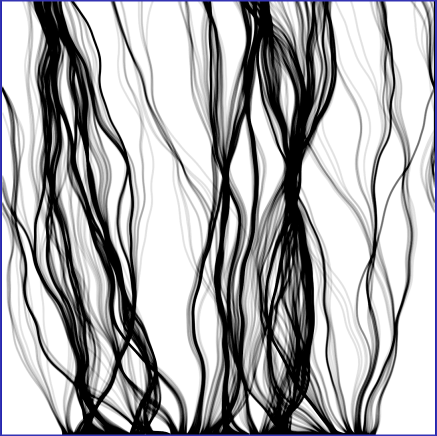
Angewandte Mathematik



Dr. rer. nat. Johannes Riesterer

Für fast alle (hinreichend große) Mathlib

Für einen Filter l bedeutet die Bedingung $\forall^f x\ p(x)$ dass die Menge der Elemente, für die $p(x)$ gilt, ein Element des Filters l ist, also $\{X \mid p(x)\} \in l$.

Big O Notation Mathlib

Für einen Filter l und Funktionen f, g definieren wir

$$f = \mathcal{O}[l]g \leftrightarrow \exists c > 0, \forall^{\mathrm{f}} x \in l, \|f(x)\| \leq c \cdot \|g(x)\|$$

Erklärung

Die Aussage besagt, dass für fast alle x in der Menge l, die Norm von $f(x)$ durch ein Vielfaches der Norm von $g(x)$ beschränkt ist. Das Vielfache wird durch die Konstante c dargestellt.

Klein-o Notation Mathlib

Für einen Filter l und Funktionen f, g definieren wir

$$f = o[l]g \leftrightarrow \forall c > 0, \forall^{\mathrm{f}} x \in l, \|f(x)\| \leq c \cdot \|g(x)\| \text{ für } x \geq N$$

Erklärung

Die Aussage besagt, dass für jede positive Konstante c und für fast alle x in der Menge l die Norm von $f(x)$ kleiner oder gleich $c \cdot \|g(x)\|$ ist. Dies beschreibt, dass $f(x)$ asymptotisch schneller gegen 0 geht als $g(x)$. In anderen Worten, der Ausdruck $\frac{\|f(x)\|}{\|g(x)\|}$ geht gegen 0 entlang l, wobei mögliche Probleme durch Division durch Null durch diese Definition vermieden werden.

Beweis: Klein-o impliziert Groß-O Mathlib

Wenn $f = o[l]g$, dann ist $f = O[l]g$.

Klassische Definition in einer Dimension:

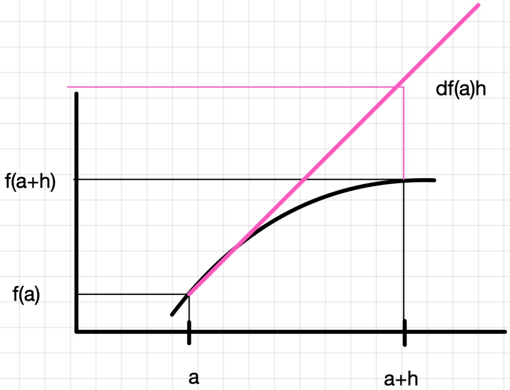
$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

Definition mit o-Kalkül:

$$f(a+h) = f(a) + f'(a)h + o(h)$$

Äquivalenz:

1. **Von $o(h)$ zur klassischen Definition:**
   $\frac{f(a+h)-f(a)}{h} = f'(a) + \frac{o(h)}{h}$
   Mit $\lim_{h\to 0} \frac{o(h)}{h} = 0$ folgt die klassische Definition.
2. **Von der klassischen Definition zur $o(h)$-Form:**
   Nehmen wir $f'(a) = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$, dann:
   $f(a+h) = f(a) + f'(a)h + r(h)$, wobei $r(h)$ der Restterm ist.
   Da $\lim_{h\to 0} \frac{r(h)}{h} = 0$, gilt $r(h) = o(h)$.



