§ 9 Höhere Ableitungen und lokale Extremstellen

- Seien V, W endlich-dimensionale, normierte $\mathbb{R}$-Vektorräume.
- Sei $U \subseteq V$ offen und $f : U \to W$ eine Funktion, die in jedem Punkt $x \in U$ differenzierbar ist.
- Dann kann die erste Ableitung als Funktion

$$f' : U \longrightarrow \mathscr{L}(V, W) \quad , \quad x \mapsto f'(x) \qquad \text{betrachtet werden.}$$

Definition (9.1)

Ist die Abbildung f' auf U stetig, dann bezeichnen wir f als stetig differenzierbare Funktion. Ist sie darüber hinaus in jedem Punkt $x \in U$ differenzierbar, dann sprechen wir von einer zweimal differenzierbaren Funktion.

Höhere Ableitungen

- Die zweite Ableitung ist eine Abbildung

$$f'' : U \to \mathscr{L}(V, \mathscr{L}(V, W)).$$

- Ist auch diese Abbildung differenzierbar, dann nennt man f dreimal differenzierbar.
- Definieren wir rekursiv $\hat{\mathscr{L}}^1(V, W) = \mathscr{L}(V, W)$ und

$$\hat{\mathscr{L}}^{n+1}(V, W) \quad = \quad \mathscr{L}(V, \hat{\mathscr{L}}^n(V, W))$$

  dann ist die n-te Ableitung eine Abbildung $f^{(n)} : U \to \hat{\mathscr{L}}^n(V, W)$.

Die totale Ableitung als multilineare Abbildung

- Sei V^n das n-fache kartesische Produkt $V \times ... \times V$.
- Eine Abbildung $\phi : V^n \to W$ heißt n-fach linear oder auch multilinear, wenn sie in jeder ihrer n Komponenten linear ist.
- Wir bezeichnen den $\mathbb{R}$-Vektorraum dieser Abbildungen mit $\mathscr{L}^n(V, W)$.

Die totale Ableitung als multilineare Abbildung

Satz (9.2)

Jedem Element $\hat{\phi} \in \hat{\mathscr{L}}^n(V, W)$ kann durch die Definition

$$\phi(v_1, ..., v_n) \quad = \quad \hat{\phi}(v_1)(v_2) \dots (v_n)$$

ein Element in $\mathscr{L}^n(V, W)$ zugeordnet werden, und die Abbildung $\Phi : \hat{\mathscr{L}}^n(V, W) \to \mathscr{L}^n(V, W)$, $\hat{\phi} \mapsto \phi$ ist ein Isomorphismus von $\mathbb{R}$-Vektorräumen.

Somit können wir die n-te Ableitung in jedem Punkt a als multilineare Abbildung $f^{(n)}(a) : V^n \to W$ betrachten.

Eindimensionale Taylorpolynome

- Bereits in der Analysis einer Variablen haben wir das p-te Taylorpolynom

$$\tau_p(f,a)(x) \quad = \quad \sum_{k=0}^{p} \frac{1}{k!} f^{(k)}(a)(x-a)^k.$$

  von f an der Stelle a definiert.

- Dieses Polynom ist charakterisiert durch die Eigenschaft $\tau_p(f,a)^{(k)}(x) = f^{(k)}(a)$ für $0 \leq k \leq p$. Denn k-fache Differentiation des Polynom ergibt

$$\tau_p(f,a)^{(k)}(x) \quad = \quad \sum_{\ell=k}^{p} \frac{1}{\ell!} f^{(k)}(a) \cdot k \cdot (k-1) \cdot \ldots \cdot (k-\ell+1) \cdot (x-a)^{k-\ell}$$

  so dass jeweils $\tau_p(f,a)^{(k)}(a) = \frac{1}{k!} f^{(k)}(a) \cdot k! \cdot (x-a)^0 = f^{(k)}(a)$ gilt.

Verallgemeinerung des Mittelwertsatzes

Satz (9.3)

Sei $p \in \mathbb{N}$ und $x \in \mathbb{R}$ ein Punkt mit $x > a$, so dass f auf dem offenen Intervall $]a, x[$ mindestens $(p-1)$-mal stetig differenzierbar und p-mal differenzierbar ist. Dann gibt es einen Punkt $\xi \in]a, x[$ mit

$$f(x) \quad = \quad \tau_{p-1}(f, a)(x) + \frac{1}{p!} f^{(p)}(\xi)(x-a)^p.$$

Mehrdimensionale Taylorpolynome

Definition (9.4)

Sei $p \in \mathbb{N}$, $U \subseteq V$ offen, $a \in U$ und $f : U \to \mathbb{R}$ eine im Punkt a p-mal differenzierbare Funktion. Dann bezeichnet man die Funktion gegeben durch

$$\tau_p(f,a)(x) \quad = \quad f(a) + \sum_{k=1}^{p} \frac{1}{k!} f^{(k)}(a)(\underbrace{x-a, ..., x-a}_{k\text{-mal}})$$

als Taylorpolynom p-ten Grades von f an der Stelle a.

