# Mathematische Grundlagen

BSc Psychologie | MSc Psychologie

MSc Klinische Psychologie und Psychotherapie

WiSe 2024/25

Prof. Dr. Dirk Ostwald

# (5) Differentialrechnung

Grundlagen der Differentialrechnung

Analytische Optimierung

Differentialrechnung multivariater reellwertiger Funktionen

Selbstkontrollfragen

**Grundlagen der Differentialrechnung**

Analytische Optimierung

Differentialrechnung multivariater reellwertiger Funktionen

Selbstkontrollfragen

## Definition (Ableitung)

Es sei $I \subseteq \mathbb{R}$ ein Intervall und

$$f : I \to \mathbb{R}, x \mapsto f(x) \tag{1}$$

eine univariate reellwertige Funktion. $f$ heißt in $a \in I$ *differenzierbar*, wenn der Grenzwert

$$f'(a) := \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \tag{2}$$

existiert. $f'(a)$ heißt dann die *Ableitung von $f$ an der Stelle $a$*. Ist $f$ differenzierbar für alle $x \in I$, so heißt $f$ *differenzierbar* und die Funktion

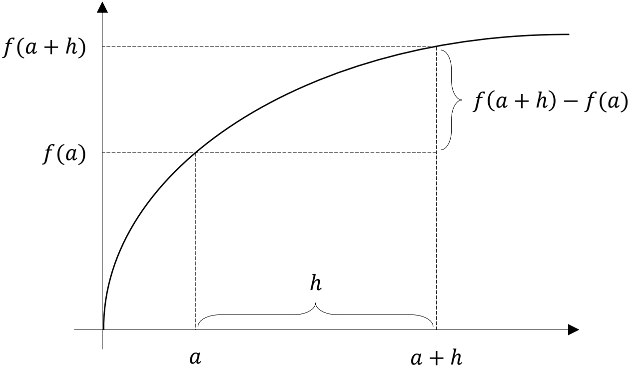
$$f' : I \to \mathbb{R}, x \mapsto f'(x) \tag{3}$$

heißt *Ableitung von $f$*.

Bemerkungen

- Für $h > 0$ heißt $\frac{f(a+h)-f(a)}{h}$ *Newtonscher Differenzquotient*.
- Der Differenzquotient misst die Änderung $f(a+h) - f(a)$ von $f$ pro Strecke $h$.
- Für $h \to 0$ misst der Differenzquotient die Änderungsrate von $f$ in $a$.
- $f'(a)$ ist eine Zahl, $f'$ ist eine Funktion.
- Wir werden keine Grenzwertbildung zur Berechnung von Ableitungen benötigen.

Newtonscher Differenzquotient



## Definition (Notation für Ableitungen univariater reellwertiger Funktionen)

Es sei $f$ eine univariate reellwertige Funktion. Äquivalente Schreibweisen für die Ableitung von $f$ und die Ableitung von $f$ an einer Stelle $x$ sind

(1) die *Lagrange-Notation* $f'$ und $f'(x)$,

(2) die *Newton-Notation* $\dot{f}$ und $\dot{f}(x)$,

(3) die *Leibniz-Notation* $\frac{df}{dx}$ und $\frac{df(x)}{dx}$ und

(4) die *Euler-Notation* $Df$ und $Df(x)$,

respektive.

Bemerkungen

- Für univariate reellwertige Funktionen benutzen wir $f'$ und $f'(x)$ als Bezeichner.
- In Berechnungen benutzen wir auch die "Operator-Schreibweise" $\frac{d}{dx}f(x)$.
- Wir verstehen $\frac{d}{dx}f(x)$ als den Auftrag, die Ableitung von $f$ zu berechnen.

## Theorem (Ableitungen elementarer Funktionen)

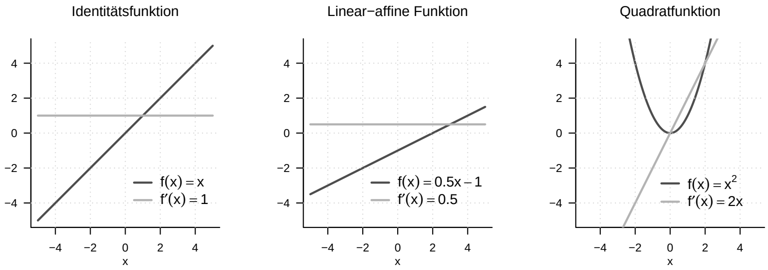
Für einige elementare Funktionen der Datenanalyse ergeben sich folgende Ableitungen

| Name | Definition | Ableitung |
|---|---|---|
| Polynomfunktionen | $f(x) := \sum_{i=0}^{n} a_i x^i$ | $f'(x) = \sum_{i=1}^{n} i a_i x^{i-1}$ |
| Konstante Funktion | $f(x) := a$ | $f'(x) = 0$ |
| Identitätsfunktion | $f(x) := x$ | $f'(x) = 1$ |
| Linear-affine Funktion | $f(x) := ax + b$ | $f'(x) = a$ |
| Quadratfunktion | $f(x) := x^2$ | $f'(x) = 2x$ |
| Exponentialfunktion | $f(x) := \exp(x)$ | $f'(x) = \exp(x)$ |
| Logarithmusfunktion | $f(x) := \ln(x)$ | $f'(x) = \frac{1}{x}$ |

