4. Mittelwertsatz und höhere Ableitungen

<u>Satz 4.1</u> (Mittelwertsatz) Seien $U \subset \mathbb{R}^n$ offen, $f: U \to \mathbb{R}$ differenzierbar und $a, b \in U$. Sei $S_{a,b} := \{x \in \mathbb{R}^n \mid \exists t \in [0,1]: x = a + t(b-a)\} \subset U$ (Strecke von a nach b liegt in U). Dann existiert ein $c \in S_{a,b}$ mit

$$f(b) = f(a) + f'(c)(b-a) = f(a) + \nabla f(c)(b-a)$$

<u>Bemerkung 4.2</u> a) Es gilt folgende Verallgemeinerung: Sei $\gamma: [0,1] \to U$ differenzierbar mit $\gamma(0) = a$, $\gamma(1) = b$, (d.h., γ ein differenzierbarer Weg von a nach b). Dann gilt $f(b) = f(a) + f'(\gamma(\vartheta))\gamma'(\vartheta)$ für ein $\vartheta \in (0,1)$.

b) Andere Formulierung des MWS: Für $x \in U$, $|h|$ klein, gilt

$$f(x+h) = f(x) + f'(x + \vartheta h) h \qquad \text{für ein } \vartheta \in (0,1).$$

c) Für $f: \mathbb{R}^n \to \mathbb{R}^m$ gilt der Mittelwertsatz in jeder Komponente, also mit einem anderen Mittelwert ϑ_j in jeder Komponente. ◊

<u>Korollar 4.3</u> Sei $U \subset \mathbb{R}^n$ offen und wegzusammenhängend. Sei $f: U \to \mathbb{R}$ differenzierbar. Dann gilt: $\nabla f = 0 \iff f$ ist konstant.

<u>Satz 4.4</u>. Seien $U \subset \mathbb{R}^n$ offen, $a \in U$, $f: U \to \mathbb{R}^m$. Falls $\partial_1 f, \ldots, \partial_n f$ ($\partial_j = \frac{\partial}{\partial x_j}$) in einer Umgebung von $U(a)$ von a existieren und in a stetig sind, so ist f differenzierbar in a.

<u>Bemerkung 4.5</u>. Falls die partiellen Ableitungen $\partial_j f$ in einer offenen Umgebung $U(a)$ von a stetig sind, so ist f in $U(a)$ differenzierbar, und f' ist in $U(a)$ stetig. ◊

Seien nun $U \subset \mathbb{R}^n$ offen, $f: U \to \mathbb{R}^m$ differenzierbar. Dann gibt es zwei Interpretationen von f': einmal als Abbildung $f': U \to L(\mathbb{R}^n; \mathbb{R}^m)$, $x \mapsto f'(x)$ oder als Abbildung $f': U \times \mathbb{R}^n \to \mathbb{R}^m$, $(x,h) \mapsto f'(x,h) := f'(x)h$. Wegen $L(\mathbb{R}^n; \mathbb{R}^m) \cong \mathbb{R}^{m \times n} \cong \mathbb{R}^{m \cdot n}$ können höhere Ableitungen wie folgt definiert werden

$f'': U \to L(\mathbb{R}^n; L(\mathbb{R}^n; \mathbb{R}^m)) \cong L(\mathbb{R}^n, \mathbb{R}^{m\times n})$,

$f''': U \to L(\mathbb{R}^n; L(\mathbb{R}^n; L(\mathbb{R}^n; \mathbb{R}^m))) \cong L(\mathbb{R}^n, L(\mathbb{R}^n; \mathbb{R}^{m\times n}))$

etc.

beziehungsweise

$f'': U \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$, $(x, h^{(1)}, h^{(2)}) \longmapsto f''(x, h^{(1)}, h^{(2)}) = (f''(x, h^{(1)}))\, h^{(2)}$.

Dabei ist $f''(x, h^{(1)}, h^{(2)})$ linear in $h^{(1)}$ und $h^{(2)}$. Höhere Ableitungen ergeben nichtlineare Abbildungen.

