Höhere Ableitungen

- Kapitel 6: Höhere Ableitungen
- Informationen zur Vorlesung:
  http://www.mathematik.uni-trier.de/∼wengenroth/

6.1 Bemerkung

(a) In Satz 5.4 haben wir als notwendiges Kriterium für Extremalstellen x einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$ die Bedingung $D_v f(x) = 0$ für alle Richtungen v und damit insbesondere $\nabla f(x) = 0$ gezeigt.
Genau wie in EA I für Funktionen $\mathbb{R} \to \mathbb{R}$ braucht man zusätzliche Kriterien, um zu entscheiden, ob tatsächlich eine Extremalstelle vorliegt.

(b) Beispiele:

- $f : \mathbb{R}^2 \to \mathbb{R}, f(x, y) = x^2 + y^2$ hat in 0 ein (sogar globales) Minimum
- $f : \mathbb{R}^2 \to \mathbb{R}, f(x, y) = xy$ hat nur einen „kritischen Punkt“ $\begin{bmatrix} x \\ y \end{bmatrix}$ mit $\nabla f(x, y) = 0$, nämlich $x = y = 0$, aber weder ein lokales Minimum noch ein lokales Maximum.

6.2 Konvexität

(a) Für ein Intervall $I \subseteq \mathbb{R}$ und $f : I \to \mathbb{R}$ wurde in EA I Konvexität definiert durch die Bedingung

$$\forall\, x, y \in I \quad \forall\, t \in [0,1] \quad f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

(b) Genauso wollen wir Konvexität für $f : A \to \mathbb{R}$ definieren, und dafür muss natürlich $tx + (1-t)y$ wieder ein Element von A sein.

(c) Eine Teilmenge $A \subseteq \mathbb{R}^n$ heißt konvex, falls für alle $x, y \in A$ und $t \in [0,1]$ der Vektor $tx + (1-t)y$ wieder in A liegt. Beachte, dass $S_{x,y} = \{tx + (1-t)y : t \in [0,1]\} = \{x + s(y-x) : s \in [0,1]\}$ das „Segment“ von x nach y beschreibt.

(d) Eine Funktion $f : A \to \mathbb{R}$ heißt konvex, falls
- $A \subseteq \mathbb{R}^n$ eine konvexe Menge ist, und es gilt
- $\forall\, x, y \in A \quad \forall\, t \in [0,1] \quad f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$.

6.3 Satz

Seien $A \subseteq \mathbb{R}^n$ eine konvexe offene Menge, $f : A \to \mathbb{R}$ konvex und $x \in A$, so dass f in x total differenzierbar ist. Falls $\nabla f(x) = 0$, so ist

$$f(x) = \min\{f(y) : y \in A\}.$$

Beweis. Für jedes $y \in A$ und $t \in [0,1]$ ist
$f(x + t(y-x)) = f(ty + (1-t)x) \leq tf(y) + (1-t)f(x)$, also

$$\frac{f(x + t(y-x)) - f(x)}{t} \leq f(y) - f(x).$$

Der Quotient links konvergiert für $t \to 0$ gegen die Richtungsableitung $D_{y-x}f(x) = \nabla f(x) \cdot (y-x) = 0$, und dies impliziert $0 \leq f(y) - f(x)$, also $f(x) \leq f(y)$.

$\square$

6.4 Höehere Ableitungen

(a) Wir wollen die Konvexität von $f : A \to \mathbb{R}$ analog zu EA I durch die Ableitungen charakterisieren. Dazu bemerken wir:
f konvex auf $A \iff$ Für alle $x, y \in A$ ist $g = f \circ \sigma$ konvex, wobei $\sigma(t) = x + t(y - x)$.
Falls $g = f \circ \sigma$ zweimal (nach (t)) differenzierbar ist, so besagt Satz 5.4.4 EA I: $g''(t) \geq 0$ für alle $t \in [0, 1] \Longrightarrow g$ konvex auf $[0, 1]$.
Ist g'' sogar stetig, so gilt auch die Umkehrung:
Beweis. Annahme, es gibt $t_0 \in (0, 1)$ mit $g''(t_0) < 0$. Wegen der Stetigkeit in t_0 gibt es $\delta > 0$, so dass $g''(t) < 0$ für alle $t \in (t_0 - \delta, t_0 + \delta) = I$. Dann ist g sowohl konvex als auch konkav auf I, und deshalb affin linear auf I. Dann ist aber $g'' = 0$ auf I und insbesondere $g''(t_0) = 0 \lightning$.

