richtungsableitungen:

Da $f(tx,ty) = t \cdot f(x,y)$ für alle t und beliebiges (x,y) gilt (man nennt eine solche Funktion ***homogen*** vom Grad 1), ist

$$D_{\mathbf{h}} f(\mathbf{0}) = \lim_{t\to 0} \frac{f(t\mathbf{h}) - f(\mathbf{0})}{t} = f(\mathbf{h}).$$

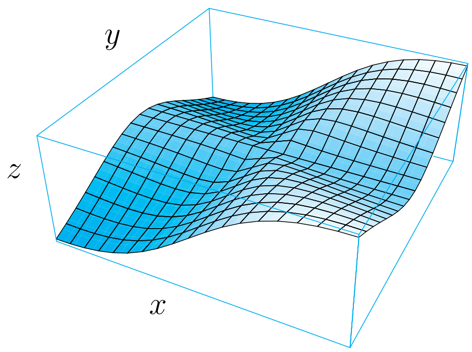
Man kann also im Nullpunkt in jeder beliebigen Richtung eine Tangente an den Graphen G_f legen.

Wäre f in $\mathbf{0}$ total differenzierbar, so müsste $Df(\mathbf{0})(\mathbf{h}) = 0$ für jedes $\mathbf{h}$ gelten. Für $\mathbf{h} := (r,r)$ ist aber

$$\frac{f(\mathbf{h}) - f(\mathbf{0}) - 0}{\|\mathbf{h}\|} = \frac{r^3}{2r^2 \cdot \sqrt{2}|r|} = \pm\frac{1}{2\sqrt{2}},$$

und dieser Ausdruck strebt für $\mathbf{h} \to \mathbf{0}$ nicht gegen Null.

Also ist f im Nullpunkt nicht total differenzierbar, und der Graph von f besitzt dort keine Tangentialebene. Wie soll man sich das vorstellen?



Da f homogen ist, gehört mit $(\mathbf{x}, z)$ auch jeder Punkt $(t\mathbf{x}, tz)$ zum Graphen von f, also die ganze Gerade durch $(\mathbf{x}, z)$ und den Nullpunkt. Diese Geraden sind dann natürlich auch Tangenten, und sie müssten daher auch in einer etwa existierenden Tangentialebene enthalten sein. Das ist nicht möglich, weil die Geraden gar nicht alle in einer Ebene liegen. Die Punkte $(1, 1, \frac{1}{2})$, $(1, -1, \frac{1}{2})$ und $(1, 0, 0)$ liegen z.B. auf G_f, sind aber linear unabhängig.

Tatsächlich hat G_f im Nullpunkt so etwas wie einen „Knick“, und dieser Mangel an Glattheit verhindert die totale Differenzierbarkeit.

3.11. Spezielle Kettenregel für differenzierbare Funktionen

Ist $B \subset \mathbb{R}^n$ offen, $\boldsymbol{\alpha} : I \to B$ in $t_0 \in I$ und $f : B \to \mathbb{R}$ in $\boldsymbol{\alpha}(t_0)$ differenzierbar, so ist auch $f \circ \boldsymbol{\alpha}$ in t_0 differenzierbar, und es gilt:

$$(f \circ \boldsymbol{\alpha})'(t_0) = \nabla f(\boldsymbol{\alpha}(t_0)) \bullet \boldsymbol{\alpha}'(t_0).$$

Beweis: Nach Voraussetzung gibt es eine in t_0 stetige Funktion $\mathbf{D} : I \to \mathbb{R}^n$ mit

$$\boldsymbol{\alpha}(t) = \boldsymbol{\alpha}(t_0) + \mathbf{D}(t) \cdot (t - t_0) \qquad \text{auf } I$$

und $\mathbf{D}(t_0) = \boldsymbol{\alpha}'(t_0)$ (siehe Analysis 1).

Außerdem gibt es eine in $\mathbf{x}_0 = \boldsymbol{\alpha}(t_0)$ stetige Funktion $\boldsymbol{\Delta} : B \to \mathbb{R}^n$, so dass $f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\Delta}(\mathbf{x})^\top$ und $\boldsymbol{\Delta}(\mathbf{x}_0) = \nabla f(\mathbf{x}_0)$ ist. Dann ist

$$\begin{aligned} f \circ \boldsymbol{\alpha}(t) - f \circ \boldsymbol{\alpha}(t_0) &= \big(\boldsymbol{\alpha}(t) - \boldsymbol{\alpha}(t_0)\big) \cdot \boldsymbol{\Delta}(\boldsymbol{\alpha}(t))^\top \\ &= \mathbf{D}(t) \cdot \boldsymbol{\Delta}(\boldsymbol{\alpha}(t))^\top \cdot (t - t_0) \\ &= \big(\boldsymbol{\Delta}(\boldsymbol{\alpha}(t)) \bullet \mathbf{D}(t)\big) \cdot (t - t_0). \end{aligned}$$