Spezialfälle: Taylor-Polynome für Grad 1 und 2

$f : \mathbb{R}^n \to \mathbb{R}$ mind. zweimal diff'bar

Taylor-Polynom vom Grad 1:

$$\tau_1(f,a) = f(a) + f'(a)(x-a) =$$

$$f(a) + (\partial_1 f(a) \dots \partial_n f(a)) \begin{pmatrix} x_1 - a_1 \\ \vdots \\ x_n - a_n \end{pmatrix} =$$

$$f(a) + \sum_{i=1}^{n} \partial_i f(a)(x_i - a_i)$$

(*) siehe unten

Taylor-Polynom vom Grad 2:

$$\tau_2(f,a) = f(a) + f'(a)(x-a) + \frac{1}{2} f''(a)(x-a, x-a)$$

$$\overset{(*)}{=} f(a) + \sum_{i=1}^{n} \partial_i f(a)(x_i - a) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \partial_{ij} f(a)(x_i - a)(x_j - a)$$

Bsp.: $f: \mathbb{R}^2 \to \mathbb{R},\ (x,y) \mapsto x^2+3y^2-7xy+5x-3y+8$

Berechnung des 1 und 2. Taylor-Pol an der Stelle $(3,5)$

$$a=(3,5) \Rightarrow f(a) = 9+75-105+15-15+8 = -13$$

$$\partial_1 f(x,y) = 2x-7y+5 \qquad \partial_1 f(3,5) = -24$$

$$\partial_2 f(x,y) = 6y-7x-3 \qquad \partial_2 f(3,5) = 6$$

$$\Rightarrow \tau_1(f,a) = -13-24(x-3)+6(y-5)$$

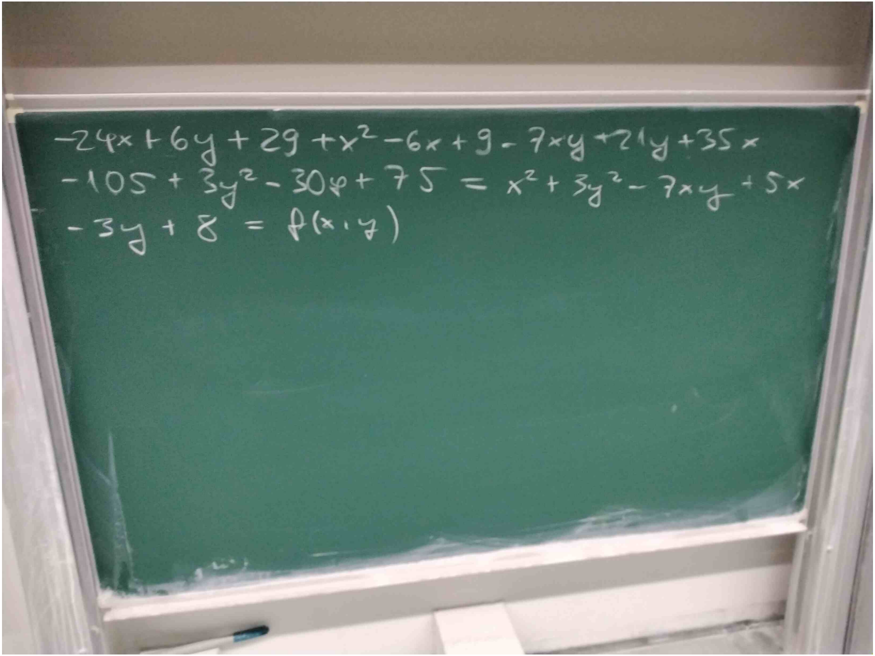
$$= -24x+6y+29$$

$$\partial_{11} f(x,y) = 2,\ \partial_{12} f(x,y) = \partial_{21} f(x,y) = -7;\ \partial_{22} f(x,y) = 6$$

$$\Rightarrow \tau_2(f,a) = \tau_1(f,a) + \tfrac{1}{2}\cdot 2\cdot(x-3)^2 + (-7)(x-3)(y-5)$$

$$+ \tfrac{1}{2}\cdot 6\,(y-5)^2 = \tau_1(f,a) + x^2-6x+9 + (-7)xy + 21y$$

$$+35x-105 + 3y^2-30y+75 =$$


$-24x + 6y + 29 + x^2 - 6x + 9 - 7xy + 21y + 35x$
$-105 + 3y^2 - 30y + 75 = x^2 + 3y^2 - 7xy + 5x$
$-3y + 8 = f(x,y)$