6.4 Höhere Ableitungen

(b) Wir wollen also die zweite Ableitung von $g(t) = f(\sigma(t))$ ausrechnen, und dazu brauchen wir zunächst $g'(t)$. Wegen der Kettenregel ist

$$\begin{aligned} g'(t) &= \nabla f(\sigma(t)) \cdot \sigma'(t) = \nabla f(\sigma(t)) \cdot (y - x) \\ &= \sum_{j=1}^{n} D_j f(\sigma(t))(y_j - x_j). \end{aligned}$$

Sind $D_j f : A \longrightarrow \mathbb{R}$ wiederum total differenzierbar, so liefern die Linearität und die Kettenregel (für $D_j f$ anstatt f)

$$g''(t) = \sum_{j=1}^{n} \sum_{k=1}^{n} D_k D_j f(\sigma(t))(y_k - x_k)(y_j - x_j)$$

6.4 Höhere Ableitungen

(c) Eine Funktion $f : A \to \mathbb{R}$ mit $A \subseteq \mathbb{R}^n$ offen, heißt zweimal stetig differenzierbar, falls alle partiellen Ableitungen $D_j f : A \to \mathbb{R}$ stetig differenzierbar sind.
Für jedes j haben wir dann die n partiellen Ableitungen $D_1 D_j f, \dots, D_n D_j f : A \to \mathbb{R}$. Für $x \in A$ heißt die Matrix

$$H_f(x) = \begin{bmatrix} D_1 D_1 f(x) & D_2 D_1 f(x) & \dots & D_n D_1 f(x) \\ D_2 D_1 f(x) & D_2 D_2 f(x) & \dots & D_n D_2 f(x) \\ \vdots & \vdots & & \vdots \\ D_n D_1 f(x) & D_n D_2 f(x) & \dots & D_n D_n f(x) \end{bmatrix}$$

die Hesse-Matrix von f im Punkt x.

6.4 Höhere Ableitungen

(d) Bezeichnen wir für $v = y - x \in \mathbb{R}^n = \mathbb{R}^{n\times 1}$ (Spaltenvektor) mit v^t den entsprechenden Zeilenvektor $v^t = [v_1, \dots, v_n]$, so ist die Doppelsumme in (c) gerade

$$\sum_{j=1}^{n} \sum_{k=1}^{n} D_k D_j f(\sigma(t)) v_k v_j = v^t \cdot H_f(\sigma(t)) \cdot v$$

(Beachte $\mathbb{R}^{1\times n} \cdot \mathbb{R}^{n\times n} \cdot \mathbb{R}^{n\times 1} \rightsquigarrow \mathbb{R}^{1\times n} \cdot \mathbb{R}^{n\times 1} \rightsquigarrow \mathbb{R}^{1\times 1} = \mathbb{R}$).

(d) Eine Matrix $H \in \mathbb{R}^{n\times n}$ heißt (streng) positiv definit, falls $v^t H v \geq 0$ (beziehungsweise > 0) für alle $v \in \mathbb{R}^n \setminus \{0\}$. Wir haben damit gezeigt:

6.5 Satz

Seien $A \subseteq \mathbb{R}^n$ konvex und offen und $f : A \to \mathbb{R}$ zweimal stetig differenzierbar, so dass die Hesse-Matrix $H_f(x)$ in jedem Punkt $x \in A$ positiv definit ist. Dann ist f konvex auf A.

6.6 Satz

Seien $A \subseteq \mathbb{R}^n$ offen, $f : A \to \mathbb{R}$ zweimal stetig differenzierbar und $x \in A$ mit $\nabla f(x) = 0$. Ist $H_f(x)$ streng positiv definit, so hat f in x ein lokales Minimum, das heißt, es gilt $\delta > 0$, so dass

$$f(x) = \min\{f(y) : y \in A \text{ und } \|y - x\| < \delta\}.$$

Gilt es v und w mit $v^t H_f(x) v > 0$ und $w^t H w < 0$, so hat f in x kein lokales Extremum.

