Allgemeiner Mittelwert-Abschätzungssatz.

Satz (Mittelwert-Abschätzungssatz)**:** *Die Funktion* $\mathbf{f} : D \to \mathbb{R}^m$ *sei differenzierbar auf der offenen Menge* $D \subset \mathbb{R}^n$*. Weiterhin seien* $\mathbf{a}, \mathbf{b}$ *Punkte in* D *mit* $[\mathbf{a}, \mathbf{b}] \subset D$*. Dann gilt für jede Vektornorm* $\|\cdot\|$ *eine Abschätzung der Form*

$$\|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})\| \leq \sup_{\xi \in [\mathbf{a},\mathbf{b}]} \|\mathbf{J}\,\mathbf{f}(\xi)\| \cdot \|\mathbf{b} - \mathbf{a}\|.$$

Beweisidee: Wie im vorigen Satz, jedoch mit der Integraldarstellung

$$(\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})) = \left(\int_0^1 \mathbf{Jf}(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))\, dt \right) \cdot (\mathbf{b} - \mathbf{a})$$

und der Standardabschätzung

$$\left\| \int_0^1 \mathbf{A}(t)\, dt \right\| \leq \int_0^1 \|\mathbf{A}(t)\|\, dt \qquad \text{für } \mathbf{A}(t) \in \mathbb{R}^{m \times n}, \quad t \in [0, 1].$$

□

Taylor-Entwicklungen: Notationen.

Zunächst definieren wir einen Multiindex $\alpha \in \mathbb{N}_0^n$ als

$$\alpha := (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$$

Weiterhin sei

$$|\alpha| := \alpha_1 + \ldots + \alpha_n \qquad \text{und} \qquad \alpha! := \alpha_1! \cdot \ldots \cdot \alpha_n!$$

Ist $f : D \to \mathbb{R}$ $|\alpha|$-mal stetig differenzierbar, so setzen wir

$$D^\alpha = D_1^{\alpha_1} D_2^{\alpha_2} \ldots D_n^{\alpha_n} = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}},$$

wobei $D_i^{\alpha_i} = \underbrace{D_i \ldots D_i}_{\alpha_i\text{-mal}}$, und wir schreiben

$$\mathbf{x}^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} \qquad \text{für } \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Der Satz von Taylor.

Satz (Satz von Taylor): *Sei $D \subset \mathbb{R}^n$ offen und konvex, $f : D \to \mathbb{R}$ eine C^{m+1}-Funktion, und sei $\mathbf{x}^0 \in D$. Dann gilt für $\mathbf{x} \in D$ die* Taylor-Entwicklung

$$
\begin{aligned}
f(\mathbf{x}) &= T_m(\mathbf{x};\mathbf{x}^0) + R_m(\mathbf{x};\mathbf{x}^0) \\
T_m(\mathbf{x};\mathbf{x}^0) &= \sum_{|\alpha|\leq m} \frac{D^\alpha f(\mathbf{x}^0)}{\alpha!}(\mathbf{x}-\mathbf{x}^0)^\alpha \\
R_m(\mathbf{x};\mathbf{x}^0) &= \sum_{|\alpha|=m+1} \frac{D^\alpha f(\mathbf{x}^0+\theta(\mathbf{x}-\mathbf{x}^0))}{\alpha!}(\mathbf{x}-\mathbf{x}^0)^\alpha
\end{aligned}
$$

mit einem geeigneten $\theta \in (0,1)$. □

Definition: *In der obigen Taylor-Entwicklung heißt $T_m(\mathbf{x};\mathbf{x}^0)$* Taylor-Polynom m-ten Grades *und $R_m(\mathbf{x};\mathbf{x}^0)$ wird als* Lagrange-Restglied *bezeichnet.* □

Herleitung der Taylorschen Formel.

