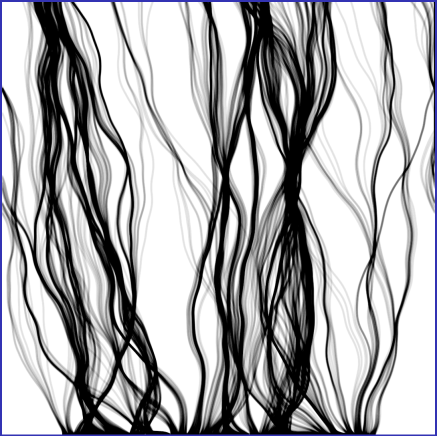
Angewandte Mathematik



Dr. rer. nat. Johannes Riesterer

Eigenschaften des Differentials

Für das Differential einer differenzierbaren Funktion $f : U \to \mathbb{R}$ gilt für alle $a \in U$:

- $df(a)(h) := df(a) \cdot h$ ist eine lineare Abbildung von $\mathbb{R}^n$ nach $\mathbb{R}$.
- $df(a) \cdot h = \partial_h f(a)$.
- $d(f \cdot g) = g(a)d(f) + f(a)dg$
- $d(f + g) = df + dg$

Beweis

- Multiplikation mit einer Matrix ist eine lineare Abbildung.
- Für die Basisvektoren ist per Definition $df(a) \cdot e_i = \partial_{e_i} f(a)$. Da jeder Vektor h eine Linearkombination der Basisvektoren ist und df linear ist, folgt die Behauptung.
- Folgt direkt aus der entsprechenden Eigenschaft reeller Funktionen.
- Folgt direkt aus der entsprechenden Eigenschaft reeller Funktionen.

Lokale Linearisierung

Ist $f : U \to \mathbb{R}$ differenzierbar, dann gilt für alle $a \in U$ und $h \in \mathbb{R}$

$$f(a+h) = f(a) + df(a) \cdot h + R(h)$$

mit $\lim_{h\to 0} \frac{R(h)}{||h||} = 0$.

Bedeutung

Eine differenzierbare Funktion kann auf hinreichend kleinen Umgebungen beliebig genau durch eine lineare Funktion approximiert werden.

Beweis

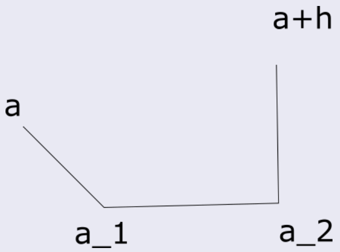


Figure: Kantenzug mit achesenparallelen Kanten

Beweis

$$a_0 := a$$
$$a_i := a_{i-1} + h_i e_i; \quad i = 1, \cdots, n$$

Beweis

- $f(a+h) - f(a) = \sum_{i=1}^{n} \big(f(a_i) - f(a_{i-1}) \big)$

Beweis

- $f(a+h) - f(a) = \sum_{i=1}^{n} \big(f(a_i) - f(a_{i-1})\big)$
- Mit $\varphi_i(t) := f(a_i + te_i)$ gilt $f(a_i) - f(a_{i-1}) = \varphi_i(h_i) - \varphi_i(0)$

Beweis

- $f(a+h) - f(a) = \sum_{i=1}^{n} \big(f(a_i) - f(a_{i-1})\big)$
- Mit $\varphi_i(t) := f(a_i + te_i)$ gilt $f(a_i) - f(a_{i-1}) = \varphi_i(h_i) - \varphi_i(0)$
- Mittelwertsatz einer Veränderlichen: Es gibt τ_i mit

$$\varphi_i(h_i) - \varphi_i(0) = h_i \varphi'(\tau_i)\ .$$

Beweis

- $f(a+h) - f(a) = \sum_{i=1}^{n} \big(f(a_i) - f(a_{i-1})\big)$
- Mit $\varphi_i(t) := f(a_i + te_i)$ gilt $f(a_i) - f(a_{i-1}) = \varphi_i(h_i) - \varphi_i(0)$
- Mittelwertsatz einer Veränderlichen: Es gibt τ_i mit

$$\varphi_i(h_i) - \varphi_i(0) = h_i \varphi'(\tau_i)\ .$$

- $$f(a+h) - f(a) - df(a) \cdot h = \sum_{i=1}^{n} \left(\frac{\partial f(\xi)}{\partial x_i} - \frac{\partial f(a)}{\partial x_i} \right) h_i$$