Beweis. Aus der Stetigkeit von $D_j D_k f$ folgt die Existenz von $\delta > 0$, so dass $H_f(y)$ für jedes $y \in K(x, \delta)$ positiv definit ist. Weil A offen ist, kann man δ verkleinern, so dass $K(x, \delta) \subseteq A$ gilt. Wegen Satz 6.5 ist dann f auf $K(x, \delta)$ konvex und wegen Satz 6.3 ist

$$f(x) = \min\{f(y) : y \in K(x, \delta)\}.$$

□

Bemerkung: Falls $-H_f(x)$ streng positiv definit ist, hat f in x ein lokales Maximum.

WARNUNG: Oft ist es nicht leicht

- alle Punkte $x \in A$ mit $\nabla f(x)$ auszurechnen und
- zu überprüfen, ob $H_f(x)$ positiv definit ist.

6.7 Beispiel

Sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch $f(x,y) = \exp(-x^2 - 1 + \cos(y))$. Dann ist

$$
\begin{aligned}
D_1 f(x,y) &= -2x \exp(-x^2 - 1 + \cos(y)) \\
D_2 f(x,y) &= -\sin(y) \exp(-x^2 - 1 + \cos(y)) \\
D_1 D_1 f(x,y) &= -2 \exp(-x^2 - 1 + \cos(y)) + 4x^2 \exp(-x^2 - 1 + \cos(y)) \\
D_1 D_2 f(x,y) &= 2x \sin(y) \exp(-x^2 - 1 + \cos(y)) \\
D_2 D_1 f(x,y) &= 2x \sin(y) \exp(-x^2 - 1 + \cos(y)) \\
D_2 D_2 f(x,y) &= -\cos(y) \exp(-x^2 - 1 + \cos(y)) + \sin^2(y) \exp(-x^2 - 1 + \cos(y))
\end{aligned}
$$

Weil exp keine Nullstellen hat, gilt $\nabla f(x,y) = 0 \iff x = 0$ und $\sin(y) = 0$ und in diesen Punkten ist

$$
H_f(x,y) = \begin{bmatrix} -2\exp(-1+\cos(y)) & 0 \\ 0 & -\cos(y)\exp(-1+\cos(y)) \end{bmatrix}
$$

6.7 Beispiel

Diese Matrix ist von der Form $H_f(x,y) = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$ mit $a < 0$ und $b < 0$, falls $\cos(y) > 0$, sowie $b > 0$, falls $\cos(y) < 0$.

Für eine so strukturierte Matrix ist $v^t H v$ leicht auszurechnen:
$[v_1, v_2] \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = av_1^2 + bv_2^2$. Also ist $H_f(x,y)$ genau dann negativ definit, falls $\cos(y) > 0$. Falls $\cos(y) < 0$ hat f in $\begin{bmatrix} 0 \\ y \end{bmatrix}$ kein lokales Extremum.

Beobachtung: In diesem Beispiel gilt $D_1 D_2 f(x,y) = D_2 D_1 f(x,y)$. Dass dies kein Zufall ist, besagt folgender Satz.

6.8 Satz (Schwarz)

Seien $A \subseteq \mathbb{R}^n$ offen und $f : A \to \mathbb{R}$ zweimal stetig differenzierbar. Für alle $x \in A$ und $j, k \in \{1, \dots n\}$ gilt dann $D_k D_j f(x) = D_j D_k f(x)$, das heißt, man kann die Ableitungsreihenfolge vertauschen.

Plausibilität für $n = 2$:

$$
\begin{aligned}
D_2 D_1 f(x,y) &= \lim_{t \to 0} \frac{D_1 f(x, y+t) - D_1 f(x,y)}{t} \\
&= \lim_{t \to 0} \lim_{s \to 0} \frac{f(x+s, y+t) - f(x, y+t) - (f(x+s, y) - f(x,y))}{st} \\
&= \lim_{s \to 0} \lim_{t \to 0} \frac{f(x+s, y+t) - f(x+s, y) - (f(x, y+t) - f(x,y))}{st} \\
&= \lim_{s \to 0} \frac{D_2 f(x+s, y) - D_2(x,y)}{s} = D_1 D_2 f(x,y)
\end{aligned}
$$

□