7.1 Lineare Approximation

Die Stetigkeit einer linearen Abbildung $L : V \longrightarrow V'$ eines normierten Raumes $(V, \| \ \|)$ in den normierten Raum $(V', \| \ \|')$ über $\mathbb{R}$ oder $\mathbb{C}$ lässt sich besonders einfach charakterisieren:

Satz 7.1.0.
äq:

(i) *L ist stetig*

(ii) *Es gibt ein $C \geq 0$, so dass $\|L\mathrm{v}\|' \leq C \cdot \|\mathrm{v}\|$ für alle $\mathrm{v} \in V$.*

Folgerung 7.1.1. *Jede lineare Abbildung $\mathbb{R}^N \longrightarrow \mathbb{R}^{N'}$ ist stetig.*

Man kann zeigen, dass

$$\|L\| := \inf\{C \geq 0 \big| \|L\mathrm{v}\|' \leq C \cdot \|\mathrm{v}\| \text{ für alle } \mathrm{v} \in V\}$$

eine Norm auf dem Vektorraum $L(V, V')$ aller stetigen, linearen Abbildungen $V \longrightarrow V'$ ist.

Im folgenden seien sämtliche normierten Räume $(V, \| \ \|)$ Banach[1]-Räume, d.h. $\| \ \|-$vollständig. Jede $\| \ \|$-Cauchy-Folge in V hat also einen $\| \ \|$-Grenzwert in V. Beispielsweise sind $(\mathbb{R}^N, \| \ \|_{1,2,\infty})$ und $(C[0,1], \| \ \|_\infty)$ Banachräume, nicht aber $(C[0,1], \| \ \|_{1,2})$. Ist V' ein Banachraum, so ist auch $L(V, V')$ ein Banachraum.

Definition 7.1.2.

$\overline{x} \in X \subset V$ offen, $f : X \longrightarrow V'$.
f linear approximierbar (oder differenzierbar) in $\overline{x}$: $\Longleftrightarrow$ Es gibt eine stetige, lineare Abbildung $L_{\overline{x}} : V \longrightarrow V'$ sowie eine Abbildung $r : X \longrightarrow V'$, so dass

D1: *$f(x) = f(\overline{x}) + L_{\overline{x}}(x - \overline{x}) + r(x)$ für alle $x \in X$,*

D2: $\dfrac{r(x)}{\|x - \overline{x}\|} \to 0$ *für $x \to \overline{x}, x \neq \overline{x}$.*

Lemma 7.1.3.

Ist f in $\overline{x}$ linear approximierbar, so sind $L_{\overline{x}}$ und r 1-deutig bestimmt.
Die 1-deutig bestimmte stetige, lineare Abbildung

$$Df(\overline{x}) := L_{\overline{x}} \in L(V, V')$$

heißt die Ableitung von f in $\overline{x}$.
Ist $V = V' = \mathbb{R}$, so ist $f'(\overline{x}) = Df(\overline{x})(1) \in \mathbb{R}$.

[1] Stefan Banach (1892 - 1945), polnischer Mathematiker

Beweis der Eindeutigkeit:

Ist $$\begin{aligned} f(x) &= f(\overline{x}) + L(x-\overline{x}) + r(x) \\ &= f(\overline{x}) + \tilde{L}(x-\overline{x}) + \tilde{r}(x), \end{aligned}$$

so ist $(L-\tilde{L})\,(x-\overline{x}) = \tilde{r}(x) - r(x)$, und damit folgt für alle $x \in X$, $x \neq \overline{x}$:

$$(L-\tilde{L})\left(\frac{x-\overline{x}}{\|x-\overline{x}\|}\right) = \frac{\tilde{r}(x)-r(x)}{\|x-\overline{x}\|}.$$

Da X offen ist und die rechte Seite gegen Null konvergiert, ist die lineare Abbildung $L-\tilde{L}=0$ auf der Sphäre $\{\mathrm{v} \in V \big| \|\mathrm{v}\| = 1\}$, also überhaupt. □

Ex:

[1] Eine stetige, lineare Abbildung $L : V \longrightarrow V'$ ist in allen Punkten linear approximierbar, $DL(\overline{x}) = L$.

[2] Eine stetige, bilineare Abbildung $b : V \times V \longrightarrow V'$, $(x,y) \mapsto b(x,y)$, ist in allen Punkten linear approximierbar und es gilt $Db(\overline{x},\overline{y}) = b(\overline{x},.) + b(.,\overline{y})$, d.h. $Db(\overline{x},\overline{y})(x,y) = b(\overline{x},y) + b(x,\overline{y})$ für alle $(x,y) \in V \times V$.

Insbesondere:

$D\det(\overline{x}_1,\cdots,\overline{x}_N)(x_1,\cdots,x_N) = \sum_{k=1}^{N} \det(\overline{x}_1,\cdots,\overline{x}_{k-1},x_k,\overline{x}_{k+1}\cdots,\overline{x}_N)$

Wegen $\|f(x)-f(\overline{x})\|' \leq (\|L\|+1)\|x-\overline{x}\|$ für alle "x nahe $\overline{x}$" (= alle x aus einer geeigneten Umgebung von x), ist eine in $\overline{x}$ differenzierbare Abbildung auch stetig in $\overline{x}$.