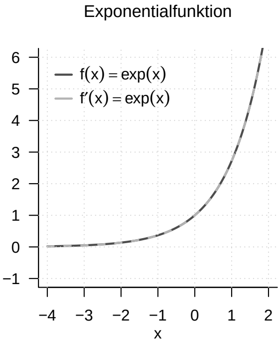
Bemerkung

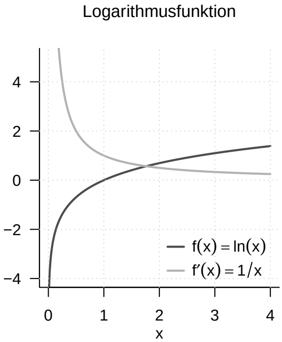
- Für Beweise verweisen wir auf die weiterführende Literatur.

## Ableitungen elementarer Funktionen



## Ableitungen elementarer Funktionen



## Definition (Höhere Ableitungen)

Es sei $f$ eine univariate reellwertige Funktion und

$$f^{(1)} := f' \tag{4}$$

sei die Ableitung von $f$. Die $k$-te Ableitung von $f$ ist rekursiv definiert durch

$$f^{(k)} := \left(f^{(k-1)}\right)' \text{ für } k > 1, \tag{5}$$

unter der Annahme, dass $f^{(k-1)}$ differenzierbar ist. Insbesondere ist die *zweite Ableitung von $f$* definiert durch die Ableitung von $f'$, also

$$f'' := (f')'. \tag{6}$$

Bemerkungen

- Wir schreiben auch $\frac{d^2}{dx^2} f(x)$ für den Auftrag, die zweite Ableitung von $f$ zu bestimmen.
- Die nullte Ableitung $f^{(0)}$ von $f$ ist $f$ selbst.
- Üblicherweise schreibt man für $k < 4$ $f', f'', f'''$ statt $f^{(1)}, f^{(2)}, f^{(3)}$.
- Im Allgemeinen benötigen wir nur $f'$ und $f''$.

## Beispiel

Es sei

$$f : \mathbb{R} \to \mathbb{R}, x \mapsto f(x) := x^2 \tag{7}$$

Dann gilt

$$f^{(1)}(x) = f'(x) = \frac{d}{dx}\left(x^2\right) = 2x \tag{8}$$

Weiterhin gelten

$$\begin{aligned}
f^{(2)}(x) &= \left(f^{(2-1)}\right)'(x) = \left(f^{(1)}\right)'(x) = \frac{d}{dx}(2x) = 2 \\
f^{(3)}(x) &= \left(f^{(3-1)}\right)'(x) = \left(f^{(2)}\right)'(x) = \frac{d}{dx}(2) = 0 \\
f^{(4)}(x) &= \left(f^{(4-1)}\right)'(x) = \left(f^{(3)}\right)'(x) = \frac{d}{dx}(0) = 0
\end{aligned} \tag{9}$$

Man berechnet also durchgängig lediglich erste Ableitungen.

## Theorem (Rechenregeln für Ableitungen)

Für $i = 1, ..., n$ seien $g_i$ reellwertige univariate differenzierbare Funktionen. Dann gelten folgende Rechenregeln für Ableitungen

(1) Summenregel

$$\text{Für } f(x) := \sum_{i=1}^{n} g_i(x) \text{ gilt } f'(x) = \sum_{i=1}^{n} g_i'(x). \tag{10}$$

(2) Produktregel

$$\text{Für } f(x) := g_1(x)g_2(x) \text{ gilt } f'(x) = g_1'(x)g_2(x) + g_1(x)g_2'(x). \tag{11}$$

(3) Quotientenregel

$$\text{Für } f(x) := \frac{g_1(x)}{g_2(x)} \text{ gilt } f'(x) = \frac{g_1'(x)g_2(x) - g_1(x)g_2'(x)}{g_2^2(x)}. \tag{12}$$

(4) Kettenregel

$$\text{Für } f(x) := g_1(g_2(x)) \text{ gilt } f'(x) = g_1'(g_2(x))g_2'(x). \tag{13}$$

Bemerkung

- Für Beweise verweisen wir auf die weiterführende Literatur.

Beispiel zur Summenregel

Es sei

$$f : \mathbb{R} \to \mathbb{R}, x \mapsto f(x) := 4x^3 + 3x^2 \tag{14}$$

Dann hat $f$ die Form

$$f(x) = \sum_{i=1}^{2} g_i(x) = g_1(x) + g_2(x) \text{ mit } g_1(x) := 4x^3 \text{ und } g_2(x) := 3x^2. \tag{15}$$

Es gelten außerdem

$$g_1'(x) = \frac{d}{dx}\left(4x^3\right) = 12x^2 \text{ und } g_2'(x) = \frac{d}{dx}\left(3x^2\right) = 6x. \tag{16}$$

Also gilt mit der Summenregel

$$f'(x) = \sum_{i=1}^{2} g_i'(x) = g_1'(x) + g_2'(x) = 12x^2 + 6x. \tag{17}$$

## Beispiel zur Kettenregel

Es sei

$$f : \mathbb{R} \to \mathbb{R}, x \mapsto f(x) := \exp\left(-x^2\right) \tag{18}$$

Dann hat $f$ die Form

$$f(x) = g_2(g_1(x)) \text{ mit } g_2(z) := \exp(z) \text{ und } g_1(x) := -x^2, \tag{19}$$

wobei man beachte, dass für das Funktionsargument $z$ in $g_2$ im Fall von $f$ gerade $z = -x^2$ eingesetzt wird.