df(a)h
f(a+h)
f(a)
a
a+h

Lokale Linearisierung

Eine Funktion $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ heisst differenzierbar in $a \in U$ falls es eine lineare Funktion $df(a) : \mathbb{R}^n \to \mathbb{R}^m$ gibt mit

$$f(a+h) = f(a) + df(a)h + o(||h||) \tag{1}$$

$$\Leftrightarrow lim_{h \to 0} \frac{f(a+h) - f(a) - df(a)h}{||h||} = 0 \tag{2}$$

für alle $h \in \mathbb{R}^n$

Bedeutung

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

Beispiel

$A \in \mathbb{R}^{n \times m},\ b \in \mathbb{R}^n$

$$f(x) := A \cdot x + b \tag{3}$$

$$df(a) := A \tag{4}$$

Beweis

$$\lim_{h \to 0} \frac{A(x+h) - A \cdot x - A \cdot h}{||h||} = \tag{5}$$

$$\lim_{h \to 0} \frac{A \cdot x + A \cdot h - A \cdot x - A \cdot h}{||h||} = 0 \tag{6}$$

Eindeutigkeit

Die Ableitung df ist eindeutig bestimmt.

Beweis

Ist df' eine weiter Abbildung mit Eigenschaft (1), so gilt für jeden Basisvektor e_i

$$\lim_{t\to 0} \frac{f(a+te_i) - f(a) - df(a)te_i}{||te_i||} = 0 \tag{7}$$

$$\lim_{t\to 0} \frac{f(a+te_i) - f(a) - df'(a)te_i}{||te_i||} = 0 \tag{8}$$

$$\Rightarrow (df(a) - df'(a))(e_i) = lim_{t\to 0} \frac{(df'(a) - df(a))(te_i)}{||te_i||} = 0 \tag{9}$$

Mehrdimensionale Differenzierbarkeit

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

$$\begin{aligned} &F_i : U \subset \mathbb{R}^n \to \mathbb{R} \\ &F_i(x) := (pr_i \circ F)(x) := pr_i(F(x)) \end{aligned}$$

differenzierbar sind.

Beweis

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

Lineare Abbildungen

Definition

Eine Abbildung $T : V \to W$ zwischen zwei Vektorräumen V und W über einem Körper K heißt **linear**, wenn für alle $v_1, v_2 \in V$ und alle $\alpha, \beta \in K$ gilt:

$$T(\alpha v_1 + \beta v_2) = \alpha T(v_1) + \beta T(v_2).$$

Eigenschaften

- Lineare Abbildungen erhalten die Vektorraumstruktur: Sie respektieren die Addition und die Skalarmultiplikation.
- Jede lineare Abbildung ist durch ihr Verhalten auf einer Basis des Vektorraums eindeutig bestimmt.
- Die Ableitung einer linearen Abbildung ist die Abbildung selbst: Für eine lineare Abbildung T gilt $D(T) = T$.

Lineare Abbildungen

Beispiele

- Die Identitätsabbildung $\mathrm{id}_V : V \to V$, definiert durch $\mathrm{id}_V(v) = v$, ist linear.
- Projektionen und Rotationen in $\mathbb{R}^n$ sind lineare Abbildungen.
- Matrizen wirken als lineare Abbildungen auf Vektorräumen.

Lineare Abbildung in Lean Mathlib

In Lean4 (Mathlib) wird eine lineare Abbildung T zwischen zwei normierten Vektorräumen V und W über $\mathbb{R}$ als eine stetige lineare Abbildung (`continuous_linear_map`) definiert:

$$T : V \to L[\mathbb{R}]W$$

Die lineare Struktur wird durch zwei Eigenschaften beschrieben:

- `map_add` : $T(v_1 + v_2) = Tv_1 + Tv_2$
- `map_smul` : $T(c \cdot v) = c \cdot Tv$

Vergleich

Im endlichdimensionalen Fall entspricht die Definition in Lean der üblichen Definition einer linearen Abbildung. Im unendlichdimensionalen Fall wird zusätzlich die Stetigkeit gefordert, da diese nicht automatisch gegeben ist.

Definition

Eine Funktion f ist an x differenzierbar, wenn:

$$f(x') - f(x) - f'(x' - x) = o[L](x' - x)$$

Dies bedeutet, dass der Restterm $f(x') - f(x) - f'(x' - x)$ schneller gegen 0 geht als $x' - x$, wenn $x' \to x$ unter einem Filter L.

Beispiele für Filter L

- **Standardfall: Filter der Umgebung von** x
  Der Filter $L = \mathcal{N}(x)$ beschreibt, dass x' beliebig nahe an x heranrückt. Dieser Filter erfasst alle offenen Umgebungen von x.
  $$o[\mathcal{N}(x)](x' - x)$$
  bedeutet, dass der Restterm verschwindet, wenn x' gegen x läuft.