Bemerkung 7.1.4.

Sind $f, g : X \longrightarrow V'$ linear approximierbar in $\overline{x} \in X$, $\lambda \in \mathbb{R}$ bzw. $\mathbb{C}$, so sind auch $f+g$ und λf in $\overline{x}$ linear approximierbar, und

$$\begin{aligned} D(f+g)(\overline{x}) &= Df(\overline{x}) + Dg(\overline{x}), \\ D(\lambda f)(\overline{x}) &= \lambda \cdot Df(\overline{x}). \end{aligned}$$

Satz 7.1.5.

$\overline{x} \in X \subset V$ offen, $\overline{y} \in Y \subset V'$ offen, $f : X \longrightarrow Y$ linear approximierbar in $\overline{x}$, $\overline{y} = f(\overline{x})$, $g : Y \longrightarrow V''$ linear approximierbar in $\overline{y}$. Dann ist die Komposition $g \circ f$ linear approximierbar in $\overline{x}$ und

$$D(g \circ f)(\overline{x}) = Dg(\overline{y}) \circ Df(\overline{x}).$$

Beweis : vgl. A I. □

Satz 7.1.6.
$X \subset V$, $Y \subset V'$ offen, $f : X \xrightarrow{\approx} Y$ Homöomorphismus, $\overline{y} = f(\overline{x})$.
äq

(i) *f linear approximierbar in $\overline{x}$, $Df(\overline{x})$ Isomorphismus (mit stetigem Inversen)*

(ii) *f^{-1} linear approximierbar in $\bar{y}$.*

<u>Zusatz</u>: $Df^{-1}(\overline{y}) = Df(\overline{x})^{-1}$.

Beweis : Nach der Kettenregel ist $id = D(id) = D(f^{-1} \circ f) = Df^{-1} \circ Df$, d.h. es genügt, (i) $\Rightarrow$ (ii) zu zeigen. Wir zeigen zunächst, dass

$$m := \inf\{\|Df(\bar{x})\mathrm{v}\|' \mid \mathrm{v} \in V,\ \|\mathrm{v}\| = 1\} > 0.$$

Sonst gäbe es nämlich eine Folge v_n von Vektoren der Norm 1 mit $\lim \|Df(\overline{x})\mathrm{v}_n\| = 0$, also $\mathrm{v}_n = Df(\overline{x})^{-1}(Df(\overline{x})\mathrm{v}_n) \to 0$, ein Widerspruch.
Ansatz: $f^{-1}(y) - f^{-1}(\overline{y}) = Df(\overline{x})^{-1}(y - \overline{y}) + R(y)$. Zum Beweis von (ii) genügt es zu zeigen, dass $\frac{R(y)}{\|y-\overline{y}\|'} \to 0$ für $y \to \overline{y}$, $y \neq \overline{y}$. Wegen $y = f(x)$, $\overline{y} = f(\overline{x})$ ist
$R(y) = x - \overline{x} - Df(\overline{x})^{-1}(y - \overline{y}) \underset{(i)}{=} x - \overline{x} - Df(\overline{x})^{-1}\big(Df(\overline{x})(x - \overline{x}) + r(x)\big) = -Df(\overline{x})^{-1}r(x)$.
Da f homöomorph ist, folgt aus $y \neq \overline{y}$, y nahe $\overline{y}$, auch $x \neq \overline{x}$, x nahe $\overline{x}$, und daher gilt für y hinreichend nahe $\overline{y}$

$$\frac{R(y)}{\|y-\overline{y}\|'} = \frac{-Df(\overline{x})^{-1}r(x)}{\|y-\overline{y}\|'} = -\frac{\|x-\overline{x}\|}{\|y-\overline{y}\|'} \cdot Df(\overline{x})^{-1}\left(\frac{r(x)}{\|x-\overline{x}\|}\right), \text{ wobei}$$

$$\frac{\|y-\overline{y}\|'}{\|x-\overline{x}\|} \underset{(i)}{=} \left\| Df(\overline{x})\left(\frac{x-\overline{x}}{\|x-\overline{x}\|}\right) + \frac{r(x)}{\|x-\overline{x}\|}\right\|' \geq \left\| Df(\overline{x})\left(\frac{x-\overline{x}}{\|x-\overline{x}\|}\right)\right\|' - \left\|\frac{r(x)}{\|x-\overline{x}\|}\right\|'$$

$$\geq m - \frac{\|r(x)\|'}{\|x-\overline{x}\|} \geq \frac{m}{2}, \text{ da } \mathrm{v} := \frac{x-\overline{x}}{\|x-\overline{x}\|} \text{ ein Vektor der Länge 1 ist, folglich}$$

$$\frac{\|R(y)\|}{\|y-\overline{y}\|'} \leq \frac{2}{m}\left\| Df(\overline{x})^{-1}\left(\frac{r(x)}{\|y-\overline{y}\|'}\right)\right\| \leq \frac{2}{m}\|Df(\overline{x})^{-1}\|\frac{\|r(x)\|'}{\|x-\overline{x}\|}.$$