Es gelten außerdem

$$g_1'(x) = \frac{d}{dx}\left(-x^2\right) = -2x \text{ und } g_2'(z) = \frac{d}{dz}\left(\exp(z)\right) = \exp(z). \tag{20}$$

Also gilt mit der Kettenregel

$$f'(x) = g_2'(g_1(x))g_1'(x) = \exp\left(-x^2\right)(-2x) = -2x\exp\left(-x^2\right). \tag{21}$$

Man kann sich die resultierende Ableitung der Kettenregel merken als

"Die Ableitung der äußeren Funktion an der Stelle der inneren Funktion mal die Ableitung der inneren Funktion."

Grundlagen der Differentialrechnung

**Analytische Optimierung**

Differentialrechnung multivariater reellwertiger Funktionen

Selbstkontrollfragen

## Definition (Extremstellen und Extremwerte)

Es sei $U \subseteq \mathbb{R}$ und $f : U \to \mathbb{R}$ eine univariate reellwertige Funktion. Dann hat $f$ an der Stelle $x_0 \in U$

- ein *lokales Minimum*, wenn es ein Intervall $I := ]a, b[$ gibt mit $x_0 \in ]a, b[$ und
$$f(x_0) \le f(x) \text{ für alle } x \in I \cap U, \tag{22}$$
- ein *globales Minimum*, wenn gilt, dass
$$f(x_0) \le f(x) \text{ für alle } x \in U, \tag{23}$$
- ein *lokales Maximum*, wenn es ein Intervall $I := ]a, b[$ gibt mit $x_0 \in ]a, b[$ und
$$f(x_0) \ge f(x) \text{ für alle } x \in I \cap U, \tag{24}$$
- *globales Maximum*, wenn gilt, dass
$$f(x_0) \ge f(x) \text{ für alle } x \in U. \tag{25}$$

Der Wert $x_0$ heißt entsprechend *lokale* oder *globale Minimalstelle* oder *Maximalstelle*, der Funktionswert $f(x_0)$ heißt entsprechend *lokales* oder *globales Minimum* oder *Maximum*. Generell heißt der Wert $x_0$ *Extremstelle* und der Funktionswert $f(x_0)$ *Extremwert*. Minimalstellen und Maximalstellen werden auch mit

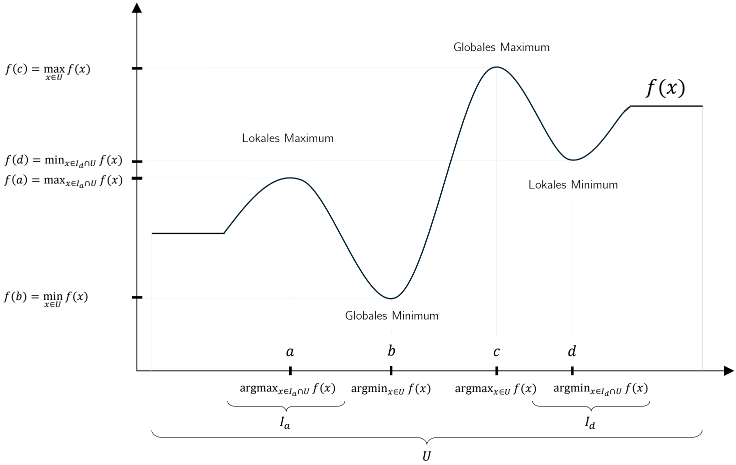
$$\arg\min_x f(x) \text{ bzw. } \arg\max_x f(x) \tag{26}$$

bezeichnet, Minima und Maxima werden auch mit

$$\min_x f(x) \text{ bzw. } \max_x f(x) \tag{27}$$

bezeichnet, wobei häufig noch die Mengenzugehörigkeit von $x$ angegeben wird.

## Extremstellen und Extremwerte



## Definition (Notwendige Bedingung für Extrema)

$f$ sei eine univariate reellwertige Funktion. Dann gilt

$$x_0 \text{ ist Extremstelle von } f \Rightarrow f'(x_0) = 0. \tag{28}$$

Bemerkungen

- Wenn $x_0$ eine Extremstelle von $f$ ist, dann ist die erste Ableitung von $f$ in $x_0$ null.
- Sei zum Beispiel $x_0$ eine lokale Maximalstelle von $f$
- Dann gilt, dass $f$ links von $x_0$ ansteigt und rechts von $x_0$ abfällt.
- In $x_0$ steigt $f$ weder an, noch fällt $f$ ab, also ist $f'(x_0) = 0$.

## Definition (Hinreichende Bedingungen für lokale Extrema)

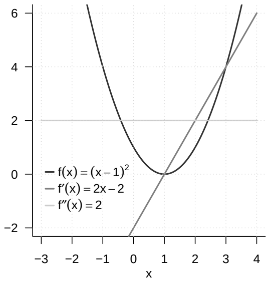
$f$ sei eine zweimal differenzierbare univariate reellwertige Funktion.