Wir definieren eine skalare Funktion einer Variablen $t \in [0, 1]$ als

$$g(t) := f(\mathbf{x}^0 + t(\mathbf{x} - \mathbf{x}^0))$$

und berechnen die Taylor-Entwicklung um $t = 0$. Es gilt:

$$g(1) = g(0) + g'(0) \cdot (1 - 0) + \frac{1}{2} g''(\xi) \cdot (1 - 0)^2 \qquad \text{für ein } \xi \in (0, 1).$$

Die Berechnung von $g'(0)$ liefert mit der Kettenregel

$$\begin{aligned} g'(0) &= \frac{d}{dt} f(x_1^0 + t(x_1 - x_1^0), x_2^0 + t(x_2 - x_2^0), \dots, x_n^0 + t(x_n - x_n^0))\Big|_{t=0} \\ &= D_1 f(\mathbf{x}^0)(x_1 - x_1^0) + \dots + D_n f(\mathbf{x}^0)(x_n - x_n^0) \\ &= \sum_{|\alpha|=1} \frac{D^\alpha f(\mathbf{x}^0)}{\alpha!} (\mathbf{x} - \mathbf{x}^0)^\alpha. \end{aligned}$$

Fortsetzung der Herleitung.

Berechnung von $g''(0)$ liefert

$$
\begin{aligned}
g''(0) &= \frac{d^2}{dt^2} f(\mathbf{x}^0 + t(\mathbf{x} - \mathbf{x}^0))\Big|_{t=0} \\
&= \frac{d}{dt} \sum_{k=1}^{n} D_k f(\mathbf{x}^0 + t(\mathbf{x} - \mathbf{x}^0))(x_k - x_k^0)\Bigg|_{t=0} \\
&= D_{11} f(\mathbf{x}^0)(x_1 - x_1^0)^2 + D_{21} f(\mathbf{x}^0)(x_1 - x_1^0)(x_2 - x_2^0) \\
&\qquad + \ldots + D_{ij} f(\mathbf{x}^0)(x_i - x_i^0)(x_j - x_j^0) + \ldots + \\
&\quad D_{n-1,n} f(\mathbf{x}^0)(x_{n-1} - x_{n-1}^0)(x_n - x_n^0) + D_{nn} f(\mathbf{x}^0)(x_n - x_n^0)^2 \\
&= \sum_{|\alpha|=2} \frac{D^\alpha f(\mathbf{x}^0)}{\alpha!} (\mathbf{x} - \mathbf{x}^0)^\alpha \qquad \text{(Vertauschungssatz von Schwarz!)}
\end{aligned}
$$

Nun: Beweis der Taylor-Formel mittels vollständiger Induktion! □

Beweis des Satzes von Taylor.

Die Funktion $g(t) := f(\mathbf{x}^0 + t(\mathbf{x} - \mathbf{x}^0))$ ist $(m+1)$-mal stetig differenzierbar, und es gilt

$$g(1) = \sum_{k=0}^{m} \frac{g^{(k)}(0)}{k!} + \frac{g^{(m+1)}(\theta)}{(m+1)!} \qquad \text{für ein } \theta \in [0,1]$$

Weiterhin gilt (per Induktion über k)

$$\frac{g^{(k)}(0)}{k!} = \sum_{|\alpha|=k} \frac{D^\alpha f(\mathbf{x}^0)}{\alpha!} (\mathbf{x} - \mathbf{x}^0)^\alpha$$

und

$$\frac{g^{(m+1)}(\theta)}{(m+1)!} = \sum_{|\alpha|=m+1} \frac{D^\alpha f(\mathbf{x}^0 + \theta(\mathbf{x} - \mathbf{x}^0))}{\alpha!} (\mathbf{x} - \mathbf{x}^0)^\alpha.$$

$\blacksquare$

Beispiel.

Berechne das Taylor-Polynom $T_2(\mathbf{x};\mathbf{x}^0)$ zweiten Grades der Funktion

$$f(x,y,z) = x\,y^2 \sin z$$

zum Entwicklungspunkt $(x,y,z) = (1,2,0)^T$.

Berechnung von $T_2(\mathbf{x};\mathbf{x}^0)$ benötigt partielle Ableitungen bis zur 2. Ordnung.

Diese Ableitungen müssen am Punkt $(x,y,z) = (1,2,0)^T$ ausgewertet werden.