Also ist $f \circ \boldsymbol{\alpha}$ in t_0 differenzierbar und $(f \circ \boldsymbol{\alpha})'(t_0) = \nabla f(\boldsymbol{\alpha}(t_0)) \bullet \boldsymbol{\alpha}'(t_0)$. ∎

Wir können jetzt das Wesen des Gradienten etwas besser ergründen:

Sei $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}$ eine stetig partiell differenzierbare Funktion. Für $c \in \mathbb{R}$ sei $F_c := \{\mathbf{x} \in B \mid f(\mathbf{x}) = c\}$ die entsprechende ***Niveaumenge*** von f.

3.12. Satz

Sei $\mathbf{a} \in B$, $f(\mathbf{a}) = c$ und $\nabla f(\mathbf{a}) \neq \mathbf{0}$.

1. *$\nabla f(\mathbf{a})$ zeigt in die Richtung, in der f am schnellsten wächst.*
2. *Ist $\boldsymbol{\alpha} : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ ein differenzierbarer Weg mit $\boldsymbol{\alpha}(0) = \mathbf{a}$, der ganz in F_c verläuft, so steht $\nabla f(\mathbf{a})$ auf $\boldsymbol{\alpha}'(0)$ senkrecht.*

BEWEIS: 1) Ist $\mathbf{v} \neq \mathbf{0}$ und $\lambda > 0$, so ist

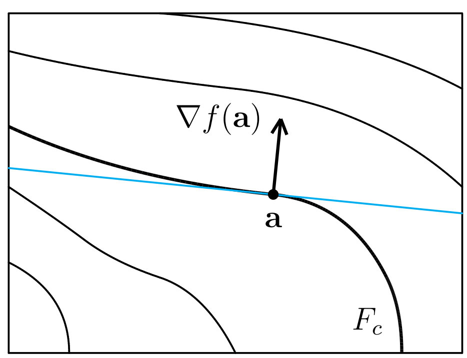
$$D_{\lambda\mathbf{v}} f(\mathbf{a}) = \lim_{t \to 0} \frac{f(\mathbf{a} + t(\lambda\mathbf{v})) - f(\mathbf{a})}{t} = \lambda \cdot \lim_{t\lambda \to 0} \frac{f(\mathbf{a} + (t\lambda)\mathbf{v}) - f(\mathbf{a})}{t\lambda} = \lambda \cdot D_{\mathbf{v}} f(\mathbf{a}).$$

Wir brauchen deshalb nur Vektoren $\mathbf{v}$ mit $\|\mathbf{v}\| = 1$ zu betrachten. Zu zeigen ist, dass $D_{\mathbf{v}} f(\mathbf{a})$ genau dann sein Maximum annimmt, wenn $\mathbf{v}$ in die Richtung des Gradienten zeigt. Tatsächlich ist

$$D_{\mathbf{v}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \bullet \mathbf{v} = \|\nabla f(\mathbf{a})\| \cdot \|\mathbf{v}\| \cdot \cos\theta,$$

wobei $\theta \in [0, \pi]$ der Winkel zwischen $\mathbf{v}$ und $\nabla f(\mathbf{a})$ ist. Dieser Ausdruck wird genau dann maximal, wenn $\theta = 0$ ist, also

$$\mathbf{v} = \frac{\nabla f(\mathbf{a})}{\|\nabla f(\mathbf{a})\|}.$$



2) Verläuft $\boldsymbol{\alpha}$ ganz in F_c, so ist $f \circ \boldsymbol{\alpha}(t) \equiv c$ und $0 = (f \circ \boldsymbol{\alpha})'(0) = \nabla f(\mathbf{a}) \bullet \boldsymbol{\alpha}'(0)$. ∎

Man sagt deshalb auch, **der Gradient steht auf der Niveaumenge senkrecht**.