Schreibe $\partial_j f = (\partial_j f_1, \dots, \partial_j f_m)^T$, $\partial_j = \frac{\partial}{\partial x_j}$ und $\nabla = (\partial_1, \dots, \partial_n)^T$. Dann ist $h^{(1)} \in \mathbb{R}^n$:

$$f'(x, h^{(1)}) = f'(x)\, h^{(1)} = \begin{pmatrix} \partial_1 f_1 \dots \partial_n f_1 \\ \vdots \\ \partial_1 f_m \dots \partial_n f_m \end{pmatrix} \begin{pmatrix} h_1^{(1)} \\ \vdots \\ h_n^{(1)} \end{pmatrix} = \sum_{j=1}^n h_j^{(1)} \partial_j f = \underbrace{((h^{(1)})^T \cdot \nabla)}_{\text{Skalarprodukt}} f.$$

Für die zweite Ableitung erhält man

$$f''(x, h^{(1)}, h^{(2)}) = \sum_{i=1}^n h_i^{(2)} \partial_i \Big(\sum_{j=1}^n h_j^{(1)} \partial_j f \Big) = \sum_{i,j=1}^n h_i^{(2)} h_j^{(1)} \partial_i \partial_j f$$
$$= ((h^{(2)})^T \cdot \nabla)((h^{(1)})^T \cdot \nabla) f.$$

Für die p-te Ableitung gilt: Für $h^{(1)}, \dots, h^{(p)} \in \mathbb{R}^n$ ist

$$f^{(p)}(x, h^{(1)}, h^{(2)}, \dots, h^{(p)}) = ((h^{(p)})^T \cdot \nabla) \dots ((h^{(1)})^T \cdot \nabla) f.$$

Beachte, dass für die zweiten Ableitungen gilt

$$\partial_i \partial_j f(x) = f''(x, e_i, e_j) = \begin{pmatrix} \partial_i \partial_j f_1 \\ \vdots \\ \partial_i \partial_j f_m \end{pmatrix}, \quad h^{(1)} = e_j,\ h^{(2)} = e_i$$

<u>Satz 4.6.</u> Seien $U \subset \mathbb{R}^n$ offen und $f: U \to \mathbb{R}^m$ zweimal differenzierbar. Dann gilt für alle $x \in U$:

a) $\forall h^{(1)}, h^{(2)} \in \mathbb{R}^n$: $f''(x, h^{(1)}, h^{(2)})$
$$= \lim_{s \searrow 0} \frac{1}{s^2}\big(f(x + s h^{(1)} + s h^{(2)}) - f(x + s h^{(1)}) - f(x + s h^{(2)}) + f(x) \big).$$

b) $\forall h^{(1)}, h^{(2)} \in \mathbb{R}^n$: $f''(x, h^{(1)}, h^{(2)}) = f''(x, h^{(2)}, h^{(1)})$.

c) $\forall i,j \in \{1, \dots, n\}$: $\partial_i \partial_j f(x) = \partial_j \partial_i f(x)$.

<u>Bemerkung 4.7.</u> a) Für die Aussage des Satzes reicht es, dass f einmal differenzierbar ist in einer Umgebung von x und dass $f''(x)$ existiert.

b) Falls die zweiten partiellen Ableitungen $\partial_i \partial_j f(x)$ in einer Umgebung $U(a)$ von a existieren und in a stetig sind, so existiert $f''(a)$. Falls die zweiten partiellen Ableitungen in einer Umgebung $U(a)$ existieren und stetig sind, ist $f''(a)$ in $U(a)$ stetig (vgl. Bemerkung 4.5). $\Diamond$

Satz 4.8. (Taylor) Seien $U \subset \mathbb{R}^n$ offen und $f: U \to \mathbb{R}$ $(p+1)$-mal differenzierbar. Seien $x \in U$ und $h \in \mathbb{R}^n$ so klein, dass die Strecke $S_{x,x+h}$ von x nach $x+h$ in U liegt. Dann gilt:

$$f(x+h) = f(x) + f'(x,h) + \dots + \frac{1}{p!} f^{(p)}(x, \underbrace{h,\dots,h}_{p\text{-mal}}) + R_p(x,h)$$

mit

$$R_p(x,h) = \frac{1}{(p+1)!} f^{(p+1)}(x+\vartheta h, \underbrace{h,\dots,h}_{p\text{-mal}}) \quad \text{für ein } \vartheta \in (0,1).$$

Bemerkung 4.9. a) Falls $f \in C^{p+1}(U,\mathbb{R})$, folgt die Restglieddarstellung

$$R_p(x,h) = \frac{1}{p!} \int_0^1 (1-t)^p f^{(p+1)}(x+th, \underbrace{h,\dots,h}_{p\text{-mal}})\, dt.$$

b) Falls $f: U \to \mathbb{R}^m$, so gilt die Taylorentwicklung komponentenweise. Aber der Zwischenwert $\vartheta = \vartheta_j$ ist in jeder Komponente verschieden. $\Diamond$