Beispiele für Filter L (Fortsetzung)

- **Filter auf einem Teilraum:**
  Wenn man Differenzierbarkeit nur auf einem Teilraum $S \subseteq E$ betrachtet, verwendet man den Filter $\mathcal{N}[S](x)$, der Umgebungen in S enthält. Damit kann man die Differenzierbarkeit von f auf S testen.
- **Filter für gerichtete Mengen:**
  Bei Funktionen auf gerichteten Mengen (z.B. in Optimierungsproblemen) verwendet man den Filter L, der beschreibt, wie x' sich entlang einer Richtung oder eines Pfades $\gamma(t) \to x$ nähert.

Zusammenfassung

Der Filter L gibt die Art und Weise an, wie x' gegen x strebt. Der häufigste Fall ist der Filter der offenen Umgebungen von x, aber auch Teilräume oder spezielle Pfade können durch Filter modelliert werden.

Kettenregel für differenzierbare Funktionen

Gegeben seien differenzierbare Funktionen $f : E \to F$ und $g : F \to G$ mit

$$f'(x) : E \to F \quad \text{und} \quad g'(f(x)) : F \to G,$$

wobei $f'(x)$ und $g'(f(x))$ die Ableitungen von f bzw. g an den jeweiligen Punkten darstellen.
Dann ist die Komposition $h = g \circ f$ differenzierbar und die Ableitung $h'(x)$ gegeben durch:

$$h'(x) = g'(f(x)) \circ f'(x).$$

Schritte des Beweises in klassischer Mathematik

1. Definition des Fehlerterms für f

Da f an x differenzierbar ist:

$$f(y) = f(x) + f'(x)(y - x) + o(\|y - x\|),$$

wobei der Fehlerterm $o(\|y - x\|)$ schneller gegen Null geht als $\|y - x\|$.

2. Definition des Fehlerterms für g

Da g an $f(x)$ differenzierbar ist:

$$g(f(y)) = g(f(x)) + g'(f(x))(f(y) - f(x)) + o(\|f(y) - f(x)\|).$$

Kombinieren der Fehlerterme

3. Einsetzen des Fehlers von f in den Fehler von g

Ersetzen von $f(y) - f(x)$ durch den Ausdruck aus Schritt 1 ergibt:

$$g(f(y)) = g(f(x)) + g'(f(x))\left(f'(x)(y-x) + o(\|y-x\|)\right) + o(\|f(y)-f(x)\|).$$

4. Vereinfachen und Abschätzen

Da die Terme $g'(f(x))(o(\|y-x\|))$ und $o(\|f(y)-f(x)\|)$ gegen Null gehen, ist

$$h(y) = h(x) + h'(x)(y-x) + \text{Fehlerterm},$$

was die Differenzierbarkeit von $h = g \circ f$ zeigt.

Schritt-für-Schritt-Übersetzung Mathlib

- `hg : HasFDerivAtFilter g g' (f x) L'` und `hf : HasFDerivAtFilter f f' x L` entsprechen den Differenzierbarkeitsannahmen von g und f.
- eq_1 und eq_2 repräsentieren die asymptotischen Abschätzungen, die den klassischen Fehlertermen entsprechen.
- `refine .of_isLittleO <| eq₂.triangle <| eq₁.congr_left` verwendet die Dreiecksungleichung und die $o(\cdot)$-Schranken, um die Differenzierbarkeit zu folgern.

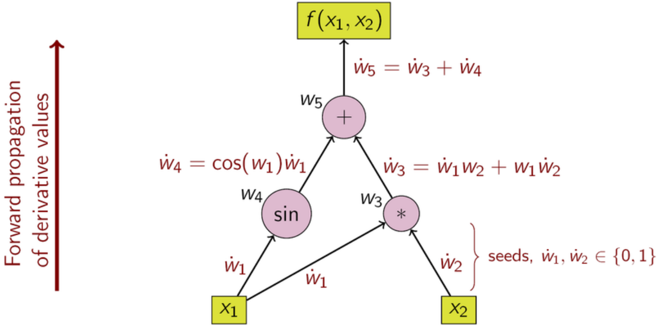


Figure: Quelle: Wikipedia

Automatisches Ableiten in Pytorch
Automatisches Ableiten in JAX

Ableitung Berechnen

Wie kann man die Ableitung einer Funktion berechnen?