Dass f in $\mathbf{x}_0$ differenzierbar ist, bedeutet, dass die affin-lineare Funktion

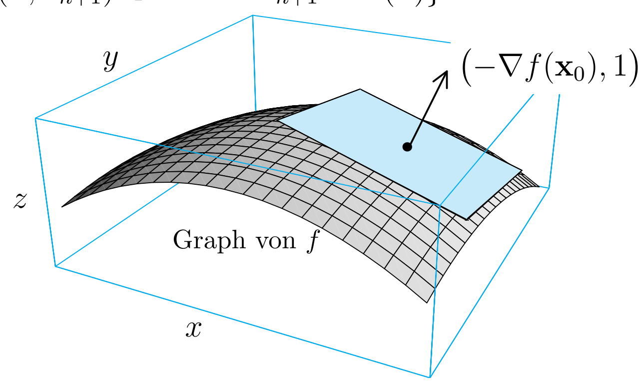
$$\mathbf{x} \mapsto \Lambda(\mathbf{x}) := f(\mathbf{x}_0) + Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \bullet (\mathbf{x} - \mathbf{x}_0)$$

die Funktion f bei $\mathbf{x}_0$ von zweiter Ordnung approximiert. Das ist das, was wir meinen, wenn wir sagen, dass die durch Λ beschriebene affine Hyperebene im $\mathbb{R}^{n+1}$ den Graphen von f im Punkte $(\mathbf{x}_0, f(\mathbf{x}_0))$ berührt.

Der Graph der Funktion f ist die Niveaumenge $\{(\mathbf{x}, y) \in B \times \mathbb{R} \,:\, F(\mathbf{x}, y) = 0\}$ der Funktion $F(\mathbf{x}, y) := f(\mathbf{x}) - y$. Die Tangentialebene an diese Menge im Punkt

$\mathbf{a} = (\mathbf{x}_0, f(\mathbf{x}_0))$ muss mit der Hyperebene $T \subset \mathbb{R}^{n+1}$ übereinstimmen, die auf $\nabla F(\mathbf{a}) = (\nabla f(\mathbf{x}_0), -1)$ senkrecht steht. Tatsächlich ist

$$\begin{aligned}
T &= \{(\mathbf{x}, x_{n+1}) \in \mathbb{R}^{n+1} \,:\, \big(\mathbf{x}-\mathbf{x}_0, x_{n+1} - f(\mathbf{x}_0)\big) \bullet \nabla F(\mathbf{x}_0, f(\mathbf{x}_0)) = 0\} \\
&= \{(\mathbf{x}, x_{n+1}) \in \mathbb{R}^{n+1} \,:\, \big(\mathbf{x}-\mathbf{x}_0, x_{n+1} - f(\mathbf{x}_0)\big) \bullet (\nabla f(\mathbf{x}_0), -1) = 0\} \\
&= \{(\mathbf{x}, x_{n+1}) \in \mathbb{R}^{n+1} \,:\, (\mathbf{x}-\mathbf{x}_0) \bullet \nabla f(\mathbf{x}_0) + f(\mathbf{x}_0) - x_{n+1} = 0\} \\
&= \{(\mathbf{x}, x_{n+1}) \in \mathbb{R}^{n+1} \,:\, x_{n+1} = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \bullet (\mathbf{x}-\mathbf{x}_0)\} \\
&= \{(\mathbf{x}, x_{n+1}) \in \mathbb{R}^{n+1} \,:\, x_{n+1} = \Lambda(\mathbf{x})\}.
\end{aligned}$$



Definition

Eine Teilmenge $M \subset \mathbb{R}^n$ heißt ***konvex***, falls mit je zwei Punkten $\mathbf{x}_0$ und $\mathbf{y}_0$ aus M auch ihre Verbindungsstrecke $S(\mathbf{x}_0, \mathbf{y}_0) := \{\mathbf{x} = \mathbf{x}_0 + t(\mathbf{y}_0 - \mathbf{x}_0) \,:\, 0 \le t \le 1\}$ zu M gehört.

3.13. Beispiel

Jede (offene oder abgeschlossene) Kugel ist konvex:

Beweis: Wir betrachten eine offene Kugel B vom Radius r um $\mathbf{0}$. Ist $\mathbf{x}_0 \in B$ und $\mathbf{y}_0 \in B$, so folgt:

$$\begin{aligned}
\|\mathbf{x}_0 + t(\mathbf{y}_0 - \mathbf{x}_0)\| &= \|(1-t)\mathbf{x}_0 + t\mathbf{y_0}\| \\
&\le (1-t)\cdot\|\mathbf{x}_0\| + t\cdot\|\mathbf{y}_0\| \\
&< (1-t)r + tr \;=\; r.
\end{aligned}$$