- Wenn für $x_0 \in U \subseteq \mathbb{R}$
$$f'(x_0) = 0 \text{ und } f''(x_0) > 0 \tag{29}$$
gilt, dann hat $f$ an der Stelle $x_0$ ein Minimum.
- Wenn für $x_0 \in U \subseteq \mathbb{R}$
$$f'(x_0) = 0 \text{ und } f''(x_0) < 0 \tag{30}$$
gilt, dann hat $f$ an der Stelle $x_0$ ein Maximum.

Bemerkung

- Eine Intuition vermittelt nachfolgende Abbildung.

# Analytische Optimierung




Hier ist offenbar $x_0 = 1$ eine lokale Minimalstelle von $f(x) = (x-1)^2$. Man erkennt:

- Links von $x_0$ fällt $f$ ab, rechts von $x_0$ steigt $f$ an.
- In $x_0$ steigt $f$ weder an, noch fällt $f$ ab, also ist $f'(x_0) = 0$.
- Links und rechts von $x_0$ und in $x_0$ ist die Änderung $f''$ von $f'$ positiv.
- Links von $x_0$ schwächt sich die Negativität von $f'$ zu $0$ ab.
- Rechts von $x_0$ verstärkt sich die Positivität von $f'$.

## Theorem (Standardverfahren der analytischen Optimierung)

$f$ sei eine univariate reellwertige Funktion. Lokale Extremstellen von $f$ können mit folgendem *Standardverfahren der analytischen Optimierung* identifiziert werden:

(1) Berechnen der ersten und zweiten Ableitung von $f$.

(2) Bestimmen der Nullstellen $x^*$ von $f'$ durch Auflösen von $f'(x^*) = 0$ nach $x^*$. Die Nullstellen von $f'$ sind dann Kandidaten für Extremstellen von $f$.

(3) Evaluation von $f''(x^*)$: Wenn $f''(x^*) > 0$, dann ist $x^*$ lokale Minimumstelle von $f$; wenn $f''(x^*) < 0$, dann ist $x^*$ lokale Maximumstelle von $f$; wenn $f''(x^*) = 0$, dann ist $x^*$ keine Extremstelle von $f$.

# Analytische Optimierung

## Beispiel

Wir betrachten die Funktion

$$f : \mathbb{R} \to \mathbb{R}, x \mapsto f(x) := (x-1)^2. \tag{31}$$

aus obiger Abbildung. Die erste Ableitung von $f$ ergibt sich mit der Kettenregel zu

$$f'(x) = \frac{d}{dx}\left((x-1)^2\right) = 2(x-1) \cdot \frac{d}{dx}(x-1) = 2x-2. \tag{32}$$

Die zweite Ableitung von $f$ ergibt sich zu

$$f''(x) = \frac{d}{dx}f'(x) = \frac{d}{dx}(2x-2) = 2 > 0 \text{ für alle } x \in \mathbb{R}. \tag{33}$$

Auflösen von $f'(x^*) = 0$ nach $x^*$ ergibt

$$f'(x^*) = 0 \Leftrightarrow 2x^* - 2 = 0 \Leftrightarrow 2x^* = 2 \Leftrightarrow x^* = 1. \tag{34}$$

$x^* = 1$ ist folglich eine Minimalstelle von $f$ mit zugehörigen Minimalwert $f(1) = 0$.

Grundlagen der Differentialrechnung

Analytische Optimierung

**Differentialrechnung multivariater reellwertiger Funktionen**

Selbstkontrollfragen

## Definition (Multivariate reellwertige Funktion)

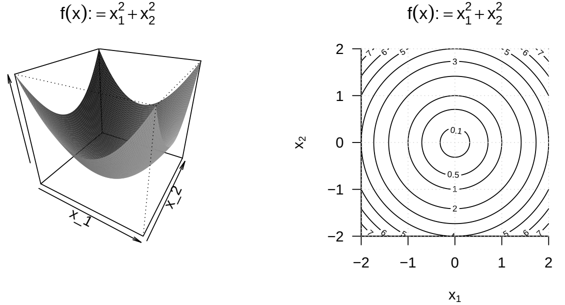
Eine Funktion der Form

$$f : \mathbb{R}^n \to \mathbb{R}, x \mapsto f(x) = f(x_1, ..., x_n) \tag{35}$$

heißt *multivariate reellwertige* Funktion.

Beispiel für $n := 2$

$$f : \mathbb{R}^2 \to \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2 \tag{36}$$



## Definition (Partielle Ableitung)

Es sei $D \subseteq \mathbb{R}^n$ eine Menge und

$$f : D \to \mathbb{R}, x \mapsto f(x) \tag{37}$$

eine multivariate reellwertige Funktion. $f$ heißt in $a \in D$ *nach* $x_i$ *partiell differenzierbar*, wenn der Grenzwert

$$\frac{\partial}{\partial x_i} f(x) := \lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h} \tag{38}$$

existiert. $\frac{\partial}{\partial x_i} f(a)$ heißt dann die *partielle Ableitung von* $f$ *nach* $x_i$ *an der Stelle* $a$. Wenn $f$ für alle $x \in D$, nach $x_i$ partiell differenzierbar ist, dann heißt $f$ *nach* $x_i$ *partiell differenzierbar* und die Funktion

$$\frac{\partial}{\partial x_i} f : D \to \mathbb{R}, x \mapsto \frac{\partial}{\partial x_i} f(x) \tag{39}$$

heißt *partielle Ableitung von* $f$ *nach* $x_i$.