Darstellung der p-ten Ableitung durch mehrfache partielle Ableitungen

Proposition (9.5)

Sei $U \subseteq V$ offen, $a \in U$ und $f : U \to \mathbb{R}$ eine p-mal in a differenzierbare Funktion. Dann gilt für alle $v_1, ..., v_p \in V$ jeweils

$$f^{(p)}(a)(v_1, ..., v_p) \quad = \quad \partial_{v_1} \cdots \partial_{v_p} f(a).$$

$i=1 \quad j=1$

Beweis von Prop. (9.5)

$U \subseteq V$ offen, $a \in U$, $f : U \to \mathbb{R}$ p-mal diff'bar in a

$v_1, \dots, v_p \in V$ z.zg.: $f^{(p)}(a)(v_1, \dots, v_p) = \partial_{v_1} \dots \partial_{v_p} f(a)$

vollst. Ind. über $p \in \mathbb{N}_0$: Ind-Anf. $p = 0$: $f(a) = f(a)$ ($\checkmark$)

Ind-schritt $p \mapsto p+1$:

Definiere $\tilde{f} = \partial_{v_2} \dots \partial_{v_p} f$ und $g = f^{(p-1)}$ Ind-V. $\Rightarrow$

$\tilde{f}(a+h) = g(a+h)(v_2, \dots, v_p)$ für h in einer Umg. von 0_V

g diff'bar in a $\Rightarrow$ $g(a+h) = g(a) + g'(a)(h) + \psi(h)$

mit $\lim_{h \to 0} \frac{\psi(h)}{\|h\|} \overset{(*2)}{=} 0$ $\longrightarrow$ $\tilde{f}(a+h) = g(a+h)(v_2, \dots, v_p) =$

$g(a)(v_2, \dots, v_p) + g'(a)(h, v_2, \dots, v_p) + \varphi(h)$ mit

$\varphi(h) = \psi(h)(v_2, \dots, v_p)$ Wegen $(*_2)$ gilt auch

$\lim_{h \to 0} \frac{\varphi(h)}{\|h\|} = 0$ (folgt aus der Def der Operatornorm)

$\Rightarrow \tilde{f}$ ist in a total diff'bar, und $\tilde{f}'(a)(v_1) = g'(a)(v_1, \dots, v_p)$

Prop (8.4) $\Rightarrow \partial_{v_1} \tilde{f}(a) = \tilde{f}'(a)(v_1)$ insgesamt:

$\partial_{v_1} \partial_{v_2} \dots \partial_{v_p} f(a) = \partial_{v_1} \tilde{f}(a) = \tilde{f}'(a)(v_1) = g'(a)(v_1, \dots, v_p)$

$= f^{(p)}(a)(v_1, \dots, v_p)$. $\square$

wichtiger Spezialfall: $f''(a)(v,w) = \sum_{i=1}^{n} \sum_{j=1}^{n} \partial_{ij} f(a) v_i w_j$

denn $f''(a)(v,w) = f''(a)\left(\sum_{i=1}^{n} v_i e_i, \sum_{j=1}^{n} w_j e_j\right) =$

$\sum_{i=1}^{n} \sum_{j=1}^{n} v_i w_j f''(a)(e_i, e_j) \overset{\text{Prop (9.5)}}{=} \sum_{i=1}^{n} \sum_{j=1}^{n} v_i w_j \partial_{e_i} \partial_{e_j} f(a)$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} v_i w_j \, \partial_{ij} f(a).$$

Mehrdimensionale Taylor-Approximation

Satz (9.6)

Sei $U \subseteq \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ eine p-mal stetig differenzierbare Funktion. Dann gibt es zu jedem Punkt $a \in U$ eine Umgebung U_0 des Nullpunkts und eine Funktion $\psi : U_0 \to \mathbb{R}$ mit

$$f(a+h) \quad = \quad \tau_p(f,a)(a+h) + \psi(h) \quad \text{für alle } h \in U_0$$

und $\lim\limits_{h \to 0} \dfrac{\psi(h)}{\|h\|^p} = 0.$

Die Darstellungsmatrix der zweiten Ableitung

Sei $f : \mathbb{R}^n \to \mathbb{R}$ zweimal differenzierbar.

