Teil 1: Differentialrechnung im $\mathbb{R}^n$

an7: Satz von Taylor, Lokale Extrema

Stichworte: Satz von Taylor, Extrema, Kritische Stellen, Kriterien, Hessematrix

Literatur: [Hoff] Kapitel, 9.6/7, [Forster] Kapitel 7

7.1. Einleitung: Der Satz von Taylor in der mehrdimensionalen Version für Skalarfelder liefert Kriterien zur Erkennung von Extrema anhand Gradienten und Hessematrix.

7.2. Vor.: $f : U \to \mathbb{R}$ mit $U \subset \mathbb{R}^n$, $f \in \ell^{m+1}(U, \mathbb{R})$, $\overline{ax} \subseteq U$.
Bezeichnung: Für $\alpha \in \mathbb{N}_0^n, \alpha = (\alpha_1, ..., \alpha_n)$ setze zur einfacheren Notation
$|\alpha|{:=}\alpha_1 + ... + \alpha_n$, $\alpha!{:=}\ \alpha_1! \cdots \alpha_n!$
$D^\alpha{:=}D_1^{\alpha_1} \circ D_2^{\alpha_2} \circ ... \circ D_n^{\alpha_n}$, $x^\alpha{:=}x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, $X^\alpha{:=}X_1^{\alpha_1} \cdots X_n^{\alpha_n}$.
Man nennt α auch einen Multi-Index.
Damit kann jedes Polynom $P \in \mathbb{R}[X_1^-, ..., X_n^-]$, deg P=m, auch in der Kurzen Multi-Index-Schreibweise notiert werden als
$P = \sum_{\alpha\in\mathbb{N}_0^n;|\alpha|\le m} c_\alpha X^a$, d.h. $P(X_1-, ..., X_n) = \sum_{0\le\alpha_1,...,\alpha_n\le m,\alpha_1+...+\alpha_n\le m} c_{(\alpha_1,...,\alpha_n)} X_1^{\alpha_1} \cdots X_n^{\alpha_n}$.

Bsp.: $\sum_{|\alpha|\le 2} \alpha! X^\alpha = \underbrace{0!X_1^0}_{\text{Grad 0}} + \underbrace{\sum_{i=1}^{n} 1!X_i^1}_{\text{Grad 1}} + \underbrace{\sum_{1\le i\ne j\le n} 1!1!X_i^1 X_j^1 + \sum_{i=1^n 2!X_i^2}}_{\text{Grad 2}}$

Damit kann der Satz von Taylor in einer Kurzgefassten Formel notiert werden:
7.3. Satz von Taylor: Unter der Vor. wie in 7.2 gilt:
Beh.: $\exists c \in \overline{ax}$:

$$\boxed{f(x) = \sum_{|\alpha|\le m} \frac{1}{\alpha!}(D^\alpha f)(a)(x-a)^\alpha + \sum_{|\alpha|=m+1} \frac{1}{\alpha!}(D^\alpha f)(c)(x-a)^\alpha}$$

7.4. Bem.: Für m=0 lautet die Beh. $f(x) = f(a) + Df(c)(x-a)$ für ein $c \in \overline{ax}$, dies is die Aussage des MWS 6.4.

7.5. Kor.: Für m=1 lautet die Beh.
$f(x) = f(a) + <\text{grad } f(a), x-a> + \frac{1}{2}(x-a)^T H(f_ic)(x-a)$
mit der (laut dem Satz von Schwarz) symmetrischen Matrix $H(f_ic){:=}(D_iD_jf(c))_{n,n}$, die Hessematrix heißt; die zugehörige quadratische Form heißt Hesseform.
(Auch: Schreibweise Hess f(c) statt $H(f_ic)$ üblich.)
(Jede symmetrische Matrix A, wo $A^T = A$, definiert über $<x, Ax> = x^T Ax$ eine quadratische Form.)

7.6. Bem.: Ist P ein Polynom, $P \in \mathbb{R}[X_1, ..., X_n]$, deg P=m, dann ist $P \in \ell^\infty(\mathbb{R}^n)$ und P ist seine (eigene) Taylorreihe.

7.7. Beweis des Satzes 7.3 von Taylor:
1. Schritt: Für $\epsilon{>}0, t\ \in]-\epsilon, 1+\epsilon[\subseteq \mathbb{R}$ setze $g(t) := f(a + t(x-a))$ für festes x und a. Es ist also $g \in \ell^{m+1}(]-\epsilon, 1+\epsilon[)$.
Beh.: Für $k \subseteq m+1$ ist $\frac{d^k g}{dt^k}(t) \overset{!}{=} \sum_{|\alpha|=k} \frac{k!}{\alpha!} D^\alpha f(a + t(x-a))$.
Bew.: Setze zunächst y:=x-a$=:(\eta_1, ..., \eta_n)^T$.