Da $\varphi_i'(t) = \frac{\partial f(a_{i-1} + te_i)}{\partial x_i}$ und mit $\xi_i := a_i + \tau_i e_i$

Beweis

$$|f(a+h)-f(a)-df(a)\cdot h| \leq ||h||_\infty \sum_{i=1}^{n}\left|\frac{\partial f(\xi)}{\partial x_i}-\frac{\partial f(a)}{\partial x_i}\right|.$$

Für $h \to 0$ gilt $\xi_i \to a$ und da die partiellen Ableitung stetig sind nach Voraussetzung und alle Normen äquivalent sind folgt

$$\lim_{h\to 0}\frac{f(a+h)-f(a)-df(a)\cdot h}{||h||}=0$$

Eindeutigkeit des Differentials

Umformuliert bedeutet der Satz über die lokale Linearisierung, dass für das Differential einer differenzierbare Funktion

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - df(a)h}{||h||} = 0 \tag{1}$$

gilt. Ist L eine weiter lineare Abbildung mit $\lim_{h \to 0} \frac{f(a+h)-f(a)-L(a)h}{||h||}$, so ist $L = df$. Das Differential ist somit eindeutig durch die Eigenschaft der lokalen Linearisierung bestimmt.

Eindeutigkeit des Differentials

Für $v \in \mathbb{R}^n$ mit $||v|| = 1$ gilt

$$\big(L(a) - df(a)\big)(v) = \lim_{t \to 0} \big(L(a) - df(a)\big)\Big(\frac{tv}{||tv||}\Big)$$
$$= \lim_{t \to 0} \frac{\big(L(a) - df(a)\big)(tv)}{||tv||} = 0$$

Da jeder Vektor als Linearkombination von Einheitsbasisvektoren dargestellt werden kann, folgt die Behauptung.

Wege

Ein Weg ist eine Abbildung

$$\gamma : [a, b] \to \mathbb{R}^n$$
$$\gamma(t) := \begin{pmatrix} \gamma_1(t) \\ \vdots \\ \gamma_n(t) \end{pmatrix}$$

mit reellen, stetigen Funktionen $y_i : [a, b] \to \mathbb{R}$ (Damit ist auch γ stetig).

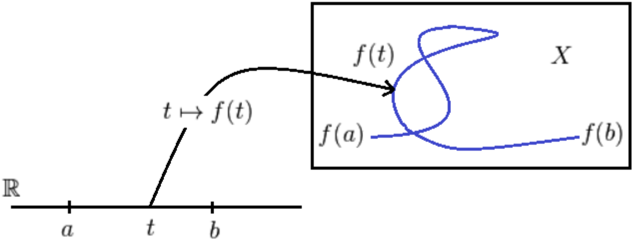


Figure: Quelle: Wikipedia:
https://de.wikipedia.org/wiki/Datei:EbeneKurve.png

Weg

Ein Weg ist eine Abbildung

$$\gamma : [a, b] \to \mathbb{R}^n$$
$$\gamma(t) := \begin{pmatrix} \gamma_1(t) \\ \vdots \\ \gamma_n(t) \end{pmatrix}$$

mit reellen, stetigen Funktionen $y_i : [a, b] \to \mathbb{R}$ (Damit ist auch γ stetig).

Differenzierterer Weg

Der Weg γ heißt differenzierbar, falls alle Ableitungen $\gamma_i'(t)$ existieren. In diesem Fall bezeichnen wir

$$\gamma'(t) := \begin{pmatrix} \gamma_1'(t) \\ \vdots \\ \gamma_n'(t) \end{pmatrix}$$

Baby Kettenregel

Sei $\gamma : I \to U$ ein differenzierterer Weg und $f : U \to \mathbb{R}$ eine differenziertere Funktion. Dann ist $f \circ \gamma : I \to \mathbb{R}$ differenzierbar und hat die Ableitung

$$\frac{d(f \circ \gamma)}{dt}(t) = df(\gamma(t))\gamma'(t) = \sum_{i=1}^{n} \frac{\partial f(\gamma(t))}{\partial x_i}\gamma_i'(t)$$