- Wie wir bereits wissen, kann die Ableitung $f'(a)$ als $(n \times 1)$-Matrix dargestellt werden.
- Die zweite Ableitung $f''(a)$ ist eine bilineare Abbildung $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, also eine Bilinearform auf dem $\mathbb{R}^n$.
- Die Darstellungsmatrix $A \in \mathscr{M}_{n,\mathbb{R}}$ dieser Bilinearform hat nach Prop. (9.5) die Einträge

$$a_{ij} = f''(a)(e_i, e_j) = \partial_{ij} f(a)$$

  für $1 \leq i, j \leq n$.
- Sie wird die Hesse-Matrix von f an der Stelle a genannt und mit $\mathscr{H}(f)(a)$ bezeichnet.

Approximation zweiter Ordnung

Folgerung (9.7)

Sei $U \subseteq \mathbb{R}^n$ offen, $f : U \to \mathbb{R}$ eine Funktion und $a \in U$ ein Punkt, in dem f zweimal differenzierbar ist. Dann gibt es eine Funktion $\psi : U_a \to \mathbb{R}$ mit

$$f(a+h) = f(a) + f'(a) \cdot h + \frac{1}{2}\, {}^{\mathrm{t}}h \cdot \mathscr{H}(f)(a) \cdot h + \psi(h)$$

und $\lim\limits_{h \to 0} \|h\|^{-2}\, \psi(h) = 0$.

Definition lokaler Extrema

Definition (9.8)

Sei $U \subseteq \mathbb{R}^n$ eine beliebige Teilmenge, $a \in U$ und $f : U \to \mathbb{R}$ eine reellwertige Funktion. Man sagt, f hat im Punkt a ein

(i) lokales Maximum, wenn eine Umgebung $U' \subseteq \mathbb{R}^n$ von a existiert, so dass $f(a) \geq f(x)$ für alle $x \in U \cap U'$ gilt,

(ii) isoliertes lokales Maximum, wenn U so gewählt werden kann, dass sogar $f(a) > f(x)$ für alle $x \in (U \cap U') \setminus \{a\}$ erfüllt ist.

Entsprechend definiert man lokale Minima und isolierte lokale Minima. Wie in der Analysis einer Variablen verwenden wir den Begriff Extremum als Oberbegriff für Minima und Maxima.

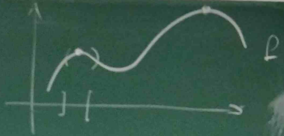
Notwendiges Kriterium für Extrema

Satz (9.9)

Sei $U \subseteq \mathbb{R}^n$ offen, $f : U \to \mathbb{R}$ differenzierbar und $a \in U$ ein lokales Extremum von f. Dann gilt $f'(a) = 0$.

Bsp: (1) $A = \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix}$ ist positiv definit, denn

Extrema im Eindimensionalen:



Beweis von Satz (9.9):

$U \subseteq \mathbb{R}^n$ offen, $f: U \to \mathbb{R}$ diff'bar, $a \in U$ lokales Extremum von f. <u>Beh.</u>: $f'(a) = 0$

Sei $v \in \mathbb{R}^n$, betrachte die Hilfsfunktion $\phi: \mathbb{R} \to \mathbb{R}^n$, $t \mapsto a + tv$ und setze $\tilde{f}(t) = (f \circ \phi)(t) = f(a + tv)$

O.B.d.A. sei a ein lokales Maximum von f. Dann gibt es ein $\varepsilon \in \mathbb{R}^+$ mit $\tilde{f}(t) = f(a + tv) \leq f(a)$ für alle $t \in \mathbb{R}$ mit $|t| < \varepsilon$. denn: a lok. Max $\rightarrow$

$\exists$ Umg. $U' \subseteq U$ mit $f(x) \le f(a)$ $\forall x \in U'$

Ist $|t|$ hinreichend klein, dann liegt $a+tv$ in U', auf Grund der Stetigkeit von ϕ.

$\Longrightarrow \tilde{f}$ hat in 0 ein lokales Maximum $\Rightarrow \tilde{f}'(0) = 0$ [Analysis einer Variablen]

$\Longrightarrow \partial_v f(a) = 0 \Rightarrow f'(a)(v) = 0$, für jedes $v \in \mathbb{R}^n$.