⌜Beh.: $\frac{d^k g}{dt^k}(t) = \sum_{i1,i,k=1}^{n} D_{i_k} D_{i_{k-1}} \cdots D_{i_1} f(a+ty)\eta_{i_1}\cdots\eta_{i_k}$.
Bew.: Vollständige Induktion über k:
k=1: $\frac{dg}{dt}(t) = \sum_{i=1}^{n} D_i f(a+ty)\eta_i$ nach Kettenregel 5.12
denn Df(z)=$(D_\eta f(z), ..., D_n f(z))$,
D(a+ty)=y=$(\eta_1, ..., \eta_n)^T$ $k \to k+1$: $\frac{d^{k+1}g}{dt^{k+1}}(t) = \frac{d}{dt}(\sum_{i1,i,k=1}^{n} D_{i_k} D_{i_{k-1}} \cdots D_{i_1} f(a+ty)\eta_{i_1}\cdots\eta_{i_k})$
$=\sum_{j=1}^{n}\sum_{i1,i,k=1}^{n} D_j D_{i_k} D_{i_{k-1}} \cdots D_{i_1} f(a+ty)\eta_{i_1}\cdots\eta_{i_k}\eta_j$, setze $i_{k+1} := j$
$\sum_{i1,i,k=1}^{n} D_{i_k+1} D_{i_k} \cdots D_{i_1} f(a+ty)\eta_{i_1}\cdots\eta_{i_k}\eta_{k+1}$ nach Kettenregel 5.12⌝

Nach dem Satz von Schwarz ist hier beliebiges Umordnen der partiellen Ableitungen möglich. Fassen daher in den $(i_1, ..., i_k) \in \{1, ..., n\}^k$ gleiche Indezes zusammen. Dabei komme der Index j darunter α_j-mal vor ($j \in \{1, ..., n\}$), und wir erhalten zu einem $(i_1, ..., i_k)$ ein bestimmtes $(\alpha_1, ..., \alpha_n)$ mit $\alpha_1 + ... + \alpha_n = k$. Zu diesem $(\alpha_1, ..., \alpha_n)$ gibt es genau $\frac{k!}{\alpha!}$ viele Möglichlkeiten, ein solches passendes $(i_1, ..., i_k)$ zu finden (beweisbar durch vollständige Induktion, s. 7.8).
Daraus folgt
$\frac{dg}{dt}(t) = \sum_{i=1}^{n} \frac{k!}{\alpha!} D_i f(a+ty) y^\alpha$ für $k \geq 1$(k=0: Def. von g)

2. Schritt: Wenden nun auf g den eindimensionalen Satz von Taylor An 19.3, an:

$$f(x) = g(1) = \sum_{j=0}^{m} \frac{g^{(j)}(0)}{j!} + \frac{1}{(m+1)!} g^{m+1}(\vartheta) \text{für } \vartheta \in]0,1[(Lagrange-RestgliedAn19.8) \tag{1}$$

$$(schritt1.) \to = \sum_{j=0}^{m} \frac{1}{j!} \sum_{|\alpha|=j} D^\alpha f(a)(x-a)^\alpha + \frac{1}{(m+1)!} \sum_{|\alpha|=m+1} \frac{(m+1!)}{\alpha!} D^\alpha f(c)(x-a)^\alpha \tag{2}$$

mit $c := a + \vartheta(x-a)$ also $c \in \overline{ax} \backslash \{a, x\}$
$= \sum_{|\alpha|\leq m} \frac{1}{\alpha!} D^\alpha f(a)(x-a)^\alpha + \sum_{|\alpha|=m+1} \frac{1}{\alpha!} D^\alpha f(c)(x-a)^\alpha$.

7.8. Beh.: Zu $(\alpha_1, ..., \alpha_n) \in \mathbb{N}_0^n$ gibt es $\frac{k!}{\alpha!}$ viele k-Tupel $(i_1, ..., i_k) \in \{1, ..., n\}^k$ so, dass ein (vorgeg.) $j \in \{1, ..., n\}$ darin α_j-mal vorkommt mit $\alpha_1 + ... + \alpha_k = k$. ⌜Was sind die 4-Tupel $(i_1, ..., i_4)$ zu $\alpha = (\alpha_1, \alpha_2, \alpha_3) = (1, 3, 0)$? Ü⌝
Bew.: durch vollständige Induktion über k:
k=1: zu (0,...,0,1,0,...,0) (an der Stelle j) gilt es $\frac{k!}{\alpha!} = 1$ einziges k-Tupel (i_1), so, dass j darin α_j-malig vorkommt, nähmlich j. ✓
$k \to k+1$: zu $(\alpha_1, ..., \alpha_n)$ mit $\alpha_1 + ... + \alpha_n = k+1$ gibt es, fallls $i_1 = j$ ist, nach Ind. vor. genau $\frac{k!}{\alpha_1! \cdots (\alpha_j - 1)! \cdots \alpha_n!}$ viele (k+1)-Tupel $(i_1, ..., i_k+1)$ der geforderten Art, insgesamt sind dies:$\sum_{j=1}^{n} \frac{k!}{\alpha_1! \cdots (\alpha_j-1)! \cdots \alpha_n!} = \frac{k!(\alpha_1 + \cdots + \alpha_n)}{\alpha_1! \cdots \alpha_n!} = \frac{(k+1)!}{\alpha!}$ ✓

