

### Mehrdimensionale Differenzierbarkeit

Für  $z := (z_1, \dots, z_k) \in \mathbb{R}^k$  bezeichnen

$$pr_i : \mathbb{R}^k \rightarrow \mathbb{R}$$

$$pr_i(z) := \langle e_i, z \rangle_2 = z_i$$

die Projektionen auf die  $i$ -te Koordinate.

### Mehrdimensionale Differenzierbarkeit

Für zwei Funktionen  $g : U \subset \mathbb{R}^n \rightarrow V \subset \mathbb{R}^m$  und  
 $f : V \subset \mathbb{R}^m \rightarrow W \subset \mathbb{R}^l$  bezeichnet

$$f \circ g : U \rightarrow W$$

$$(f \circ g)(x) := f(g(x))$$

die Hintereinanderausführung (Verkettung) von  $f$  und  $g$ .

### Mehrdimensionale Differenzierbarkeit

Eine Funktion  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  ist genau dann differenzierbar, wenn alle Koordinatenfunktionen

$$F_i : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$$

$$F_i(x) := (pr_i \circ F)(x) := pr_i(F(x))$$

differenzierbar sind.

### Beweis

Betrachte  $dF = (dF_1, \dots, dF_m)$  zusammen mit dem Restglied  $R(h) = (R_1(h), \dots, R_m(h))$  definiert jeweils durch die rechte oder linke Seite.

### Kettenregel

Sind  $g : U \subset \mathbb{R}^n \rightarrow V \subset \mathbb{R}^m$  und  $f : V \subset \mathbb{R}^m \rightarrow W \subset \mathbb{R}^l$  differenzierbar, so ist  $g \circ f$  differenzierbar und es gilt

$$d(f \circ g)(a) = df(b) \cdot dg(a)$$

mit  $b = g(a)$ .

### Beweis

Nach Voraussetzung gilt

$$g(a+h) = g(a) + dg(a)h + \|h\|r_1(h); \quad \lim_{h \rightarrow 0} r_1(h) = 0$$

$$f(b+k) = f(b) + df(b)k + \|k\|r_2(k); \quad \lim_{k \rightarrow 0} r_2(k) = 0$$

### Beweis Kettenregel

Einsetzen ergibt

$$(f \circ g)(a+h) = (f \circ g)(a) + (df(b) \cdot dg(a))h + R(h)$$

mit  $R(h) := \|h\|df(b)r_1(h) + \|k\|r_2(k)$  und  
 $k = dg(a)h + \|h\|r_1(h)$ .

### Beweis Kettenregel

Müssen nur noch zeigen, dass  $\lim_{h \rightarrow 0} \frac{R(h)}{\|h\|} = 0$  gilt.

### Beweis Kettenregel

Da  $dg(a)$  eine lineare Abbildung ist, gibt es ein  $c \in \mathbb{R}$  mit

$$\|k\| \leq \|h\|(c + \|r_1(h)\|)$$

womit die Behauptung folgt.

### Satz von Schwarz

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

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

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

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

Für einen Vektor  $z \in \mathbb{R}^n$  definieren wir

$$d^p f(a)z^p := d^p f(a) \underbrace{(z, \dots, z)}_{p\text{-mal}}.$$

### Hessematrix

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

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

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

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

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 Kettenregel mit  $\gamma(t) := a + th$  und  $F(t) = f(\gamma(t))$  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  $T_p(x, a) = f(a) + \sum_{k=1}^p \frac{1}{k!} d^k f(a)(x - a)^k$  die Taylorapproximation einer  $\mathcal{C}^p$ -Funktion. Dann gilt:

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

### 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

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

Mit obiger Abschätzung folgt die Behauptung.

### Extrema

Sei  $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$  eine reelle 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.