Als Ergebnis erhält man $T_2(\mathbf{x};\mathbf{x}^0)$ in der Form

$$T_2(\mathbf{x};\mathbf{x}^0) = 4z(x+y-2)$$

$\square$

Bemerkung. Das Restglied eines Taylor-Polynoms enthält **alle** partiellen Ableitungen der Ordnung $(m+1)$:

$$R_m(\mathbf{x};\mathbf{x}^0) = \sum_{|\alpha|=m+1} \frac{D^\alpha f(\mathbf{x}_0 + \theta(\mathbf{x}-\mathbf{x}^0))}{\alpha!}(\mathbf{x}-\mathbf{x}^0)^\alpha$$

Sind alle diese Ableitungen in der Nähe von $\mathbf{x}^0$ beschränkt durch eine Konstante C, so gilt die Restgliedabschätzung

$$|R_m(\mathbf{x};\mathbf{x}^0)| \leq \frac{n^{m+1}}{(m+1)!} C \, \|\mathbf{x}-\mathbf{x}^0\|_\infty^{m+1}$$

Für die Approximationsgüte des Taylor-Polynoms einer C^{m+1}-Funktion folgt

$$f(\mathbf{x}) = T_m(\mathbf{x};\mathbf{x}^0) + \mathcal{O}\left(\|\mathbf{x}-\mathbf{x}^0\|^{m+1}\right).$$

Spezialfall $m=1$: Für eine C^2-Funktion $f(\mathbf{x})$ bekommt man

$$f(\mathbf{x}) = f(\mathbf{x}^0) + \text{grad}(f)(\mathbf{x}^0)(\mathbf{x}-\mathbf{x}^0) + \mathcal{O}(\|\mathbf{x}-\mathbf{x}^0\|^2).$$

Die Hesse-Matrix.

Man nennt die Matrix

$$\mathbf{H}_f(\mathbf{x}^0) := \begin{pmatrix} f_{x_1x_1}(\mathbf{x}^0) & \dots & f_{x_1x_n}(\mathbf{x}^0) \\ \vdots & & \vdots \\ f_{x_nx_1}(\mathbf{x}^0) & \dots & f_{x_nx_n}(\mathbf{x}^0) \end{pmatrix}$$

die Hesse-Matrix von $f(\mathbf{x})$ im Punkt $\mathbf{x}^0$.

$$\text{Hesse-Matrix} = \text{Jacobi-Matrix des Gradienten } \nabla f$$

Die Taylor-Entwicklung einer C^3-Funktion lautet

$$f(\mathbf{x}) = f(\mathbf{x}^0) + \operatorname{grad} f(\mathbf{x}^0)(\mathbf{x}-\mathbf{x}^0) + \frac{1}{2}(\mathbf{x}-\mathbf{x}^0)^T \mathbf{H}_f(\mathbf{x}^0)(\mathbf{x}-\mathbf{x}^0) + \mathcal{O}(\|\mathbf{x}-\mathbf{x}^0\|^3)$$

Die Hesse-Matrix einer C^2-Funktion ist symmetrisch. $\square$

18 Anwendungen der Differentialrechnung mehrerer Variabler

18.1 Extrema von Funktionen mehrerer Variablen

Definition: *Sei* $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$, *und* $\mathbf{x}^0 \in D$. *Dann hat* $f(\mathbf{x})$ *in* $\mathbf{x}^0$

- *ein* globales Maximum, *falls* $f(\mathbf{x}) \leq f(\mathbf{x}^0)$ *für alle* $\mathbf{x} \in D$.
- *ein* strenges globales Maximum, *falls* $f(\mathbf{x}) < f(\mathbf{x}^0)$ *für alle* $\mathbf{x} \in D$.
- *ein* lokales Maximum, *falls es ein* $\varepsilon > 0$ *gibt mit*

$$f(\mathbf{x}) \leq f(\mathbf{x}^0) \qquad \text{für alle } \mathbf{x} \in D \text{ mit } \|\mathbf{x} - \mathbf{x}^0\| < \varepsilon.$$

- *ein* strenges lokales Maximum, *falls es ein* $\varepsilon > 0$ *gibt mit:*

$$f(\mathbf{x}) < f(\mathbf{x}^0) \qquad \text{für alle } \mathbf{x} \in D \text{ mit } \|\mathbf{x} - \mathbf{x}^0\| < \varepsilon.$$

Analoge Definitionen gelten für Minima. □

Notwendige Bedingung für lokale Extrema.