Weiter folgt aus $y \to \overline{y}$, $y \neq \overline{y}$ auch $x \to \overline{x}$, $x \neq \overline{x}$, also $\dfrac{\|r(x)\|'}{\|x-\overline{x}\|} \to 0$ und somit $\dfrac{\|R(y)\|}{\|y-\overline{y}\|'} \to 0$.
□

Ex:

[1] $\overline{x} \in X \subset \mathbb{R}^N$ offen, $y \in Y \subset \mathbb{R}^{N'}$ offen, $f : X \to Y$ stetig, $g : Y \to X$ stetig, so dass

(1) $g \circ f = \mathrm{id}_X$, $f \circ g = \mathrm{id}_Y$,

(2) f linear approximierbar in $\overline{x}$, g linear approximierbar in $\overline{y} = f(\overline{x})$.

Dann ist $N = N'$ (Invarianz der Dimension), da

$$\mathrm{id}_{\mathbb{R}^n}(\overline{x}) = D(\mathrm{id})(\overline{x}) = D(g \circ f)(\overline{x}) = Dg(\overline{y}) \circ Df(\overline{x}),$$

d.h. die lineare Abbildung $Df(\overline{x}) : \mathbb{R}^N \longrightarrow \mathbb{R}^{N'}$ ist invertierbar, also $N = N'$. Beachte: $\big(Df(\overline{x})\big)^{-1}$ existiert genau dann, wenn $\det Df(\overline{x}) \neq 0$.

[2] $\overline{x} \in X \subset \mathbb{R}^N$ offen, $\mathrm{v} \in \mathbb{R}^N$, $\|\mathrm{v}\|_2 = 1$, $f : X \to \mathbb{R}$ linear approximierbar in $\overline{x}$. Für $0 < \varepsilon \ll 1$ verläuft die Kurve

$$c : (-\varepsilon, \varepsilon) \longrightarrow X, \quad t \mapsto c(t) := \overline{x} + t \cdot \mathrm{v}$$

ganz in X, und für den Tangentialvektor in $\overline{x}$ gilt

$$\dot{c}(0) = Dc(0)(1) = \mathrm{v}.$$

Da $Df(\overline{x}) : \mathbb{R}^N \longrightarrow \mathbb{R}$ ein lineares Funktional ist, gibt es einen 1-deutig bestimmten Vektor $\mathrm{grad} f(\overline{x}) \in \mathbb{R}^N$, so dass

$$Df(\overline{x})(w) = (\mathrm{grad} f(\overline{x}) | w)$$

für alle $w \in \mathbb{R}^N$.

$$D_{\mathrm{v}} f(\overline{x}) := \lim_{t \to 0} \frac{1}{t}(f(\overline{x} + t\mathrm{v}) - f(\overline{x}))$$

ist die sogenannte **Richtungsableitung** von f in $\overline{x}$ längs der Kurve c. Die Kettenregel liefert:

$$D_{\mathrm{v}} f(\overline{x}) = \frac{d}{dt} f \circ c(t)\big|_{t=0} = D(f \circ c)(0)(1) = \big(Df(\overline{x}) \circ Dc(0)\big)(1) = Df(\overline{x})(\mathrm{v}) = (\mathrm{grad} f(\overline{x}) | \mathrm{v})$$

und wegen der Cauchy-Schwarzschen Ungleichung ist die Richtungsableitung dem Betrag nach

$$|D_{\mathrm{v}} f(\overline{x})| \leq \|\mathrm{grad} f(\overline{x})\|_2 \cdot \|\mathrm{v}\|_2 = \|\mathrm{grad} f(\overline{x})\|_2.$$

$D_{\mathrm{v}} f(\overline{x})$ ist also maximal genau dann, wenn

$$\mathrm{v} = \pm \frac{\mathrm{grad} f(\overline{x})}{\|\mathrm{grad} f(\overline{x})\|_2},$$

d.h. der Gradient gibt die Richtung des größten Anstiegs von f an der Stelle $\overline{x}$ an.

Folgerung 7.1.7. $\overline{x} \in X$ *relatives Extremum von* $f \Longrightarrow \mathrm{grad} f(\overline{x}) = 0$.

Beweis : Für alle $\|\mathrm{v}\|_2 = 1$ ist $D_{\mathrm{v}} f(x) = (\mathrm{grad} f(\overline{x}) | \mathrm{v}) = 0$, d.h. $\mathrm{grad} f(\overline{x}) \perp \mathbb{R}^N$. □

Theorem 7.1.8.

$a < b$, $f : [a,b] \to V$ *stetig,* $g : [a,b] \to \mathbb{R}$ *stetig,* $A \subset [a,b]$ *höchstens abzählbar, so dass*

(1) f, g *differenzierbar in* $[a,b] \setminus A$

(2) $\|Df(t)\| \leq g'(t)$ *für alle* $t \notin A$.