Daraus folgt $f'(a) = 0$. $\square$

Bsp: (i) $A = \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix}$ ist positiv definit, denn für alle Vektoren $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ gilt $(v_1\ v_2)\begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} =$

$(v_1\ v_2)\begin{pmatrix} 3v_1 \\ 5v_2 \end{pmatrix} = 3v_1^2 + 5v_2^2 > 0$

(ii) $B = \begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix}$ ist indefinit, da $(1\ 0)\begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = -2 < 0$, $(0\ 1)\begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 1 > 0$

Eigenschaften symmetrischer Matrizen

Definition (9.10)

Wir bezeichnen eine symmetrische Matrix $A \in \mathscr{M}_{n,\mathbb{R}}$ als

- negativ definit, wenn ${}^{\mathrm{t}}vAv > 0$ für alle $v \in \mathbb{R}^n \setminus \{0\}$, als
- positiv semidefinit, wenn ${}^{\mathrm{t}}vAv \geq 0$ für alle $v \in \mathbb{R}^n$, und als
- negativ semidefinit, wenn ${}^{\mathrm{t}}vAv \leq 0$ für alle $v \in \mathbb{R}^n$ erfüllt ist.

Eine Matrix, die weder positiv noch negativ semidefinit ist, bezeichnen wir als indefinit.

Hinreichendes Kriterium für Extrema

Satz (9.11)

Sei $U \subseteq \mathbb{R}^n$ offen, $f : U \to \mathbb{R}$ eine zweimal stetig differenzierbare Funktion und $a \in U$ eine kritische Stelle von f.

- Ist $\mathscr{H}(f)(a)$ positiv definit, dann besitzt f in a ein isoliertes lokales Minimum.
- Ist $\mathscr{H}(f)(a)$ negativ definit, dann besitzt f in a ein isoliertes lokales Maximum.
- Ist $\mathscr{H}(f)(a)$ indefinit, dann hat f in a kein lokales Extremum.

Bew. von Satz (9.11) $U \subseteq \mathbb{R}^n$ offen, $f: U \to \mathbb{R}$ zweimal stetig diff'bar, $a \in U$ mit $f'(a) = 0$

Sei $A \in M_{n,\mathbb{R}}$ die Hesse-Matrix von f in a.

zu (i) Vor.: A positiv definit z.zg.: a ist lokales Minimum

A positiv definit $\Rightarrow$ ${}^t h A h > 0$ für alle $h \in \mathbb{R}^n \setminus \{0\}$

Sei $S = \{ h \in \mathbb{R}^n \mid \|h\| = 1 \}$, betrachte $S \to \mathbb{R}$, $h \mapsto {}^t h A h$ Maximumsprinzip $\Rightarrow$ Die Funktion nimmt auf S ein globales, positives Minimum m an.

$\lim_{h \to 0} \frac{\psi(h)}{\|h\|^2} = 0 \Rightarrow \exists \varepsilon \in \mathbb{R}^+$ mit $|\psi(h)| \leq \frac{1}{4} m \|h\|^2$ für alle $h \in \mathbb{R}^n$ mit $\|h\| < \varepsilon$

Sei $A \in M_{n,\mathbb{R}}$ die Hesse-Matrix ...

(ausführlich: $\lim\limits_{h \to 0} \frac{\psi(h)}{\|h\|^2} = 0 \Rightarrow$ Es gibt ein $\varepsilon \in \mathbb{R}^+$ mit $\frac{|\psi(h)|}{\|h\|^2} < \frac{1}{4} m \quad \forall h \in \mathbb{R}^n$ mit $\|h\| < \varepsilon \Rightarrow |\psi(h)| \leq \frac{1}{4} m \|h\|^2$ für diese h.) Sei $h \in \mathbb{R}^n$ mit $0 < \|h\| < \varepsilon$.

Sei $h_0 = \|h\|^{-1} h \in S \Rightarrow {}^t h_0 A h_0 \geq m \Rightarrow$
${}^t h A h \geq m \|h\|^2 \Rightarrow$

$$f(a+h) = f(a) + f'(a) h + \frac{1}{2} {}^t h A h + \psi(h)$$
$$\geq f(a) + \frac{1}{2} m \|h\|^2 - \frac{1}{4} m \|h\|^2 = f(a) + \frac{1}{4} m \|h\|^2 > f(a) \quad \square$$