Satz: *Besitzt* $f(\mathbf{x})$, $f \in C^1$, *in einem Punkt* $\mathbf{x}^0 \in D^0$ *ein lokales Extremum (Minimum oder Maximum), so gilt*

$$\operatorname{grad}(f)(\mathbf{x}^0) = 0.$$

Beweis: Für ein beliebige Richtung $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} \neq 0$ ist die Funktion

$$\varphi(t) := f(\mathbf{x}^0 + t\mathbf{v})$$

in einer Umgebung von $t^0 = 0$ stetig differenzierbar.

Weiterhin hat $\varphi(t)$ bei $t^0 = 0$ ein lokales Extremum. Damit folgt:

$$\varphi'(0) = \operatorname{grad}(f)(\mathbf{x}^0) \cdot \mathbf{v} = 0$$

Da dies für alle $\mathbf{v} \neq 0$ gilt (insbesondere $\mathbf{v} = \operatorname{grad}(f)(\mathbf{x}^0)$), folgt die Bedingung

$$\operatorname{grad}(f)(\mathbf{x}^0) = 0.$$

■

Bemerkungen.

- Die Bedingung $\text{grad}(f)(\mathbf{x}^0) = 0$ liefert ein *nichtlineares* Gleichungssystem zur Berechnung von $\mathbf{x} = \mathbf{x}^0$ mit n Gleichungen und n Unbekannte.
- Die Punkte $\mathbf{x}^0 \in D^0$ mit $\text{grad}(f)(\mathbf{x}^0) = 0$ nennt man stationäre Punkte von $f(\mathbf{x})$. Stationäre Punkte sind **nicht** notwendigerweise lokale Extrema. Zum Beispiel besitzt die Funktion

  $$f(x, y) := x^2 - y^2$$

  den Gradienten

  $$\text{grad}(f)(x, y) = 2(x, -y)$$

  und hat daher nur einen stationären Punkt: $\mathbf{x}^0 = (0, 0)^T$. Der Punkt $\mathbf{x}^0 = (0, 0)^T$ ist jedoch ein Sattelpunkt von f.

Definition: Ein stationärer Punkt $\mathbf{x}^0$ von f heißt Sattelpunkt, falls es in **jeder** Umgebung von $\mathbf{x}^0$ zwei Punkte $\mathbf{x}^1$ und $\mathbf{x}^2$ gibt mit $f(\mathbf{x}^1) < f(\mathbf{x}^0) < f(\mathbf{x}^2)$. $\square$

Klassifikation stationärer Punkte.

Satz: *Sei* $f(\mathbf{x})$ *eine* C^2*-Funktion auf* D^0 *und* $\mathbf{x}^0 \in D^0$ *ein stationärer Punkt von* $f(\mathbf{x})$*, d.h.* $\mathrm{grad}(f)(\mathbf{x}^0) = 0$.

(a) **Notwendige Bedingung**

Ist $\mathbf{x}^0$ *ein lokales Extremum von* $f(\mathbf{x})$*, so gilt:*

$$\begin{aligned} \mathbf{x}^0 \text{ lokales Minimum} \quad &\Longrightarrow \quad \mathbf{H}_f(\mathbf{x}^0) \text{ positiv semidefinit;} \\ \mathbf{x}^0 \text{ lokales Maximum} \quad &\Longrightarrow \quad \mathbf{H}_f(\mathbf{x}^0) \text{ negativ semidefinit;} \end{aligned}$$

(b) **Hinreichende Bedingung**

Ist $\mathbf{H}_f(\mathbf{x}^0)$ *positiv definit (bzw. negativ definit), so ist* $\mathbf{x}^0$ *ein strenges lokales Minimum (bzw. Maximum) von* $f(\mathbf{x})$.