Bemerkung 4.10 Speziell für $p=2$ erhalten wir

$$f'(x,h) = f'(x)h = (\partial_1 f(x), \dots, \partial_n f(x)) \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = \nabla f(x)^T h$$

und

$$f''(x,h,h) = \sum_{i,j=1}^n \partial_i \partial_j f(x) h_i h_j = (h_1,\dots,h_n) \begin{pmatrix} \partial_1^2 f(x) & \dots & \partial_1\partial_n f(x) \\ \vdots & & \vdots \\ \partial_n\partial_1 f(x) & \dots & \partial_n^2 f(x) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$$
$$= h^T \operatorname{Hess} f(x)\, h = h^T \nabla^2 f(x)\, h$$

mit der Hessematrix von f an der Stelle x

$$\operatorname{Hess} f(x) := \begin{pmatrix} \partial_1^2 f(x) & \dots & \partial_1\partial_n f(x) \\ \vdots & & \vdots \\ \partial_n\partial_1 f(x) & \dots & \partial_n^2 f(x) \end{pmatrix},$$

Nach Satz 4.6 ist $\operatorname{Hess} f(x)$ symmetrisch. Mit dieser Bezeichnung lautet die Taylorformel für dreimal differenzierbares $f: U \to \mathbb{R}$

$$f(x+h) = f(x) + \nabla f(x)^T h + h^T \nabla^2 f(x)\, h + R_2(x,h). \qquad \Diamond$$

Bemerkung 4.11. (Multiindex-Schreibweise). Sei $f: \mathbb{R}^n \supset U \to \mathbb{R}$ hinreichend oft differenzierbar. Im Satz von Taylor tauchen Ausdrücke der Form

$$f^{(k)}(x, \underbrace{h, \ldots, h}_{k\text{-mal}}) = \sum_{i_1, \ldots, i_k = 1}^{n} \partial_{i_1} \ldots \partial_{i_k} f(x)\, h_{i_1} \ldots h_{i_k}$$

auf. Dabei treten einige Terme in der Summe öfter auf, z. B.

$$\partial_1 \partial_2 f(x)\, h_1 h_2 = \partial_2 \partial_1 f(x)\, h_2 h_1$$

nach Satz 4.6.

Sei $\alpha \in (\mathbb{N}_0)^n$ (α heißt Multiindex) und $h \in \mathbb{R}^n$. Definiere

$$\partial^\alpha := \partial_1^{\alpha_1} \ldots \partial_n^{\alpha_n}, \quad h^\alpha := h_1^{\alpha_1} \ldots h_n^{\alpha_n}, \quad \alpha! = \alpha_1! \ldots \alpha_n!, \quad |\alpha| = \sum_{i=1}^{n} \alpha_i.$$

Zu $\alpha \in (\mathbb{N}_0)^n$ und $k := |\alpha|$ gibt es

$$\frac{k!}{\alpha_1! \ldots \alpha_n!} = \frac{k!}{\alpha!}$$

Summanden mit $\alpha_1 + \ldots + \alpha_n = k$, welche die Form

$$\frac{1}{k!} f^{(k)}(x, \underbrace{h, \ldots, h}_{k\text{-mal}}) = \sum_{|\alpha| = k} \frac{1}{\alpha!} \partial^\alpha f(x)\, h^\alpha.$$

Die Taylorformel lautet somit

$$f(x+h) = \sum_{k=0}^{p} \sum_{|\alpha| = k} \frac{1}{\alpha!} \partial^\alpha f(x)\, h^\alpha + R_p(x,h)$$
$$= \sum_{|\alpha| \le k} \frac{1}{\alpha!} \partial^\alpha f(x)\, h^\alpha + R_p(x,h)$$

$\Diamond$

<u>Definition 4.12.</u> Seien $U \subset \mathbb{R}^n$ offen und $f: U \to \mathbb{R}$ differenzierbar. Dann heißt $a \in U$ eine <u>kritische Stelle</u> von f, falls $\nabla f(a) = 0$ gilt, d.h.,

$$f'(a) = 0 \iff \forall h \in \mathbb{R}^n: f'(a,h) = 0$$

<u>Satz 4.13.</u> Seien $U \subset \mathbb{R}^n$ offen, $f: U \to \mathbb{R}$ differenzierbar und a lokales Extremum von f. Dann ist a kritische Stelle von f.