□

7.9. Bezeichnung: Sei $a \in D \subset \mathbb{R}^n, f : D \to \mathbb{R}$. Sei $\mathcal{U}_a$:=$\{U \subset \mathbb{R}^n; a \in U\}$.

7.10. Def.: f hat in a ein relatives $\left\{ \begin{array}{c} \text{Maximum} \\ \text{Minimum} \end{array} \right\}$:⇔ $\exists U \in \mathcal{U}_a : f_{rU} \left\{ \begin{array}{c} \leq f(a) \\ \geq f(a) \end{array} \right\}$.

Def.: f hat in a ein striktes $\left\{ \begin{array}{c} \text{Maximum} \\ \text{Minimum} \end{array} \right\}$:⇔ $\exists U \in \mathcal{U}_a : f_{rU\backslash\{a\}} \left\{ \begin{array}{c} < f(a) \\ > f(a) \end{array} \right\}$.
Ein relatives Extremum heißt auch lokales Extremum, Globales Extremum, falls U=D wählbar.

7.11. Erste notwendige Bedingung: Sei $f : D \to \mathbb{R}, D \subset \mathbb{R}^n$.
Vor.: f habe in $a \in D$ relatives Extremum, $\forall j \in \{1, ..., n\} : D_j f(a)$ ex.

Beh.: $D_j f(a) = o$ bzw. grad f(a)=O.
Bew.: Sei $\varphi(t) := f(a+te_j)$, für $|t|$ hinreichend klein:
φ hat in O ein relatives Extremum
$\Rightarrow \varphi'(t) = O$
$\Rightarrow D_j f(a) = O$.

□

7.12. Def.: a heißt Kritischer Punkt von f, falls $f \in l^1(D), D \subset \mathbb{R}^n$, und grad f(a)=0 gilt.

7.13. Zweite notwendige Bedingung:
Vor.: f habe in a ein relatives Maximum(/Minimum), $f \in l^2(D), D \subset \mathbb{R}^n$.
Beh.: H(f;a) ist negativ(/positiv) semidefiniert. [Bew. vgl. 7.17]

7.14. Bezeichnung(vgl.Lin Algebra II): A symmetrische Bilinearform, dann heißt A positiv semidefiniert $:\Leftrightarrow \forall h \in \mathbb{R}^n \backslash \{0\} : <h, Ah> = h^T Ah \begin{Bmatrix} \geq 0 \\ \leq 0 \\ > 0 \\ > 0 \end{Bmatrix}$

A positiv definiert:"""""""
1.positiv semidefiniert
2.negativ semidefiniert
3.positiv definiert
4.negativ definiert bei der geschweiften Klammer.

7.15. EQ-Kriterium für Definierbarkeit (vgl. Lineare Algebra *II, Kor.6.5.7, SoSe25*): Sei $A \in \mathbb{R}^{nxn}$ symmetrisch, d.h. $A^T = A$.
Dann gilt:A positiv (negativ) definiert $\Leftrightarrow$ alle EWe >0 (<0).
A positiv (negativ) semidefiniert $\Leftrightarrow$ alle EWe $\geq$0($\leq$0).

7.16. Hauptminoranten-Kriterium (auch: Hurwitz-Kriterium): Sei $A \in \mathbb{R}^{nxn}$ symmetrisch. Dann A positiv definiert $\Leftrightarrow$ alle Hauptminoren det A_k>0, wo A_k die Matrix sei, die aus den ersten k Zeilen und k Spalten von A besteht, $k \in \{1, ..., n\}$.

7.17. Bsp.: $g(t) = f(a + t(x-a))$ habe rel. Maximum in t=0.
Dann $0 \geq g"(t) = h^T H(f;a)h$ mit $h = x - a$ laut 7.13.
$\bullet : D = \mathbb{R}^2, f\begin{pmatrix} x \\ y \end{pmatrix} = y^2 + x^4 + x^3$. Dann ist $D_1 f \begin{pmatrix} x \\ y \end{pmatrix} = 4x^3 + 3x^2, D_2 f \begin{pmatrix} x \\ y \end{pmatrix} = 2y$.
Kritische Punkte: (0,0), $(-\frac{3}{4}, 0)$, sind mögliche Extremstellen laut 7.11.