Partielle Ableitung

Für eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ definert man die partielle Ableitung

$$D_i f(x) = \lim_{t \to 0} \frac{f(x + t \cdot e_i) - f(x)}{t}$$

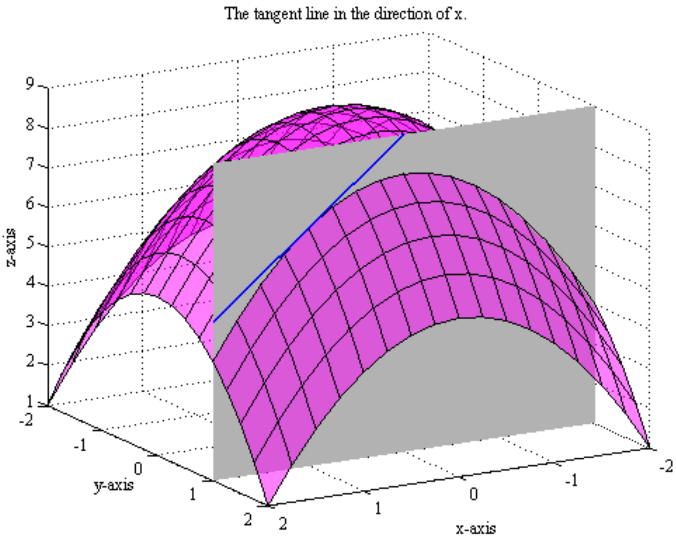
wobei e_i der i-te Einheitsvektor ist.
$\frac{\partial f(x)}{\partial x_i} := D_i f(x)$.

Richtungsableitung Mathlib

Allgemeiner definiert man für $f : \mathbb{R}^n \to \mathbb{R}$ und $v \in \mathbb{R}^n$

$$D_v f(x) = \lim_{t \to 0} \frac{f(x + t \cdot v) - f(x)}{t}$$

die Richtungsableitung (Linienableitung) von f an der Stelle x in Richtung v.



The tangent line in the direction of x.
z-axis
9
8
7
6
5
4
3
2
1
-2
-1
0
y-axis
1
2
2
1
0
-1
-2
x-axis

Gradient

Der Vektor

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

wird als Gradient bezeichnet.

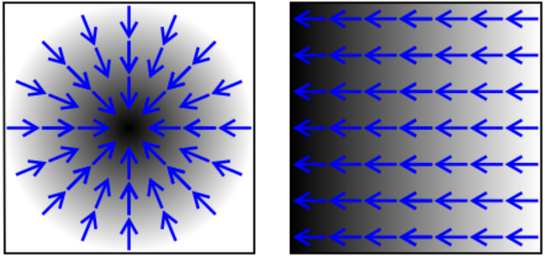


Figure: Quelle: Wikipedia:

Beispiel 1: Quadratische Funktion in 2D

$$f(x,y) = x^2 + y^2$$

Gradient:

$$\nabla f(x,y) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$$

Beispiel 2: Quadratische Funktion in 3D

$$f(x,y,z) = x^2 + y^2 + z^2$$

Gradient:

$$\nabla f(x,y,z) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}$$

Beispiel 3: Exponentialfunktion

$$f(x,y) = e^{x^2+y^2}$$

Gradient:

$$\nabla f(x,y) = \begin{pmatrix} \frac{\partial}{\partial x} e^{x^2+y^2} \\ \frac{\partial}{\partial y} e^{x^2+y^2} \end{pmatrix} = \begin{pmatrix} 2xe^{x^2+y^2} \\ 2ye^{x^2+y^2} \end{pmatrix}$$

Beispiel 4: Logarithmische Funktion

$$f(x,y) = \ln(x^2+y^2)$$

Gradient:

$$\nabla f(x,y) = \begin{pmatrix} \frac{\partial}{\partial x} \ln(x^2+y^2) \\ \frac{\partial}{\partial y} \ln(x^2+y^2) \end{pmatrix} = \begin{pmatrix} \frac{2x}{x^2+y^2} \\ \frac{2y}{x^2+y^2} \end{pmatrix}$$

Kettenregel

Gegeben seien Funktionen

$$f : \mathbb{R}^2 \to \mathbb{R} \quad \text{und} \quad g : \mathbb{R} \to \mathbb{R}^2,$$

die an den Punkten $g(x)$ bzw. x differenzierbar sind. Die Komposition $h(x) = f(g(x))$ ist dann ebenfalls differenzierbar mit:

$$h'(x) = \nabla f(g(x)) \cdot g'(x),$$

wobei $\nabla f(g(x))$ der Gradient von f an der Stelle $g(x)$ ist und $g'(x)$ die Jacobimatrix von g an x.