Ist $\mathbf{H}_f(\mathbf{x}^0)$ *indefinit, so ist* $\mathbf{x}^0$ *ein Sattelpunkt, d.h. es gibt in jeder Umgebung von* $\mathbf{x}^0$ *Punkte* $\mathbf{x}^1$ *und* $\mathbf{x}^2$ *mit* $f(\mathbf{x}^1) < f(\mathbf{x}^0) < f(\mathbf{x}^2)$.

Beweis zu (a): Sei $\mathbf{x}^0$ ein lokales Minimum. Für $\mathbf{v} \neq 0$ und $\varepsilon > 0$ hinreichend klein folgt aus der Taylor-Formel

$$f(\mathbf{x}^0 + \varepsilon\mathbf{v}) - f(\mathbf{x}^0) = \frac{1}{2}(\varepsilon\mathbf{v})^T \mathbf{H}_f(\mathbf{x}^0 + \theta\varepsilon\mathbf{v})(\varepsilon\mathbf{v}) \geq 0 \tag{1}$$

mit $\theta = \theta(\varepsilon, \mathbf{v}) \in (0, 1)$.

Der Gradient in der Taylorentwicklung verschwindet, $\text{grad}(f)(\mathbf{x}^0) = 0$, denn $\mathbf{x}^0$ ist stationär. Aus (1) folgt

$$\mathbf{v}^T \mathbf{H}_f(\mathbf{x}^0 + \theta\varepsilon\mathbf{v})\mathbf{v} \geq 0 \tag{2}$$

Da $f(\mathbf{x})$ eine C^2-Funktion ist, ist die Hesse-Matrix eine **stetige** Abbildung. Im Grenzwert $\varepsilon \longrightarrow 0$ folgt daher

$$\mathbf{v}^T \mathbf{H}_f(\mathbf{x}^0)\mathbf{v} \geq 0$$

aus (2), d.h. $\mathbf{H}_f(\mathbf{x}^0)$ ist positiv semidefinit. ■

Beweis zu (b): Ist $\mathbf{H}_f(\mathbf{x}^0)$ positiv definit, so ist $\mathbf{H}_f(\mathbf{x})$ ebenfalls in einer hinreichend kleinen Umgebung $\mathbf{x} \in K_\varepsilon(\mathbf{x}^0) \subset D$ um $\mathbf{x}^0$ positiv definit. Dies folgt aus der Stetigkeit der zweiten partiellen Ableitungen.

Für $\mathbf{x} \in K_\varepsilon(\mathbf{x}^0)$, $\mathbf{x} \neq \mathbf{x}^0$ gilt damit

$$f(\mathbf{x}) - f(\mathbf{x}^0) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^0)^T \mathbf{H}_f(\mathbf{x}^0 + \theta(\mathbf{x} - \mathbf{x}^0))(\mathbf{x} - \mathbf{x}^0) > 0$$

mit $\theta \in (0, 1)$, d.h. $f(\mathbf{x})$ hat in $\mathbf{x}^0$ ein strenges lokales Minimum.

Ist $\mathbf{H}_f(\mathbf{x}^0)$ indefinit, so existieren zu Eigenwerten von $\mathbf{H}_f(\mathbf{x}^0)$ mit verschiedenen Vorzeichen bestimmte Eigenvektoren $\mathbf{v}, \mathbf{w}$ mit

$$\mathbf{v}^T \mathbf{H}_f(\mathbf{x}^0)\mathbf{v} > 0 \qquad \text{und} \qquad \mathbf{w}^T \mathbf{H}_f(\mathbf{x}^0)\mathbf{w} < 0$$

und somit ist $\mathbf{x}^0$ ein Sattelpunkt. ■

Geometrische Interpretation: Die Hesse-Matrix kann positive und negative Eigenwerte besitzen. Die zugehörigen Eigenvektoren geben dabei Richtungen an, in denen die Funktion wächst beziehungsweise fällt. □

Bemerkungen.