Dann ist $\|f(b) - f(a)\| \leq g(b) - g(a)$.

Beweis : Vgl. A I. □

Folgerung 7.1.9.

Ist $\|Df(\overline{x})\| \leq M$ *in* $[a,b] \setminus A$*, so ist* $\|f(b) - f(a)\| \leq M(b-a)$.
Insbesondere ist $f = const$*, falls* $Df\big|_{[a,b]\setminus A} = 0$.

Folgerung 7.1.10.

Sei $\overline{x} \in X \subset V$ *offen,* $F : X \to V$ *stetig,* $\mathrm{v} \in V$ *so dass die Strecke* $S := \{\overline{x} + t \cdot \mathrm{v} | t \in [0,1]\}$ *ganz in* X *liegt.*
Ist F *differenzierbar in allen* $s \in S$ *und gilt* $\|DF(s)\| \leq C$ *für alle* $s \in S$*, dann ist*

$$\|F(\overline{x} + \mathrm{v}) - F(\overline{x})\| \leq \|\mathrm{v}\| \cdot \sup_{s\in S} \|DF(s)\|.$$

Beweis : Sei $f : [0,1] \longrightarrow V$, $f(t) := F(\overline{x} + t\mathrm{v})$. Dann ist $Df(t) = DF(\overline{x} + t\mathrm{v})(\mathrm{v})$, also

$$\|Df(t)\| \leq \|DF(\overline{x} + t\mathrm{v})\| \cdot \|\mathrm{v}\| \leq \|\mathrm{v}\| \cdot \sup_{s\in S} \|DF(s)\|.$$

□

Sind $(E_1, \| \ \|_1)$ bzw. $(E_2, \| \ \|_2)$ Banach-Räume, so ist $E_1 \times E_2$ ein Banach-Raum bzgl. der Supremumsnorm

$$\|(x,y)\|_\infty = \sup\{\|x\|_1, \|y\|_2\}$$

Ist $X \subset E_1 \times E_2$ offen und $f : X \longrightarrow V'$ linear approximierbar in $(\overline{x}, \overline{y}) \in X$, so leben die partiellen Abbildungen $x \mapsto f(x, \overline{y})$ bzw. $y \mapsto f(\overline{x}, y)$ in offenen Umgebungen von $\overline{x} \in E_1$ bzw. $\overline{y} \in E_2$ und sind in $\overline{x}$ bzw. $\overline{y}$ linear approximierbar mit Ableitungen

$$D_1 f(\overline{x}, \overline{y}) : E_1 \longrightarrow V' \text{ bzw. } D_2 f(\overline{x}, \overline{y}) : E_2 \longrightarrow V',$$

den sogenannten **partiellen Ableitungen** von f. Es gilt

$$Df(\overline{x}, \overline{y})(x,y) = D_1 f(\overline{x}, \overline{y})(x) + D_2 f(\overline{x}, \overline{y})(y) \text{ für alle } (x,y) \in E_1 \times E_2.$$

Umgekehrt ist aber f im allgemeinen nicht linear approximierbar in $(\overline{x}, \overline{y})$, falls in diesem Punkt die partiellen Ableitungen von f existieren.

WARNUNG: Die Funktion $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$,

$$f(x,y) = \begin{cases} \dfrac{2xy^2}{x^2+y^2} & x^2 + y^2 > 0 \\ 0 & \text{sonst} \end{cases}$$

ist in 0 partiell ableitbar aber nicht linear approximierbar.

Beweis : Da $f(x,0) = f(0,y) = 0$ für alle x bzw. y in $\mathbb{R}$, ist nämlich $D_1 f(0) = D_2 f(0) = 0$, also, – falls f linear approximierbar wäre –, $Df(0) = 0$. Für die Richtungsableitung $D_\mathrm{v} f(0)$ längs $\mathrm{v} = (1,1)$ folgt aber dann $0 = Df(0)(\mathrm{v}) = D_\mathrm{v} f(0) = \lim_{t\to 0} \dfrac{f(t\mathrm{v})}{t} = 1$, ein Widerspruch.□

Lemma 7.1.11.
Ist f partiell ableitbar in einer offenen Umgebung U von $\overline{u} = (\overline{x}, \overline{y}) \in E_1 \times E_2$ und $C \geq 0$ derart, dass $\|D_1 f(u)\| \leq C$ und $\|D_2 f(u)\| \leq C$ für alle $u \in U$, dann ist f stetig in $\overline{u}$.

Beweis : Wähle zu $\varepsilon > 0$ ein $0 < \delta < \frac{\varepsilon}{2C+1}$, so dass $\overline{u} + h \in U$ für alle $h \in B_\delta(0) \times B_\delta(0)$. Dann ist $f(\overline{u}+h) - f(\overline{u}) = f(\overline{u} + (h_1, h_2)) - f(\overline{u} + (h_1, 0)) + f(\overline{u} + (h_1, 0)) - f(\overline{u})$, also nach dem MWS $\|f(\overline{u}+h) - f(\overline{u})\|' \leq C\|h_1\| + C\|h_2\| \leq 2C \cdot \delta < \varepsilon$. □

Satz 7.1.12.