Verkettung

$(f \circ \gamma)(t) := f(\gamma(t))$

Beweis

- γ und f differenzierbar

$$\gamma(t+k) = \gamma(t) + k\gamma'(t) + r_1(k)|k|$$
$$f(\gamma(t)+h) = f(\gamma(t)) + df(\gamma(t)) \cdot h + r_2(h)||h||$$

$lim_{k\to 0} r_1(k) = 0,\ lim_{h\to 0} r_2(h) = 0.$

Beweis

- γ und f differenzierbar

$$\gamma(t+k) = \gamma(t) + k\gamma'(t) + r_1(k)|k|$$
$$f(\gamma(t)+h) = f(\gamma(t)) + df(\gamma(t)) \cdot h + r_2(h)||h||$$

  $lim_{k\to 0} r_1(k) = 0$, $lim_{h\to 0} r_2(h) = 0$.

- $h := \gamma(t+k) - \gamma(t)$

$$f(\gamma(t+k)) = f(\gamma(t)) + df(\gamma(t)) \cdot \gamma'(t)k + R(k)$$

  mit dem Restglied

$$R(k) := df(\gamma(t)) + r_1(k)|k| + r_2\big(\gamma(t+k) - \gamma(t)\big)||\gamma'(t)k + r_1(k)|k|||$$

Beweis

- $\lim_{k\to 0} R(k) = 0.$

Satz von Schwarz

Wenn Für eine Funktion $f : U \to \mathbb{R}$ die Ableitungen $\frac{\partial}{\partial x_i} f(a)$, $\frac{\partial}{\partial x_j} f(a)$ und $\frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} f(a)$ existieren und letztere stetig ist, dann existiert auch $\frac{\partial}{\partial x_j} \frac{\partial}{\partial x_i} f(a)$ und es gilt

$$\frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} f(a) = \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_i} f(a)$$

Bedeutung

Die Reihenfolge spielt beim wiederholten ableiten keine Rolle.

$\mathcal{C}^k$-Funktionen

Eine Funktion $f : U \to \mathbb{R}$ für die alle partiellen Ableitungen

$$\frac{\partial}{\partial x_{i_1}} \cdots \frac{\partial}{\partial x_{i_k}} f(a)$$

mit $i_1 + \cdots + i_k \leq k$ existieren und stetig sind heißt $\mathcal{C}^k$-Funktion oder k-mal stetig differenzierbar.

$\mathcal{C}^k$-Funktionen

Eine $\mathcal{C}^1$-Funktion ist also eine differenzierbare Funktion.

p-te Ableitung

Für eine Funktion $f : U \to \mathbb{R}$, $a \in U$ und Vektoren $v^1, \cdots, v^p \in \mathbb{R}^n$ heißt

$$d^p f(a)\big(v^1, \cdots, v^p\big) := \partial_{v^1} \dots \partial_{v^p} f(a)$$

die p-te Richtungsableitung von f. Sie ist wegen dem Satz von Schwarz wohldefiniert.

$$d^p f(a)\big(v^1, \cdots, v^p\big) = \sum_{i_1=1}^{n} \cdots \sum_{i_p=1}^{n} \frac{\partial}{\partial x_{i_1}} \cdots \frac{\partial}{\partial x_{i_p}} f(a) \cdot v^1_{i_1} \cdots v^p_{i_p} \ .$$

Für einen Vektor $z \in \mathbb{R}^n$ definieren wir $d^p f(a) z^p := d^p f(a) \underbrace{(z, \cdots, z)}_{p-mal}$.

Hessematrix

Für $p = 2$ und $u, v \in \mathbb{R}^n$ ist

$$d^2 f(a)(u, v) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} f(a) v_i u_i$$

und mit

$$f''(a) := \begin{pmatrix} \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_1} f(a) & \cdots & \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_n} f(a) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_n} \frac{\partial}{\partial x_1} f(a) & \cdots & \frac{\partial}{\partial x_n} \frac{\partial}{\partial x_n} f(a) \end{pmatrix}$$

ist $d^2 f(a)(u, v) = u^T \cdot f''(a) \cdot v$. Die Matrix $f''(a)$ wird auch Hesse-Matrix genannt. Nach dem Satz von Schwarz ist sie symmetrisch.