$f$ heißt *partiell differenzierbar in* $x \in D$, wenn $f$ für alle $i = 1, ..., n$ in $x \in D$ nach $x_i$ partiell differenzierbar ist, und $f$ heißt partiell differenzierbar, wenn $f$ für alle $i = 1, ..., n$ in allen $x \in D$ nach $x_i$ partiell differenzierbar ist.

Bemerkungen

- $e_i \in \mathbb{R}^n$ bezeichnet den *i-ten kanonischen Einheitsvektor*.
- Es gilt also $e_{i_j} = 1$ für $i = j$ und $e_{i_j} = 0$ für $i \neq j$ mit $j = 1, ..., n$.
- $\frac{f(x+he_i)-f(x)}{h}$ misst die Änderung $f(x + he_i) - f(x)$ von $f$ pro Strecke $h$ in Richtung $e_i$.
- Für $h \to 0$ misst der Differenzquotient die Änderungs<u>rate</u> von $f$ in $x$ in Richtung $e_i$.
- $\frac{\partial}{\partial x_i} f(x)$ ist eine Zahl, $\frac{\partial}{\partial x_i} f$ ist eine Funktion.
- Praktisch berechnet man $\frac{\partial}{\partial x_i} f$ als die (einfache) Ableitung

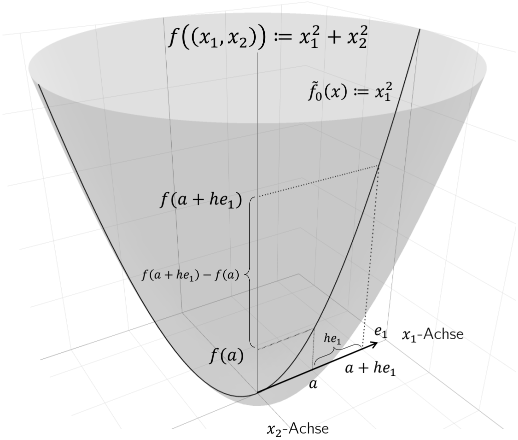
$$\frac{d}{dx_i} \tilde{f}_{x_1, \ldots x_{i-1}, x_{i+1}, \ldots, x_n}(x_i) \tag{40}$$

der univariaten reellwertigen Funktion

$$\tilde{f} : \mathbb{R} \to \mathbb{R}, x_i \mapsto \tilde{f}_{x_1, \ldots x_{i-1}, x_{i+1}, \ldots, x_n}(x_i) := f(x_1, ..., x_i, ..., x_n). \tag{41}$$

- Man betrachtet alle $x_j$ mit $j \neq i$ also als Konstanten.
- Wir werden also keine Grenzwertbildung zur Berechnung von partiellen Ableitungen benötigen.

Partieller Newtonscher Differenzquotient



## Beispiel (1)

Wir betrachten die Funktion

$$f : \mathbb{R}^2 \to \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2. \tag{42}$$

Weil die Definitionsmenge dieser Funktion zweidimensional ist, kann man zwei partielle Ableitungen berechnen

$$\frac{\partial}{\partial x_1} f : \mathbb{R}^2 \to \mathbb{R}, x \mapsto \frac{\partial}{\partial x_1} f(x) \text{ und } \frac{\partial}{\partial x_2} f : \mathbb{R}^2 \to \mathbb{R}, x \mapsto \frac{\partial}{\partial x_2} f(x). \tag{43}$$

Um die erste dieser partiellen Ableitungen zu berechnen, betrachtet man die Funktion

$$f_{x_2} : \mathbb{R} \to \mathbb{R}, x_1 \mapsto f_{x_2}(x_1) := x_1^2 + x_2^2, \tag{44}$$

wobei $x_2$ hier die Rolle einer Konstanten einnimmt. Um explizit zu machen, dass $x_2$ kein Argument der Funktion ist, die Funktion aber weiterhin von $x_2$ abhängt haben wir die Subskriptnotation $f_{x_2}(x_1)$ verwendet. Um nun die partielle Ableitung zu berechnen, berechnen wir die (einfache) Ableitung von $f_{x_2}$,

$$f'_{x_2}(x) = 2x_1. \tag{45}$$

Es ergibt sich also

$$\frac{\partial}{\partial x_1} f : \mathbb{R}^2 \to \mathbb{R}, x \mapsto \frac{\partial}{\partial x_1} f(x) = \frac{\partial}{\partial x_1} (x_1^2 + x_2^2) = f'_{x_2}(x) = 2x_1. \tag{46}$$

Analog gilt mit der entsprechenden Formulierung von $f_{x_1}$, dass

$$\frac{\partial}{\partial x_2} f : \mathbb{R}^2 \to \mathbb{R}, x \mapsto \frac{\partial}{\partial x_2} f(x) = \frac{\partial}{\partial x_2} (x_1^2 + x_2^2) = f'_{x_1}(x) = 2x_2. \tag{47}$$

## Definition (Zweite partielle Ableitungen)

$f : \mathbb{R}^n \to \mathbb{R}$ sei eine multivariate reellwertige Funktion und $\frac{\partial}{\partial x_i} f$ sei die partielle Ableitung von $f$ nach $x_i$. Dann ist die zweite partielle Ableitung von $f$ nach $x_i$ und $x_j$ definiert als

$$\frac{\partial^2}{\partial x_j \partial x_i} f(x) := \frac{\partial}{\partial x_j} \left( \frac{\partial}{\partial x_i} f \right). \tag{48}$$

Bemerkungen

- Wie die zweite Ableitung ist auch die zweite partielle Ableitung rekursiv definiert.
- Zu jeder partiellen Ableitung $\frac{\partial}{\partial x_i} f$ gibt es $n$ zweite partiellen Ableitungen $\frac{\partial^2}{\partial x_j \partial x_i} f, j = 1, ..., n$.