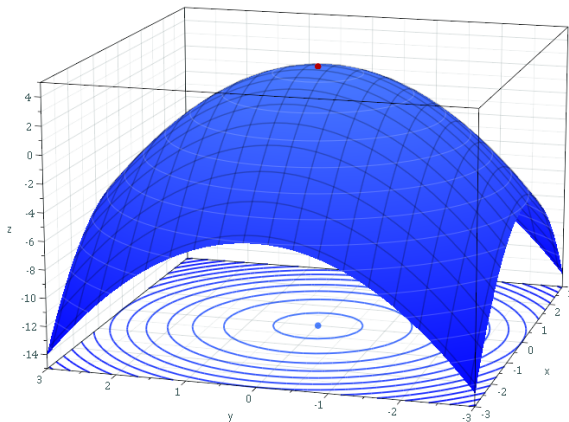


Figure: Quelle: Wikipedia:  
<https://en.wikipedia.org/wiki/File:MaximumParaboloid.png>

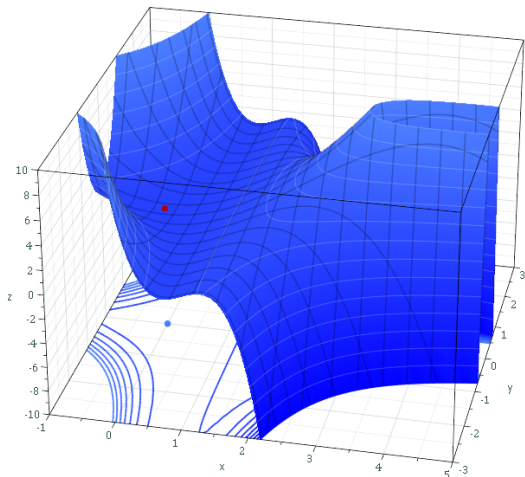


Figure: Quelle: Wikipedia:

<https://en.wikipedia.org/wiki/File:MaximumCounterexample.png>

### Extrema

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

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

Sind die partiellen Ableitungen stetig, ist dies 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  $F'_k(0) = 0$ . Da  $\frac{\partial}{\partial x_k} f(a) = F'_k(0)$  folgt die Behauptung.

### Extrema

Ist  $f : U \rightarrow \mathbb{R}$  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.
- $f''(a) \geq 0 \Rightarrow f$  hat in  $a$  einen Sattelpunkt.

### Extrema

$f''(a) > 0 \Leftrightarrow x^t f''(a) x > 0 \quad \forall x \in \mathbb{R}^n \Leftrightarrow$  alle Eigenwerte sind positiv  $\Leftrightarrow$  Alle Hauptminoren sind positiv .

### Extrema

$f''(a) < 0 \Leftrightarrow x^t f''(a) x < 0 \quad \forall x \in \mathbb{R}^n \Leftrightarrow$  alle Eigenwerte sind negativ  $\Leftrightarrow$  Alle Hauptminoren sind alternierend.

### 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  $\lim_{h \rightarrow 0} \frac{R(h)}{\|h\|} = 0$ . 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  $\epsilon$  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  $f''(a) \geq 0$  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

$$F'_v(t) = 0; \quad F''_v(0) = v^T f''(a)v > 0$$

$$F'_w(t) = 0; \quad F''_w(0) = w^T f''(a)w < 0$$

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

### Mittelwertsatz

Ist  $f : U \rightarrow \mathbb{R}$  eine differenzierbare Funktion und  $a, b \in U$  Punkte, deren Verbindungsstrecke in  $U$  verläuft. Dann gibt es einen Punkt  $\xi$  auf dieser Verbindungsstrecke mit

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

### Beweis Mittelwertsatz

Die Verbindungsstrecke ist gegeben durch  $\gamma(t) := a + t(b - a)$  mit  $t \in [0, 1]$ . Für  $F := f \circ \gamma : [0, 1] \rightarrow \mathbb{R}$  gilt  $f(b) - f(a) = F(1) - F(0)$ . Nach der Kettenregel ist  $F$  differenzierbar. Mit dem eindimensionalen Mittelwertsatz gibt es also ein  $\tau \in (0, 1)$  mit  $F(1) - F(0) = F'(\tau)$ . Mit der Kettenregel folgt  $F'(\tau) = df(\gamma(\tau))(b - a)$  und somit folgt mit  $\xi = \gamma(\tau)$  die Behauptung.