$(\overline{x}, \overline{y}) \in X \subset E_1 \times E_2$ *offen,* $f : X \longrightarrow V'$.

(1) *Ist f linear approximierbar in $(\overline{x}, \overline{y})$, so ist f partiell ableitbar in $(\overline{x}, \overline{y})$.*

(2) *Ist f partiell ableitbar in einer Umgebung von $(\overline{x}, \overline{y})$ und sind die partiellen Ableitungen $(x, y) \mapsto D_1 f(x, y)$, $(x, y) \mapsto D_2 f(x, y)$ stetig in $(\overline{x}, \overline{y})$, so ist f linear approximierbar in $(\overline{x}, \overline{y})$.*

Zusatz: $Df(\overline{x}, \overline{y})(x, y) = D_1 f(\overline{x}, \overline{y})(x) + D_2 f(\overline{x}, \overline{y})(y)$.

Beweis : Mit den obigen Bezeichnungen ist

$$f(\overline{u}+h) - f(\overline{u}) = \Big(f\big(\overline{u} + (h_1, 0)\big) - f(\overline{u})\Big) + \Big(f\big(\overline{u} + (h_1, h_2)\big) - f\big(\overline{u} + (h_1, 0)\big)\Big).$$

Dann gibt es zu $\varepsilon > 0$ ein $0 < \delta$, so dass zunächst

$$\big\|f\big(\overline{u} + (h_1, 0)\big) - f(\overline{u}) - D_1 f(\overline{u})(h_1)\big\|' \leq \varepsilon\|h_1\| \text{ für alle } \|h_1\| \leq \delta.$$

Wendet man den MWS auf die Abbildung $g(h_2) := f\big(\overline{u} + (h_1, h_2)\big) - D_2 f\big(\overline{u} + (h_1, 0)(h_2)\big)$ an, so ist

$$\Big\|f\big(\overline{u} + (h_1, h_2)\big) - f\big(\overline{u} + (h_1, 0)\big) - D_2 f\big(\overline{u} + (h_1, 0)\big)(h_2)\Big\|' \\ \leq \|h_2\| \sup_{\|z\| \leq \|h_2\|} \Big\|D_2 f\big(\overline{u} + (h_1, z)\big) - D_2 f\big(\overline{u} + (h_1, 0)\big)\Big\|.$$

Wegen der Stetigkeit der partiellen Ableitungen in $\overline{u}$ gibt es ein $0 < \eta \leq \delta$, so dass das obige Supremum $\leq \varepsilon$ für alle $\|h\|_\infty < \eta$. Dann ist aber

$$\big\|f(\overline{u}+h) - f(\overline{u}) - D_1 f(\overline{u})(h_1) - D_2 f(\overline{u})(h_2)\big\|' \leq 4\varepsilon \cdot \|h\|_\infty.$$

□

In naheliegender Weise lassen sich die vorangegangenen Überlegungen auf Abbildungen f N-facher Produkte $E_1 \times \ldots \times E_N$ von Banach-Räumen übertragen:

$$Df(\overline{x})(x_1, \ldots, x_N) = \sum_{k=1}^{N} D_k f(\overline{x})(x_k).$$

Ex: $D \det(\overline{x}_1, \ldots, \overline{x}_n)(x_1, \ldots, x_n) = \sum_{k=1}^N \det(\overline{x}_1, \ldots, \overline{x}_{k-1}, x_k, \overline{x}_{k+1}, \ldots, \overline{x}_n)$, d.h. $D_k(\overline{x}_1, \ldots, \overline{x}_n)(x_k) = det(\overline{x}_1, \ldots, \overline{x}_{k-1}, x_k, \overline{x}_{k+1}, \ldots, \overline{x}_n)$.

Ist insbesondere $f : X \longrightarrow \mathbb{R}^M$ eine Abbildung eines offenen Teiles $X \subset \mathbb{R}^N$ und ist f linear approximierbar bzw. existieren die partiellen Ableitungen und sind stetig, ist weiterhin $p_i : \mathbb{R}^M \longrightarrow \mathbb{R}$ die Projektion $(y_1, \cdots, y_M) \mapsto y_i$ auf die i-te Koordinate, $f_i := p_i \circ f$, d.h. $f = (f_1, \ldots, f_M)$ und $e_k = (0, \ldots, 1, \ldots, 0)$ der k-te Einheitsvektor, so ist

$$D(p_1 \circ f)(\overline{x})(e_k) = p_i \circ Df(\overline{x})(e_k) = p_i \circ D_k f(\overline{x})(1) = D_k f_i(\overline{x})(1) = \lim_{t \mapsto 0} \frac{f_i(\overline{x} + te_k) - f_i(\overline{x})}{t}$$

wofür man üblicherweise

$$\frac{\partial f_i}{\partial x_k}(\overline{x}) := D_k f_i(\overline{x})(1)$$

schreibt. Beachte, dass die 'klassische' partielle Ableitung $\frac{\partial f_i}{\partial x_k}(\overline{x})$ gerade die Richtungsableitung der Funktion f_i in Richtung des k-ten Einheitsvektors ist. Die $M \times N$-Matrix

$$Jf(\overline{x}) := \left(\frac{\partial f_i}{\partial x_k}(\overline{x})\right) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\overline{x}) & \ldots & \frac{\partial f_1}{\partial x_N}(\overline{x}) \\ \vdots & & \vdots \\ \frac{\partial f_M}{\partial x_1}(\overline{x}) & \ldots & \frac{\partial f_M}{\partial x_N}(\overline{x}) \end{pmatrix}$$

der partiellen Ableitungen ist die sogenannte Jacobi[2]-Matrix von f in $\overline{x}$.