7.18. Hinreichende Bedingung:
Vor.: $f \in l^2(D), D \subset \mathbb{R}^n$, a Kritischer Punkt in f, H(f;a)$\begin{Bmatrix} negativ \\ positiv \end{Bmatrix}$. definiert.
Beh.: f hat in a striktes lokales $\begin{Bmatrix} Max \\ Min \end{Bmatrix}$..
Bew.: Ans 7.3, dem Satz von Taylor für m=2, folgt:
f(a+h)=f(a)+ $\underbrace{0}_{grad f(a)=\text{oda Krit. Pkt}}$ $+\frac{1}{2}h^T H(f;c)h$ für $c \in \overline{a(a+h)}$ geeignet.
Da H(f;c) stetig und c nahe bei a liegt (wenn h klein ist), ist auch H(f;c) negativ definiert.
Daher ist f(a+h)<f(a), d.h. f hat in a ein striktes lokales Maximum.

Bem.: Dieser Beweisansatz zeigt auch 7.17. $\square$

7.19. Bsp.: $D=\mathbb{R}^2, f\begin{pmatrix}x\\y\end{pmatrix}=y^2+x^4+x^3 \Rightarrow D_2D_1f\begin{pmatrix}x\\y\end{pmatrix}=0, D_1^2f\begin{pmatrix}x\\y\end{pmatrix}=6x(2x+1), D_2^2f\begin{pmatrix}x\\y\end{pmatrix}=2.$
$\Rightarrow H(f;0)=\begin{pmatrix}0&0\\0&2\end{pmatrix}$ positiv semidefiniert,
$H(f;(-\frac{3}{4},0))=\begin{pmatrix}+&0\\0&2\end{pmatrix}$ positiv definiert $\Rightarrow$ in $(-\frac{3}{4})$ ex. ein striktes lok. Min.
Ferner ex. in o kein Extremum: Denn es ist
$f\begin{pmatrix}x\\0\end{pmatrix}=x^3(x+1)\left\{\begin{array}{c}<0=f(o)\\>0=f(o)\end{array}\right\}$ falls $\left\{\begin{array}{c}-1<x<0\\x>0\end{array}\right\}$

7.20. Bem.: Ein Anschluss von Extrema ist wie folgt möglich (Kor. aus 7.13):
Ist f auf $U\subset\mathbb{R}^n$ zweimal stetig diff'bar, grad f(a)=0, H(f;a) indefinit, so hat f in a kein lokales Extremum.

7.21. Praktisches Vorgehen: • Als Extremstellen kommen die Punkte in Frage,
-die Fritisch sind (d.h. grad f verschwindet dort),
-die Randpunkte von U sind (falls U nicht offen sein sollte),
-oder die singulär sind (wo f nicht diff'bar ist).
• Sie Kritischen Punkte ermittelt man durch lösen des Gleichungssystems (i.a. nicht lin.) grad f(x)=0, welches n Gleichungen in n Unbekannt hat. Dieses hat oft nur endlich viele Lösungen. Der Vergleich der Funktionswerte dort reicht aber nicht aus, deshalb sind Kriterien nötig.
• Dann jeweils Hessematrix berechnen und Kriterien testen. Hilft das nicht, müssen die Stellen anderweitig untersucht werden.

7.22. Bsp.: Betr. $f_1,f_2,f_3:\mathbb{R}^2\to\mathbb{R}, f_1(x,y)=x^2+y^4, f_2(x,y)=x^2, f_3(x,y)=x^2+y^3$.
Für alle $i\in\{1,2,3\}$ ist grad $f_i(o)=o$ und $H(f_i;o)=\begin{pmatrix}2&0\\0&2\end{pmatrix}$ positiv semidefinit.
• Die Fkt. f_1 hat in o ein striktes lokales Minimum. $(\forall\begin{pmatrix}x\\y\end{pmatrix}\neq o: f_1(x,y)=x^2+y^2>0=f_1(o).)$
• Die Fkt. f_2 hat in o lokales Minimum $(\forall\begin{pmatrix}x\\y\end{pmatrix}: f_2\begin{pmatrix}x\\y\end{pmatrix}\geq 0)$, das nicht strikt ist, denn in allen Punkten der y-Achse hat f_2 denselben Wert wie in o. $(\forall y: f_2(0;y)=0=f_2(0,0))$
• Die Fkt. f_3 hat in o weder ein lokales Minimum noch lokales Maximum.
$(\forall y<0: f_3(0,y)=y^3<0=f_3(o), \forall y>0: f_3(0,y)=y^3>0=f_3(o).)$