<u>Satz 4.14.</u> Seien $U \subset \mathbb{R}^n$ offen, $f \in C^2(U; \mathbb{R})$ und $a \in U$ eine kritische Stelle von f. Falls $\operatorname{Hess} f(a)$ positiv (negativ) definit ist, so ist a lokales Minimum (Maximum) von f. Falls $\operatorname{Hess} f(a)$ indefinit, so ist a kein lokales Extremum (a heißt dann Sattelpunkt).

Beispiel 4.15. Welcher Quader hat das größte Volumen? Sei $f:(0,3)\to\mathbb{R}$ mit $f(x_1,x_2):=x_1x_2(3-x_1-x_2)$, also Volumen des Quaders mit Summe der Kantenlänge gleich 3. Wir berechnen

$$\nabla f(x_1,x_2)=\begin{pmatrix}3x_2-2x_1x_2-x_2^2\\ 3x_1-2x_1x_2-x_1^2\end{pmatrix}.$$

Wegen $x_1,x_2\in(0,3)$ ist $\nabla f(x)=0$ nur für $a=\binom{1}{1}$. Ferner

$$\operatorname{Hess} f(x)=\left(\begin{array}{c|c}-2x_2 & 3-2x_1-2x_2\\ 3-2x_1-2x_2 & -2x_1\end{array}\right),$$

das heißt,

$$\operatorname{Hess} f(a)=\begin{pmatrix}-2 & -1\\ -1 & -2\end{pmatrix}.$$

Also $\langle h,\operatorname{Hess} f(a)h\rangle=-2h_1^2-2h_1h_2-2h_2^2=-h_1^2-h_2^2-(h_1+h_2)^2<0$ für alle $h\in\mathbb{R}^2\setminus\{0\}$. Also ist $\operatorname{Hess} f(a)$ negativ definit, d.h., an $\binom{1}{1}$ liegt ein lokales Maximum vor. Am Rand $x_1\in\{0,3\}$, $x_2\in\{0,3\}$ gilt $f(x)\leq 0$. Wegen $f(a)=1$ liegt also am Rand auch kein Maximum vor. Somit hat der Würfel das größte Volumen. $\diamond$

Beispiel 4.16. (Lineare Ausgleichsrechnung, Methode der kleinsten Quadrate). Seien $A\in\mathbb{R}^{m\times n}$ mit $\operatorname{rk} A=n$ (d.h., $m\geq n$) und $b\in\mathbb{R}^m$. Setze

$$f(x)=|Ax-b|^2=(Ax-b)^T(Ax-b)=x^TA^TAx-b^TAx-x^TA^Tb+b^Tb.$$

Um ∇f zu berechnen, beachte folgende Ableitungen

$$\begin{aligned}
f_1(x)&=c^Tx,\ c\in\mathbb{R}^n &&\Rightarrow f_1'(x)=c^T\\
f_2(x)&=x^Td,\ d\in\mathbb{R}^n &&\Rightarrow f_2'(x)=d^T\\
f_3(x)&=x^TBx,\ B\in\mathbb{R}^{n\times n} &&\Rightarrow f_3'(x)\underset{\text{Produktregel}}{=}x^TB+(Bx)^T=x^TB+x^TB^T
\end{aligned}$$

Damit erhalten wir $f'(x)=2x^TA^TA-2b^TA$, $f''(x)=2A^TA$. Es gilt

$$\nabla f(x)=0\iff A^TAx=A^Tb$$

Das sind die sogenannten Normalgleichungen. Wegen $\operatorname{rk} A=n$ ist A^TA invertierbar. Der einzige kritische Punkt ist

$$x_0:=(A^TA)^{-1}A^Tb=A^+b$$

mit der Pseudoinversen $A^+ = (A^T A)^{-1} A^T$. Wegen $\langle h, \text{Hess} f(x_0) h \rangle = 2 \langle h, A^T A h \rangle = 2 |Ah|^2 > 0$ für $h \in \mathbb{R}^n \setminus \{0\}$ ist $\text{Hess} f(x_0)$ positiv definit, d.h., x_0 ist ein (strikter) lokales Minimum von f. $\Diamond$