- Ein stationärer Punkt $\mathbf{x}^0$ mit $\det(\mathbf{H}_f(\mathbf{x}^0)) = 0$ heißt ausgeartet. Die Hesse-Matrix besitzt dann den Eigenwert $\lambda = 0$.
- Ist $\mathbf{x}^0$ **nicht** ausgeartet, so gibt es 3 Fälle für die Eigenwerte von $\mathbf{H}_f(\mathbf{x}^0)$:

$$\begin{array}{rcl} \text{alle Eigenwerte sind positiv} & \Longrightarrow & \mathbf{x}^0 \text{ ist strenges lokales Minimum;} \\ \text{alle Eigenwerte sind negativ} & \Longrightarrow & \mathbf{x}^0 \text{ ist strenges lokales Maximum;} \\ \text{es gibt positive und negative Eigenwerte} & \Longrightarrow & \mathbf{x}^0 \text{ ist Sattelpunkt.} \end{array}$$

- Es gelten die folgenden Implikationen:

$$\begin{array}{ccc} \mathbf{x}^0 \text{ lokales Minimum} & \Longleftarrow & \mathbf{x}^0 \text{ strenges lokales Minimum} \\ \Downarrow & & \Uparrow \\ \mathbf{H}_f(\mathbf{x}^0) \text{ positiv semidefinit} & \Longleftarrow & \mathbf{H}_f(\mathbf{x}^0) \text{ positiv definit} \end{array}$$

Bemerkung: Für keine der obigen Implikationen gilt die Umkehrung. □

Weitere Bemerkungen.

- Ist $f(\mathbf{x})$ eine C^3-Funktion, $\mathbf{x}^0$ ein stationärer Punkt von $f(\mathbf{x})$ und $\mathbf{H}_f(\mathbf{x}^0)$ positiv definit, so gilt die Abschätzung:

$$(\mathbf{x}-\mathbf{x}^0)^T\,\mathbf{H}_f(\mathbf{x}^0)\,(\mathbf{x}-\mathbf{x}^0) \geq \lambda_{\min}\cdot\|\mathbf{x}-\mathbf{x}^0\|^2$$

wobei $\lambda_{\min}$ den **kleinsten** Eigenwert der Hesse-Matrix bezeichnet.

Nach dem Satz von Taylor gilt dann

$$\begin{aligned} f(\mathbf{x})-f(\mathbf{x}^0) &\geq \frac{1}{2}\lambda_{\min}\|\mathbf{x}-\mathbf{x}^0\|^2 + R_3(\mathbf{x};\mathbf{x}^0) \\ &\geq \|\mathbf{x}-\mathbf{x}^0\|^2\left(\frac{\lambda_{\min}}{2} - C\|\mathbf{x}-\mathbf{x}^0\|\right) \end{aligned}$$

mit einer geeigneten Konstanten $C > 0$.

Um $\mathbf{x}^0$ wächst $f(\mathbf{x})$ somit mindestens quadratisch mit dem Abstand zu $\mathbf{x}^0$.

$\square$

Beispiel. Wir betrachten die Funktion

$$f(x, y) := y^2(x-1) + x^2(x+1)$$

und suchen die stationären Punkte:

$$\operatorname{grad}(f)(x, y) = (y^2 + x(3x+2), 2y(x-1))$$

Die Bedingung $\operatorname{grad}(f)(x, y) = 0$ liefert die beiden stationären Punkte

$$\mathbf{x}^0 = (0, 0)^T \qquad \text{und} \qquad \mathbf{x}^1 = (-2/3, 0).$$

Die jeweiligen Hesse-Matrizen von f an den Stellen $\mathbf{x}^0$ und $\mathbf{x}^1$ lauten

$$\mathbf{H}_f(\mathbf{x}^0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \qquad \text{und} \qquad \mathbf{H}_f(\mathbf{x}^1) = \begin{pmatrix} -2 & 0 \\ 0 & -10/3 \end{pmatrix}$$

$\mathbf{H}_f(\mathbf{x}^1)$ ist negativ definit, somit ist $\mathbf{x}^1$ ein strenges lokales Maximum von $f(\mathbf{x})$.

Die Matrix $\mathbf{H}_f(\mathbf{x}^0)$ ist indefinit, somit ist $\mathbf{x}^0$ ein Sattelpunkt. $\square$