Taylorapproximation

Sei $f : U \to \mathbb{R}$ eine $\mathcal{C}^{p+1}$-Funktion und $x, a \in U$, so dass die Verbindung zwischen x und a in U liegt. Dann gilt

$$f(x) = f(a) + \sum_{k=1}^{p} \frac{1}{p!} d^k f(a)(x-a)^k + R_{p+1}(x;a)$$

mit dem Restglied $R_{p+1}(x;a) := \frac{1}{(p+1)!} d^{p+1} f(\xi)(x-a)^{p+1}$ für ein $\xi \in [a, x]$.

Beispiel

Wiki

Beweis

Sei $F(t) := f(a + th)$ mit $t \in [0,1]$. Wiederholte Anwendung der (Baby) Kettenregel mit $\gamma(t) := a + th$ ergibt

$$F'(t) = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(a + th) h_i$$

$$F''(t) = \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_i} f(a + th) h_i h_j$$

$$\vdots$$

$$F^p(t) = \sum_{i_1=1}^{n} \cdots \sum_{i_p=1}^{n} \frac{\partial}{\partial x_{i_1}} \cdots \frac{\partial}{\partial x_{i_p}} f(a + th) h_{i_1} \cdots h_{i_p} .$$

Beweis

Mit der Taylorapproximation für Funktionen einer Veränderlichen folgt für $h := (x - a)$

$$F(1) = F(0) + F'(0) + \frac{1}{2!}F''(0) + \cdots + \frac{1}{p!}F^p(0) + R_{p+1}$$

mit dem Restglied $R_{p+1} := \frac{1}{(p+1)!}F^p(\tau)$ mit $\tau \in [0, 1]$. Da nach Konstruktion $F(0) = f(a)$ und $F(1) = f(x)$ folgt insgesamt die Behauptung.

Taylorapproximation

Sei $T_p(x,a) = f(a) + \sum_{k=1}^{p} \frac{1}{p!} d^k f(a)(x-a)^k$ die Taylorraproximation einer $\mathcal{C}^p$-Funktion. Dann gilt:

$$\lim_{x \to a} \frac{f(x) - T_p(x;a)}{||x-a||^p} = 0 \ .$$

Bedeutung

Die Taylorapproximation vom Grad p konvergiert polynominell vom Grad p gegen 0.

Beweis

Da die partiellen Ableitungen stetig sind, gibt es für jedes $\epsilon > 0$ ein Radius $r > 0$, dass für alle $y \in K_r(a)$

$$\frac{1}{p!}(d^p f(y) - d^p f(a))h^p \leq \epsilon ||h||_\infty^p \; .$$

Mit der Taylorapproximation ist

$$\begin{aligned} f(x) =& T_{p-1}(x, a) + \frac{1}{p!} d^p f(\xi)(x-a)^p \\ =& T_p(x, a) + \frac{1}{p!}\left(d^p f(\xi) - d^p f(a)\right) h^p (x-a)^p \end{aligned}$$

Mit obiger Abschätzung folgt die Behauptung.

Extrema

Sei $f : X \subset \mathbb{R}^n \to \mathbb{R}$ eine relle Funktion. Ein Punkt $a \in X$ heißt lokales Maximum bzw. Minimum, falls eine Umgebung U von a existiert, so dass $f(x) \leq f(a)$ bzw. $f(x) \geq f(a)$ für alle $x \in U$ gilt. Liegt einer der beiden Fälle vor, so spricht man von einem lokalen Extremum. Gilt strikt $f(x) < f(a)$ bzw. $f(x) > f(a)$, so nennt man das Extremum isoliert. Ist $U = X$ so nennt man es auch globales Maximum bzw. Minimum.

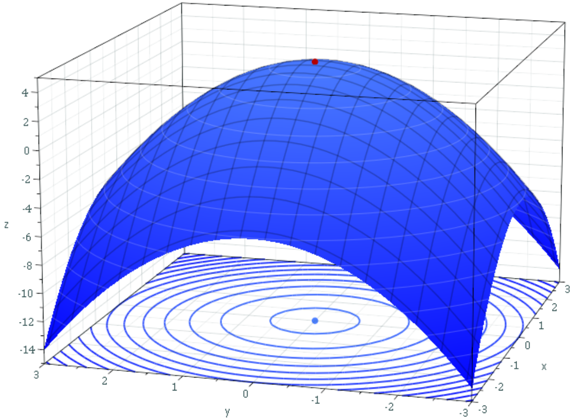


Figure: Quelle: Wikipedia:
https://en.wikipedia.org/wiki/File:MaximumParaboloid.png