Für offene Kugeln mit beliebigem Mittelpunkt und für beliebige abgeschlossene Kugeln wird der Beweis sinngemäß geführt. ■

3.14. Der Mittelwertsatz

Sei $B \subset \mathbb{R}^n$ offen und konvex, $f : B \to \mathbb{R}$ eine differenzierbare Funktion. Zu je zwei Punkten $\mathbf{a}, \mathbf{b} \in B$ gibt es ein $t \in (0,1)$ mit

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{a} + t(\mathbf{b}-\mathbf{a})) \bullet (\mathbf{b} - \mathbf{a}).$$

BEWEIS: Wir setzen $\boldsymbol{\alpha}(t) := \mathbf{a} + t(\mathbf{b} - \mathbf{a})$ und $h(t) := f \circ \boldsymbol{\alpha}(t)$. Dies ist eine auf $[0, 1]$ differenzierbare Funktion. Nach dem Mittelwertsatz in einer Veränderlichen gibt es ein $t \in (0, 1)$, so dass $h(1) - h(0) = h'(t) \cdot (1 - 0) = h'(t)$ ist. Es ist aber $h(1) - h(0) = f(\mathbf{b}) - f(\mathbf{a})$ und

$$h'(t) = \nabla f(\boldsymbol{\alpha}(t)) \bullet \boldsymbol{\alpha}'(t) = \nabla f(\mathbf{a} + t(\mathbf{b} - \mathbf{a})) \bullet (\mathbf{b} - \mathbf{a}).$$

■

3.15. Folgerung

*Sei $G \subset \mathbb{R}^n$ ein **Gebiet**, $f : G \to \mathbb{R}$ differenzierbar und $\nabla f(\mathbf{x}) = \mathbf{0}$ für alle $\mathbf{x} \in G$. Dann ist f konstant.*

BEWEIS: Sei $\mathbf{x}_0 \in G$ und $c := f(\mathbf{x}_0)$. Dann ist die Menge

$$M := \{\mathbf{x} \in G : f(\mathbf{x}) = c\}$$

nicht leer. Ist $\mathbf{y} \in M$, so gibt es eine kleine Kugel $U = U_\varepsilon(\mathbf{y})$, die noch ganz in G liegt. Für $\mathbf{x} \in U$ ist $f(\mathbf{x}) - f(\mathbf{y}) = \nabla f(\mathbf{y} + t(\mathbf{y} - \mathbf{x})) \bullet (\mathbf{y} - \mathbf{x}) = 0$, also $f(\mathbf{x}) = f(\mathbf{y}) = c$. Damit gehört U zu M und M ist offen.

Weil f stetig ist, ist auch die Menge $G \setminus M = \{\mathbf{x} \in G : f(\mathbf{x}) \neq c\}$ offen. Also muss $M = G$ sein. ■

Ist $f : B \to \mathbb{R}$ in jedem Punkt von B differenzierbar, so kann man die abgeleitete Funktion $Df : B \to L(\mathbb{R}^n, \mathbb{R})$ betrachten. Dabei sieht man ein großes Problem der Differentialrechnung von mehreren Veränderlichen: Während f noch eine skalare Funktion ist, ist DF bereits vektorwertig.

3.16. Schrankensatz

Sei $B \subset \mathbb{R}^n$ offen und konvex, $f : B \to \mathbb{R}$ differenzierbar und $\|Df(\mathbf{x})\|_{\text{op}} \leq C$ für alle $\mathbf{x} \in B$. Dann ist

$$|f(\mathbf{b}) - f(\mathbf{a})| \leq C \cdot \|\mathbf{b} - \mathbf{a}\| \text{ für alle } \mathbf{a}, \mathbf{b} \in B.$$

BEWEIS: Es interessiert nur der Fall, dass $\mathbf{a} \neq \mathbf{b}$ ist. Wegen der Konvexität von B liegt die Verbindungsstrecke von $\mathbf{a}$ und $\mathbf{b}$ ganz in B. Dann gibt es auf der Verbindungsstrecke ein $\mathbf{c}$ mit $f(\mathbf{b}) - f(\mathbf{a}) = Df(\mathbf{c})(\mathbf{b} - \mathbf{a})$, und es ist

$$|f(\mathbf{b}) - f(\mathbf{a})| = |Df(\mathbf{c})(\mathbf{b} - \mathbf{a})| \leq \|Df(\mathbf{c})\|_{\text{op}} \cdot \|\mathbf{b} - \mathbf{a}\| \leq C \cdot \|\mathbf{b} - \mathbf{a}\|.$$ ■