## Theorem (Satz von Schwarz)

$f : \mathbb{R}^n \to \mathbb{R}$ sei eine partiell differenzierbare multivariate reellwertige Funktion. Dann gilt

$$\frac{\partial^2}{\partial x_j \partial x_i} f(x) = \frac{\partial^2}{\partial x_i \partial x_j} f(x) \text{ für alle } 1 \leq i, j \leq n. \tag{49}$$

Bemerkungen

- Für einen Beweis verweisen wir auf die weiterführende Literatur.
- Das Theorem besagt, dass die Reihenfolge des partiellen Ableitens irrelevant ist
- Das Theorem erleichtert die Berechnung von zweiten partiellen Ableitungen.
- Das Theorem hilft, Fehler bei der Berechnung zweiter partieller Ableitungen aufzudecken.

## Beispiel (1) (fortgeführt)

Wir wollen die partiellen Ableitungen zweiter Ordnung der Funktion

$$f : \mathbb{R}^2 \to \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2. \tag{50}$$

berechnen. Mit den Ergebnissen für die partiellen Ableitungen erster Ordnung dieser Funktion ergibt sich

$$\begin{aligned}
\frac{\partial^2}{\partial x_1 \partial x_1} f(x) &= \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_1} f(x)\right) = \frac{\partial}{\partial x_1}(2x_1) = 2 \\
\frac{\partial^2}{\partial x_1 \partial x_2} f(x) &= \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_2} f(x)\right) = \frac{\partial}{\partial x_1}(2x_2) = 0 \\
\frac{\partial^2}{\partial x_2 \partial x_1} f(x) &= \frac{\partial}{\partial x_2}\left(\frac{\partial}{\partial x_1} f(x)\right) = \frac{\partial}{\partial x_2}(2x_1) = 0 \\
\frac{\partial^2}{\partial x_2 \partial x_2} f(x) &= \frac{\partial}{\partial x_2}\left(\frac{\partial}{\partial x_2} f(x)\right) = \frac{\partial}{\partial x_2}(2x_2) = 2
\end{aligned} \tag{51}$$

Offenbar gilt

$$\frac{\partial^2}{\partial x_1 \partial x_2} f(x) = \frac{\partial^2}{\partial x_2 \partial x_1} f(x). \tag{52}$$

Beispiel (2)

Wir wollen die partiellen Ableitungen erster und zweiter Ordnung der Funktion

$$f : \mathbb{R}^3_{>0} \to \mathbb{R}, x \mapsto f(x) := x_1^2 + x_1 x_2 + x_2\sqrt{x_3}. \tag{53}$$

berechnen.

Mit den Rechenregeln für Ableitungen ergibt sich für die partiellen Ableitungen erster Ordnung

$$\begin{aligned}
\frac{\partial}{\partial x_1} f(x) &= \frac{\partial}{\partial x_1}\left(x_1^2 + x_1 x_2 + x_2\sqrt{x_3}\right) = 2x_1 + x_2, \\
\frac{\partial}{\partial x_2} f(x) &= \frac{\partial}{\partial x_2}\left(x_1^2 + x_1 x_2 + x_2\sqrt{x_3}\right) = x_1 + \sqrt{x_3}, \\
\frac{\partial}{\partial x_3} f(x) &= \frac{\partial}{\partial x_3}\left(x_1^2 + x_1 x_2 + x_2\sqrt{x_3}\right) = \frac{x_2}{2\sqrt{x_3}}.
\end{aligned} \tag{54}$$

Beispiel (2) (fortgeführt)

Für die zweiten partiellen Ableitungen hinsichtlich $x_1$ ergibt sich

$$\begin{aligned}
\frac{\partial^2}{\partial x_1 \partial x_1} f(x) &= \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_1} f(x)\right) = \frac{\partial}{\partial x_1}(2x_1 + x_2) = 2, \\
\frac{\partial^2}{\partial x_2 \partial x_1} f(x) &= \frac{\partial}{\partial x_2}\left(\frac{\partial}{\partial x_1} f(x)\right) = \frac{\partial}{\partial x_2}(2x_1 + x_2) = 1, \\
\frac{\partial^2}{\partial x_3 \partial x_1} f(x) &= \frac{\partial}{\partial x_3}\left(\frac{\partial}{\partial x_1} f(x)\right) = \frac{\partial}{\partial x_3}(2x_1 + x_2) = 0.
\end{aligned} \tag{55}$$