Folgerung 7.1.13.

$$Df(\overline{x})(x) = Jf(\overline{x}) \cdot x$$

d.h. $Jf(\overline{x})$ ist die Matrix der linearen Abbildung $Df(\overline{x})$ bzgl. der kanonischen Basis. Insbesondere ist der Gradient einer Funktion $g : X \longrightarrow \mathbb{R}$ der Vektor

$$\text{grad } g(\overline{x}) = \left(\frac{\partial g}{\partial x_1}(\overline{x}), \ldots, \frac{\partial g}{\partial x_N}(\overline{x})\right).$$

Satz 7.1.14. (Leibniz[3]-Regel)

$X \subset \mathbb{R}^N$ offen, $a < b$, $f : X \times [a, b] \longrightarrow \mathbb{R}$ stetig, $F : X \longrightarrow \mathbb{R}$, $x \mapsto F(x) := \int_a^b f(x, t)dt$, so dass sämtliche partiellen Ableitungen

$$\frac{\partial f}{\partial x_i} : X \times [a, b] \longrightarrow \mathbb{R},\ i = 1, \ldots, N$$

existieren und stetig sind.
Dann ist auch F stetig nach x_i partiell ableitbar, und es gilt

$$\frac{\partial F}{\partial x_i}(x) = \int_a^b \frac{\partial f}{\partial x_i}(x, t)dt.$$

Beweis : Œ $N = 1$. $x \in X$ fest, $0 \neq h \in \mathbb{R}$, $|h| \ll 1$.
Definiere

$$g(h, t) := \begin{cases} \dfrac{\partial f}{\partial x}(x, t) & h = 0 \\ \frac{1}{h}\big(f(x + h, t) - f(x, t)\big) & h \neq 0. \end{cases}$$

[2] Carl Gustav Jacob Jacobi (1804 - 1851), deutscher Mathematiker

[3] Gottfried Wilhelm Leibniz (1646 - 1716) deutscher Philosoph, Mathematiker und Naturforscher

Nach dem MWS ist $(h,t) \mapsto g(h,t)$ stetig und damit auch $h \mapsto \int_a^b g(h,t)dt$, also

$$\frac{\partial F}{\partial x}(x) = \lim_{\substack{h\to 0\\ h\neq 0}} \int_a^b g(h,t)dt = \int_a^b g(0,t)dt = \int_a^b \frac{\partial f}{\partial x}(x,t)dt.$$

□

Satz 7.1.15.

Ist f 2-mal stetig partiell ableitbar, so ist

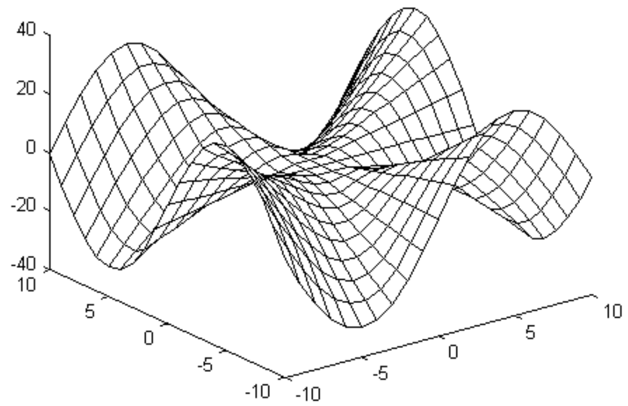
$$\frac{\partial}{\partial x_i}\,\frac{\partial}{\partial x_k} f = \frac{\partial}{\partial x_k}\,\frac{\partial}{\partial x_i} f,$$

d.h. die partiellen Ableitungen sind vertauschbar.

Ex: Die Funktion $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$,

$$f(x,y) := \begin{cases} xy \cdot \frac{x^2-y^2}{x^2+y^2}, & x^2+y^2 > 0 \\ 0 & \text{sonst} \end{cases}$$

ist 2-mal partiell ableitbar, aber nicht stetig partiell ableitbar, da $f_{xy}(0,0) = -1$, aber $f_{yx}(0,0) = 1$, wobei $f_x := \frac{\partial}{\partial x} f$.



Folgerung 7.1.16.