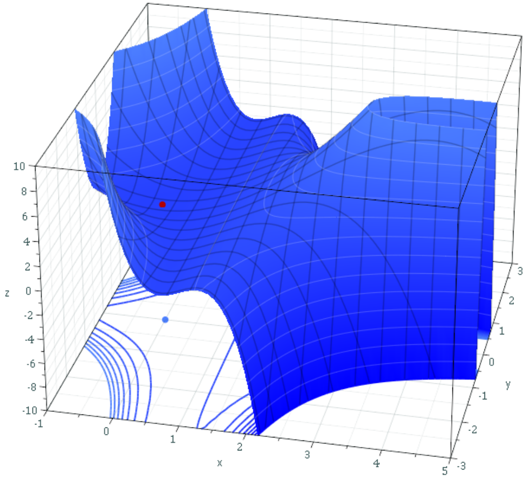


Figure: Quelle: Wikipedia:
https://en.wikipedia.org/wiki/File:MaximumCounterexample.png

Extrema

Ist $f : U \to \mathbb{R}$ differenzierbar und hat f in $a \in U$ ein lokales Extremum, so gilt

$$\frac{\partial}{\partial x_1} f(a) = \cdots = \frac{\partial}{\partial x_n} f(a) = 0 \,.$$

Dies ist gleichbedeutend mit $df(a) = 0$.

Kritischer Punkt

Ein Punkt a mit $df(a) = 0$ wird kritischer Punkt genannt.

Beweis

Setze $F_k(t) := f(a + te_k)$. Da f ein Extremum in a hat, hat F_k in einer hinreichend kleinen Umgebung um 0 ein Extremum. Da F_k eine Funktion einer Veränderlichen ist, gilt $F'(0) = 0$. Da $\frac{\partial}{\partial x_k} f(a) = F_k'(0)$ folgt die Behauptung.

Extrema

Ist $f : U \to \mathbb{R}$ eine $\mathcal{C}^2$-Funktion und ist $f'(a) = 0$ für ein $a \in U$. Dann gilt:

- $f''(a) > 0 \Rightarrow f$ hat in a ein isoliertes lokales Minimum.
- $f''(a) < 0 \Rightarrow f$ hat in a ein isoliertes lokales Maximum.
- $f''(a) \gtrless 0 \Rightarrow f$ hat in a einen Sattelpunkt.

Beweis

Sei $f'(a) = 0$ und $f''(a) > 0$. Mit der Taylorformel gilt für hinreichend kurze Vektoren $h \in \mathbb{R}^n$

$$f(a+h) = f(a) + \frac{1}{2} h^T f''(a) h + R(h)$$

mit $\lim_{h \to 0} \frac{R(h)}{||h||^2} = 0$. Für $||h|| \leq 1$ hat $h^T f''(a) h$ sein Maximum m auf dem Einheitskreis $\{h \in \mathbb{R}^n \mid ||h|| = 1\}$ da $f''(a) > 0$.

$$h^T f''(a) h = ||h|| \frac{1}{||h||} h^t f''(a) ||h|| \frac{1}{||h||} h \geq m ||h||^2 \, .$$

Beweis

Wir wählen ϵ so klein, dass $R(h) \leq \frac{m}{2}||h||^2$ gilt für $||h|| < \epsilon$ (was geht wegen Taylorformel). Damit erhalten wir

$$f(a+h) \geq f(a) + m||h||^2 \, .$$

und damit hat f ein lokales Minimum in a.
Der Fall $f''(a) < 0$ wird mit Betrachtung von $-f$ durch den vorigen Fall bewiesen.

Beweis

Es sei nun $f''(a) \gtrless 0$ und v mit $v^T f''(a)v > 0$ und w mit $w^T f''(a)w > 0$. Betrachten wir die Funktionen

$$F_v(t) := f(a + tv)$$
$$F_w(t) := f(a + tw)$$

dann ist

$$F_v'(t) = 0; \; F_v''(0) = v^T f''(a)v > 0$$
$$F_w'(t) = 0; \; F_w''(0) = w^T f''(a)w < 0$$

und somit hat F_v ein isoliertes lokales Maximum und F_w ein isoliertes lokales Minimum und damit f kein lokales Extremum in a.