Funktionen und Komposition

Sei

$$f(x,y) = x^2 + y^2 \quad \text{und} \quad g(x) = (\cos(x), \sin(x)).$$

Dann ist die Komposition $h(x) = f(g(x)) = \cos^2(x) + \sin^2(x)$.

Schritte zur Anwendung der Kettenregel

1. Berechne den Gradienten von f:

$$\nabla f(x, y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}.$$

Setze $(x, y) = g(x) = (\cos(x), \sin(x))$:

$$\nabla f(g(x)) = \begin{pmatrix} 2\cos(x) \\ 2\sin(x) \end{pmatrix}.$$

2. Berechne die Ableitung von $g(x) = (\cos(x), \sin(x))$, also die Jacobimatrix $g'(x)$:

$$g'(x) = \begin{pmatrix} -\sin(x) \\ \cos(x) \end{pmatrix}.$$

3. Wende die Kettenregel an:

$$h'(x) = \nabla f(g(x)) \cdot g'(x) = \begin{pmatrix} 2\cos(x) & 2\sin(x) \end{pmatrix} \begin{pmatrix} -\sin(x) \\ \cos(x) \end{pmatrix}.$$

Zusammenfassung

Berechne das Skalarprodukt, um $h'(x)$ zu erhalten:

$$h'(x) = 2\cos(x) \cdot (-\sin(x)) + 2\sin(x) \cdot \cos(x) = 0.$$

Da $\cos^2(x) + \sin^2(x) = 1$, ergibt sich, dass $h(x) = 1$ konstant ist und daher $h'(x) = 0$.

Funktion und Ziel

Sei $h(x, y) = e^{x^2+y^2}$. Wir möchten den Gradienten von h berechnen, d.h.:

$$\nabla h(x, y) = \left(\frac{\partial h}{\partial x}, \frac{\partial h}{\partial y} \right).$$

Statt dies direkt zu tun, können wir die Kettenregel verwenden und die Berechnung vereinfachen, indem wir h als Komposition zerlegen.

Kompositionsansatz

Definiere:

$$g(x, y) = x^2 + y^2 \quad \text{und} \quad f(u) = e^u.$$

Dann ist $h(x, y) = f(g(x, y)) = e^{x^2+y^2}$.

Verwendung der Kettenregel

Nach der Kettenregel gilt:

$$\nabla h(x, y) = f'(g(x, y)) \cdot \nabla g(x, y).$$

Schritte zur Berechnung

1. Berechne $f'(u)$:
$$f'(u) = e^u.$$
Damit ist
$$f'(g(x,y)) = e^{x^2+y^2}.$$
2. Berechne den Gradienten von $g(x,y) = x^2 + y^2$:
$$\nabla g(x,y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}.$$
3. Setze die Ergebnisse ein:
$$\nabla h(x,y) = f'(g(x,y)) \cdot \nabla g(x,y) = e^{x^2+y^2} \begin{pmatrix} 2x \\ 2y \end{pmatrix}.$$

Gradient von $h(x, y)$

Der Gradient von $h(x, y) = e^{x^2+y^2}$ ist:

$$\nabla h(x, y) = \begin{pmatrix} 2xe^{x^2+y^2} \\ 2ye^{x^2+y^2} \end{pmatrix}.$$

Vergleich: Lineare Abbildungen in Mathematik und Lean (Mathlib)

Richtungsableitungv und Ableitung

$$D_v f(x) = df(x)v$$

und damit insbesondere $df(a) \cdot h = \langle \nabla f(a), h \rangle$.

Beweisg

Folgt aus der Eindeutigkeit der Ableitung.

Ziel: Linienableitung innerhalb einer Menge zeigen

Lemma: HasFDerivWithinAt.hasLineDerivWithinAt

Wenn f an x innerhalb der Menge s differenzierbar ist mit Ableitung L, dann hat f eine Richtungsableitung (Linienableitung) in Richtung v, die durch $L(v)$ gegeben ist:

$$\text{HasFDerivWithinAt}\ f\ L\ s\ x \Rightarrow \text{HasLineDerivWithinAt}\ \mathbb{K}\ f\ (L(v))\ s\ x\ v.$$

Beweisschritte: Definition der Hilfsfunktion F

1. Definition einer Hilfsfunktion F

Um die Richtungsableitung in Richtung v zu bestimmen, definieren wir die Funktion

$$F(t) := x + t \cdot v.$$

Damit repräsentiert F eine Bewegung entlang der Richtung v vom Punkt x aus.