Ist f 2-mal stetig partiell ableitbar, so ist die **Hesse**[4]**-Matrix**

$$H(f)(x) := \left(\frac{\partial}{\partial x_i}\,\frac{\partial}{\partial x_k} f(x)\right)$$

der 2-ten partiellen Ableitungen symmetrisch, d.h. ${}^tH(f) = H(f)$.

[4]Ludwig Otto Hesse (1811 - 1874), deutscher Mathematiker

Beweis des Satzes: Œ $N = 2$. $a, b \in \mathbb{R}$ fest, $x \neq a$, $y \neq b$,

$$H(x,y) := (y-b)^{-1} \cdot \frac{\big((f(x,y) - f(x,b)\big) - \big(f(a,y) - f(a,b)\big)}{x-a}.$$

Für festes y liefert der MWS, angewandt auf $x \mapsto f(x,y) - f(x,b)$, ein $\xi = \xi_y$ zwischen a und x, so dass

$$H(x,y) = (y-b)^{-1} \left[\frac{\partial f}{\partial x}(\xi, y) - \frac{\partial f}{\partial x}(\xi, b)\right]$$

Wendet man jetzt den MWS auf $y \mapsto \frac{\partial f}{\partial x}(\xi, y)$ an, so gibt es ein η zwischen y und b, so dass $H(x,y) = \dfrac{\partial}{\partial y}\dfrac{\partial}{\partial x} f(\xi, \eta)$. Analog folgt $H(x,y) = \dfrac{\partial}{\partial x}\dfrac{\partial}{\partial y} f(\xi', \eta')$ und damit aufgrund der Stetigkeit $f_{xy} = f_{yx}$. □

Bezeichnung 7.1.17.

Für eine hinreichend oft stetig partiell ableitbare Funktion $f : X \longrightarrow \mathbb{R}$ auf einem offenen Teil $X \subset \mathbb{R}^N$, einen 'Multi-Index' $\alpha = (\alpha_1, \dots, \alpha_N) \in \mathbb{N}^N$ und $x \in X$ sei

$$\begin{aligned}
|\alpha| \ &:= \alpha_1 + \ldots + \alpha_N = \|\alpha\|_1 \\
\alpha! &:= \prod \alpha_i! \\
\alpha \leq \beta &:\Longleftrightarrow \alpha_i \leq \beta_i \textit{ für alle } i \\
x^\alpha &:= \prod x_i^{\alpha_i} \\
D^\alpha f &:= \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdot \ldots \cdot \left(\frac{\partial}{\partial x_N}\right)^{\alpha_N} f, \ \textit{und} \\
j_{\overline{x}}^k f(x) &:= \sum_{|\alpha| \leq k} \frac{D^\alpha f(\overline{x})}{\alpha!} (x - \overline{x})^\alpha
\end{aligned}$$

der **k-Jet** *oder das* **k-te Taylorpolynom**[5] *von f in $\overline{x} \in X$ (vgl. A I).*
Beachte, dass wegen der stetigen partiellen Differenzierbarkeit

$$D^\alpha f = D^{\alpha_\pi} f$$

für alle Permutationen π, $\alpha_\pi = \big(\alpha_{\pi(1)}, \dots, \alpha_{\pi(N)}\big)$, und $|x^\alpha| \leq \|x\|_\infty^{|\alpha|}$.

[5]Brook Taylor (1685 - 1731), englischer Jurist, veröffentlichte als Privatgelehrter mathematische, physikalische, philosophische und religiöse Schriften

Satz 7.1.18. (Taylor-Formel)

$f : X \longrightarrow \mathbb{R}$ $(k+1)$-mal stetig partiell ableitbar, $x, \overline{x} \in X$ so dass

$$S := \{\lambda x + (1-\lambda)\overline{x} \,|\, \lambda \in [0,1]\} \subset X.$$

Dann gibt es ein $\tau \in (0,1)$ so dass

$$f(x) = j_{\overline{x}}^k f(x) + \sum_{|\alpha|=k+1} \frac{1}{\alpha!} D^\alpha f(\tau x + (1-\tau)\overline{x}) \cdot (x-\overline{x})^\alpha.$$

Folgerung 7.1.19.

$f(x) = j_{\overline{x}}^{k+1} f(x) + \varepsilon(x)$, $\dfrac{\varepsilon(x)}{\|x-\overline{x}\|^{k+1}} \longrightarrow 0$ für $x \to \overline{x}$, $x \neq \overline{x}$.

Ex: f $(k+1)$-mal stetig partiell ableitbar, $D^\alpha f = 0$ für alle $|\alpha| = k+1$. Dann ist f ein Polynom in $x_1, \ldots, x_N$ vom Grade $\leq k$:

$$f(x_1, \ldots, x_N) = \sum_{\substack{k_1+\ldots+k_N \leq k \\ k_1, \ldots, k_N \geq 0}} a_{k_1 \ldots k_N} x_1^{k_1} \cdot \ldots \cdot x_N^{k_N}$$

Beweis : Anwendung der 1-dimensionalen Taylorformel auf die Funktion $\psi : [0,1] \longrightarrow \mathbb{R}$, $t \mapsto \psi(t) := f(\overline{x} + t(x-\overline{x}))$ liefert die Aussage. □

Satz 7.1.20.