Für die zweiten partiellen Ableitungen hinsichtlich $x_2$ ergibt sich

$$\begin{aligned}
\frac{\partial^2}{\partial x_1 \partial x_2} f(x) &= \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_2} f(x)\right) = \frac{\partial}{\partial x_1}(x_1 + \sqrt{x_3}) = 1, \\
\frac{\partial^2}{\partial x_2 \partial x_2} f(x) &= \frac{\partial}{\partial x_2}\left(\frac{\partial}{\partial x_2} f(x)\right) = \frac{\partial}{\partial x_2}(x_1 + \sqrt{x_3}) = 0, \\
\frac{\partial^2}{\partial x_3 \partial x_2} f(x) &= \frac{\partial}{\partial x_3}\left(\frac{\partial}{\partial x_2} f(x)\right) = \frac{\partial}{\partial x_3}(x_1 + \sqrt{x_3}) = \frac{1}{2\sqrt{x_3}}.
\end{aligned} \tag{56}$$

## Beispiel (2) (fortgeführt)

Für die zweiten partiellen Ableitungen hinsichtlich $x_3$ ergibt sich

$$
\begin{aligned}
\frac{\partial^2}{\partial x_1 \partial x_3} f(x) &= \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_3} f(x)\right) = \frac{\partial}{\partial x_1}\left(\frac{x_2}{2}\sqrt{x_3}\right) = 0, \\
\frac{\partial^2}{\partial x_2 \partial x_3} f(x) &= \frac{\partial}{\partial x_2}\left(\frac{\partial}{\partial x_3} f(x)\right) = \frac{\partial}{\partial x_2}\left(\frac{x_2}{2\sqrt{x_3}}\right) = \frac{1}{2\sqrt{x_3}}, \\
\frac{\partial^2}{\partial x_3 \partial x_3} f(x) &= \frac{\partial}{\partial x_3}\left(\frac{\partial}{\partial x_3} f(x)\right) = \frac{\partial}{\partial x_3}\left(x_2 \frac{1}{2} x_3^{-\frac{1}{2}}\right) = -\frac{1}{4} x_2 x_3^{-\frac{3}{2}}.
\end{aligned}
\tag{57}
$$

Weiterhin erkennt man, dass die Reihenfolge der partiellen Ableitungen irrelevant ist, denn es gilt

$$
\begin{aligned}
\frac{\partial^2}{\partial x_1 \partial x_2} f(x) &= \frac{\partial^2}{\partial x_2 \partial x_1} f(x) = 1, \\
\frac{\partial^2}{\partial x_1 \partial x_3} f(x) &= \frac{\partial^2}{\partial x_3 \partial x_1} f(x) = 0, \\
\frac{\partial^2}{\partial x_2 \partial x_3} f(x) &= \frac{\partial^2}{\partial x_3 \partial x_2} f(x) = \frac{1}{2\sqrt{x_3}}.
\end{aligned}
\tag{58}
$$

## Definition (Gradient)

$f : \mathbb{R}^n \to \mathbb{R}$ sei eine multivariate reellwertige Funktion. Dann ist der *Gradient* $\nabla f(x)$ von $f$ an der Stelle $x \in \mathbb{R}^n$ definiert als

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{pmatrix} \in \mathbb{R}^n. \tag{59}$$

Bemerkungen

- $\nabla f(x)$ fasst die partiellen Ableitungen von $f$ an der Stelle $x \in \mathbb{R}^n$ in einem Vektor zusammen.
- Gradienten sind multivariate vektorwertige Abbildungen der Form $\nabla f : \mathbb{R}^n \to \mathbb{R}^n, x \mapsto \nabla f(x)$.
- $-\nabla f(x)$ zeigt die Richtung des steilsten Abstiegs von $f$ in $\mathbb{R}^n$ an. Diese Einsicht ist aber nicht trivial.
- Für $n = 1$ gilt $\nabla f(x) = f'(x)$.

Der Erfolg neuronaler Netze (aka *AI* oder *Deep Learning*) basiert auf effizienten Gradientenberechnungen. Eine Einführung dazu geben Ostwald & Usée (2021) *An Induction Proof of the Backpropagation Algorithm in Matrix Notation.* https://arxiv.org/abs/2107.09384

Beispiele

Für die in Beispiel (1) betrachtete Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ gilt

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \in \mathbb{R}^2. \tag{60}$$

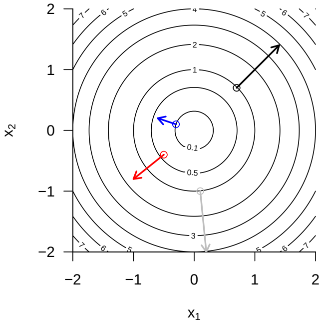
Für die in Beispiel (2) betrachtete Funktion $f : \mathbb{R}^3_{>0} \to \mathbb{R}$ gilt

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \frac{\partial}{\partial x_3} f(x) \end{pmatrix} = \begin{pmatrix} 2x_1 + x_2 \\ x_1 + \sqrt{x_3} \\ \frac{x_2}{2\sqrt{x_3}} \end{pmatrix} \in \mathbb{R}^3. \tag{61}$$

Beispiel (1) (fortgeführt)