2. Darstellung von x durch F

Wir erkennen, dass $x = F(0)$, also der Punkt x durch die Funktion F bei $t = 0$ beschrieben werden kann. Dies verwenden wir, um x im Beweis durch $F(0)$ zu ersetzen:

$$\text{rewrite: } x = F(0) \text{ in } \texttt{hf}.$$

Beweisschritte: Berechnung der Ableitung von F

3. Ableitung von F innerhalb der Menge s

Da $F(t) = x + t \cdot v$, können wir die Ableitung von F an $t = 0$ innerhalb der Menge $F^{-1}(s)$ berechnen. Diese Ableitung ist:

$$\text{HasDerivWithinAt } F \, (0 + 1 \cdot v) \, (F^{-1}(s)) \, 0.$$

Diese Ableitung resultiert aus der Kombination der konstanten Funktion x und der Identitätsfunktion, multipliziert mit v.

4. Vereinfachung der Ableitung

Die Berechnung vereinfacht sich zu:

$$\text{HasDerivWithinAt } F \, v \, (F^{-1}(s)) \, 0.$$

durch die Gleichungen $1 \cdot v = v$ und $0 + v = v$.

5. Kombination der Ableitungen

Mit der Kettenregel schließen wir, dass die Richtungsableitung von f in Richtung v durch $L(v)$ gegeben ist:

$$\mathrm{HasLineDerivWithinAt}\ \mathbb{K}\, f\, (L(v))\, s\, x\, v.$$

Dies zeigt, dass die Ableitung von f in Richtung v der Anwendung von L auf v entspricht.

Gradient

Sei $f : U \to \mathbb{R}$ differenzierbare Funktion, $a \in U$ und $v := \operatorname{argmax}_{||h||=1}\{\partial_h f(a)\}$. Dann gilt

$$||\nabla f(a)||v = \nabla f(a)\ .$$

Gradient

Der Gradient zeigt in die Richtung des steilsten Anstiegs.

Beweis

Für beliebiges h gilt

$$\partial_h f(a) = df(a)h = \langle \nabla f(a), h \rangle = ||\nabla f(a)|| \cdot ||h|| \cdot \cos(\varphi)$$

wobei φ den Innenwinkel zwischen $\nabla f(a)$ und h bezeichnet. Für $||h|| = 1$ wird somit $\partial_h f(a)$ maximal, wenn $\varphi = 0$ und somit$h = \frac{\nabla f(a)}{||\nabla f(a)||}$ ist.

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.

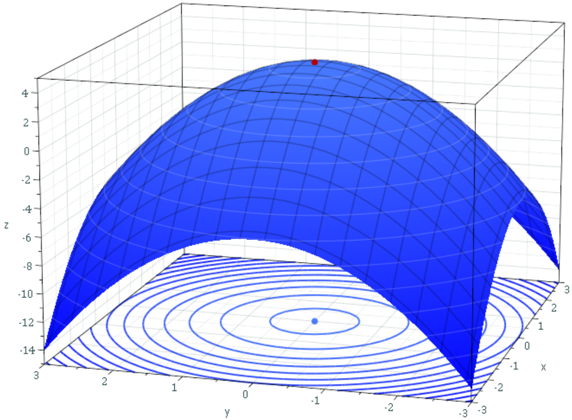


Figure: Quelle: Wikipedia:
https://en.wikipedia.org/wiki/File:MaximumParaboloid.png

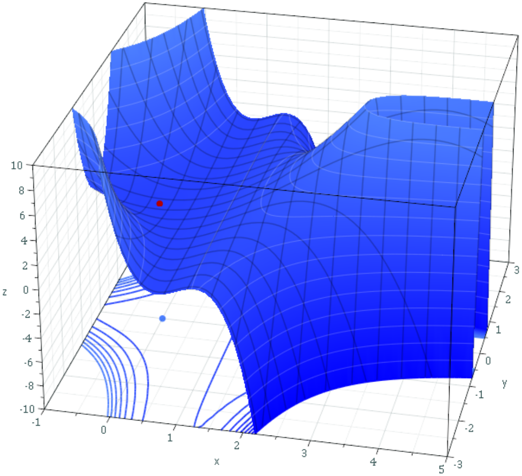


Figure: Quelle: Wikipedia:
https://en.wikipedia.org/wiki/File:MaximumCounterexample.png

Extrema Mathlib

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

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

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 $F'(0) = 0$. Da $\frac{\partial}{\partial x_k} f(a) = F_k'(0)$ folgt die Behauptung.