$\overline{x} \in X \subset \mathbb{R}^N$ offen, $f : X \longrightarrow \mathbb{R}$ 2-mal stetig partiell ableitbar.

(1) *Ist $\overline{x}$ ein lokales Minimum von f, so ist* $\operatorname{grad} f(\overline{x}) = 0$ *und $H(f)(\overline{x})$ positiv-semidefinit.*

(2) *Ist* $\operatorname{grad} f(\overline{x}) = 0$ *und $H(f)(\overline{x})$ positiv-definit, so ist $\overline{x}$ ein lokales Minimum von f.*

Bemerkung 7.1.21.

Für eine symmetrische Matrix $A \in M_N(\mathbb{R})$ sind folgende Aussagen äquivalent:

(i) *A ist positiv definit,*

(ii) *$(Ax|x) > 0$ für alle $0 \neq x \in \mathbb{R}^N$,*

(iii) *Es gibt eine Basis aus Eigenvektoren mit positiven Eigenwerten,*

(iv) *Es gibt eine Matrix $B \in M_N(\mathbb{R})$, so dass $A = {}^tBB$ und* $\det B \neq 0$,

(v) *det* $\begin{pmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \ldots & a_{kk} \end{pmatrix}$ *>0 für $k = 1, \ldots, N$.*

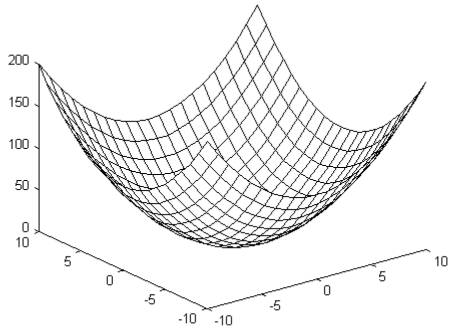
Beweis : $0 \neq h := x - \overline{x}$. Dann ist $f(x) - f(\overline{x}) = (\mathrm{grad} f(\overline{x})|h) + \frac{1}{2}(Hf(\overline{x})h|h) + \varepsilon(x)$, also

$$\frac{f(x) - f(\overline{x})}{\|h\|^2} = \frac{1}{|h\|^2}(\mathrm{grad} f(\overline{x})|h) + \frac{1}{2}\left(Hf(\overline{x})\frac{h}{\|h\|}\Big|\frac{h}{\|h\|}\right) + \frac{\varepsilon(x)}{\|h\|^2}.$$

□

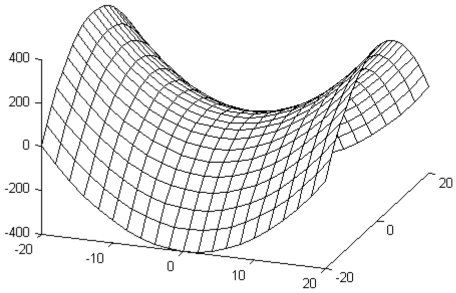
Ex:

[1] $f(x,y) := x^2 + y^2$ hat in 0 ein lokales Minimum, da $\mathrm{grad} f(0) = 0, \quad Hf(0) = 2E_2$.



Graph von f

[2] Die Funktion $g(x,y) := x^2 - y^2$ hat in 0 einen sogenannten Sattelpunkt, d.h. $Hg(0)$ ist indefinit, aber nicht ausgeartet.



Graph von g

[3] $0 \neq a \in \mathbb{R}^N$, $f : B^N \longrightarrow \mathbb{R}$, $x \mapsto (a|x) + \lambda(x), \quad \lambda(x) = \sqrt{(1 - x_1^2 - x_2^2 - \ldots - x_N^2)}$.

Für $\|x\|_2 < 1$ ist $\mathrm{grad} f(x) = a - \dfrac{x}{\lambda(x)}$ und $-(Hf(x)\mathrm{v}|\mathrm{v}) = \dfrac{1}{\lambda(x)}\left[\left(\dfrac{(x|\mathrm{v})}{\lambda(x)}\right)^2 + (\mathrm{v}|\mathrm{v})\right]$,

d.h. $Hf(x)$ ist stets negativ definit.

Daher ist $\overline{x} := \dfrac{a}{\sqrt{(1 + \|a\|_2^2)}}$ ein relatives Maximum von f, $f(\overline{x}) = \sqrt{(1 + \|a\|_2^2)} > \|a\|_2$.

Auf der $(N-1)$-Sphäre S^{N-1} ist $|f(x)| = |(a|x)| \leq \|a\|_2$, also ist $\overline{x}$ sogar ein absolutes Maximum von f auf B^N. Da der Gradient nur in $\overline{x}$ verschwindet, muss das Minimum auf S^{N-1} angenommen werden:

$$-\|a\|_2 \leq f(x) \leq \|a\|_2 \text{ auf } S^{N-1}.$$

Für $x_* := -\dfrac{a}{\|a\|_2}$ ist $f(x_*) = -\|a\|_2$, d.h. x_* ist das absolute Minimum von f.