Gradienten von $f : \mathbb{R}^2 \to \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2$ bei

$$x = \begin{pmatrix} 0.7 \\ 0.7 \end{pmatrix} \qquad x = \begin{pmatrix} -0.3 \\ 0.1 \end{pmatrix} \qquad x = \begin{pmatrix} -0.5 \\ -0.4 \end{pmatrix} \qquad x = \begin{pmatrix} 0.1 \\ -1.0 \end{pmatrix}$$



## Definition (Hesse-Matrix)

$f : \mathbb{R}^n \to \mathbb{R}$ sei ein multivariate reellwertige Funktion. Dann ist die *Hesse-Matrix* $\nabla^2 f(x)$ von $f$ an der Stelle $x \in \mathbb{R}^n$ definiert als

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \frac{\partial^2}{\partial x_n \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n \partial x_n} f(x) \end{pmatrix} \in \mathbb{R}^{n \times n}. \tag{62}$$

Bemerkung

- $\nabla^2 f(x)$ fasst die partiellen Ableitungen zweiter Ordnung von $f$ in einer Matrix zusammen.
- Hesse-Matrizen sind multivariate matrixwertige Abbildungen der Form $\nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n \times n}, x \mapsto \nabla^2 f(x)$.
- Für $n = 1$ gilt $\nabla^2 f(x) = f''(x)$.
- Mit $\frac{\partial^2}{\partial x_i \partial x_j} f(x) = \frac{\partial^2}{\partial x_j \partial x_i} f(x)$ für $1 \le i, j \le n$ folgt, dass $\left(\nabla^2 f(x)\right)^T = \nabla^2 f(x)$.

## Beispiel

Für die in Beispiel (1) betrachtete Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ gilt

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x) \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

Für die in Beispiel (2) betrachtete Funktion $f : \mathbb{R}^3_{>0} \to \mathbb{R}$ gilt

$$\begin{aligned} \nabla^2 f(x) &:= \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \frac{\partial^2}{\partial x_1 \partial x_3} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x) & \frac{\partial^2}{\partial x_2 \partial x_3} f(x) \\ \frac{\partial^2}{\partial x_3 \partial x_1} f(x) & \frac{\partial^2}{\partial x_3 \partial x_2} f(x) & \frac{\partial^2}{\partial x_3 \partial x_3} f(x) \end{pmatrix} \\ &:= \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & \frac{1}{2\sqrt{3}} \\ 0 & \frac{1}{2\sqrt{3}} & -\frac{1}{4} x_2 x_3^{-3/2} \end{pmatrix}. \end{aligned}$$

Grundlagen der Differentialrechnung

Analytische Optimierung

Differentialrechnung multivariater reellwertiger Funktionen

**Selbstkontrollfragen**

# Selbstkontrollfragen

1. Geben Sie die Definition des Begriffs der Ableitung $f'(a)$ einer Funktion $f$ an einer Stelle $a$ wieder.
2. Geben Sie die Definition des Begriffs der Ableitung $f'$ einer Funktion $f$.
3. Erläutern Sie die Symbole $f'(x), \dot{f}(x), \frac{df(x)}{dx}$, und $\frac{d}{dx}f(x)$.
4. Geben Sie die Definition des Begriffs der zweiten Ableitung $f''$ einer Funktion $f$ wieder.
5. Geben Sie die Summen-, Produkt-, Quotienten-, und Kettenregel für Ableitungen wieder.
6. Geben Sie die Definition der Begriffe des globalen und lokalen Maximums/Minimums einer univariaten reellwertigen Funktion wieder.
7. Geben Sie die notwendige Bedingung für ein Extremum einer Funktion wieder.
8. Geben Sie die hinreichende Bedingung für ein lokales Extremum einer Funktion wieder.
9. Geben Sie das Standardverfahren der analytischen Optimierung wieder.
10. Geben Sie die Definition der partiellen Ableitung $\frac{\partial}{\partial x_i}f(a)$ einer Funktion $f$ an einer Stelle $a$ wieder.
11. Geben Sie die Definition der partiellen Ableitung $\frac{\partial}{\partial x_i}f$ einer Funktion $f$ wieder.
12. Geben Sie die Defnition der zweiten partiellen Ableitung $\frac{\partial^2}{\partial x_i \partial x_j}f$ einer Funktion $f$ wieder.
13. Geben Sie den Satz von Schwarz wieder.
14. Geben Sie die Definition des Gradienten einer multivariaten reellwertigen Funktion wieder.
15. Geben Sie die Definition der Hesse-Matrix einer multivariaten reellwertigen Funktion wieder.

# Selbstkontrollfragen

1. Siehe Definition (Ableitung).
2. Siehe Definition (Ableitung).
3. Siehe Definition (Notation für Ableitungen univariater reellwertiger Funktionen).
4. Siehe Definition (Höhere Ableitungen).
5. Siehe Theorem (Rechenregeln für Ableitungen).
6. Siehe Definition (Extremstellen und Extremwerte).
7. Siehe Definition (Notwendige Bedingung für Extrema).
8. Siehe Definition (Hinreichende Bedingung für lokale Extrema).
9. Siehe Theorem (Standardverfahren der analytischen Optimierung)
10. Siehe Definition (Partielle Ableitung).
11. Siehe Definition (Partielle Ableitung).
12. Siehe Definition (Zweite partielle Ableitungen).
13. Siehe Theorem (Satz von Schwarz).
14. Siehe Definition (Gradient).
15. Siehe Definition (Hesse-Matrix).