2.1 Partielle und totale Differenzierbarkeit

Hier betrachten wir Funktionen

$$f : \mathbb{R}^n \supset \Omega \to \mathbb{R},\ x = (x_1, \dots, x_n)^T \longmapsto f(x),$$

wobei Ω eine offene Teilmenge des $\mathbb{R}^n$ ist. Offen bedeutet: Zu jedem $x \in \Omega$ gibt es eine Kugel $B_\varepsilon(x) := \{y \in \mathbb{R}^n : |x-y| < \varepsilon\}$, die vollständig in Ω enthalten ist. Zu jedem $x \in \Omega$ gibt es also ein $\varepsilon > 0$, so dass für alle $h \in \mathbb{R}^n$ mit $|h| < \varepsilon$ gilt $x + h \in \Omega$, so dass $f(x+h)$ definiert ist.

Nun seien $e_k = (0, \dots, 0, \overset{k\text{-te Stelle}}{1}, 0 \dots, 0)^T$ die kanonischen Basisvektoren des $\mathbb{R}^n$.

<u>Def.</u>: (1) $f : \Omega \to \mathbb{R}$ heißt in $x \in \Omega$ partiell nach (der Variable) x_k differenzierbar, wenn der Grenzwert

$$\lim_{t \to 0} \frac{1}{t}(f(x + te_k) - f(x)) =: \frac{\partial f}{\partial x_k}(x) \left(= \frac{d}{dt} f(x + te_k)\Big|_{t=0} \right)$$

existiert. In diesem Fall heißt $\frac{\partial f}{\partial x_k}(x)$ die partielle Ableitung von f nach x_k im Punkt $x \in \Omega$.

(2) Ist f in jedem $x \in \Omega$ partiell nach x_k differenzierbar, so heißt die Funktion

$$\frac{\partial f}{\partial x_k} : \Omega \to \mathbb{R},\quad x \longmapsto \frac{\partial f}{\partial x_k}(x)$$

die partielle Ableitung von f nach x_k.

(3) Wir nennen f stetig partiell differenzierbar nach x_k, wenn $\frac{\partial f}{\partial x_k} : \Omega \to \mathbb{R}$ stetig ist.

(4) f heißt (stetig) partiell differenzierbar, wenn f nach allen x_k (stetig) partiell differenzierbar ist.

Bem.: Sei $x \in \Omega$ fixiert, so dass

- für ein $\varepsilon > 0$ $B_\varepsilon(x) \subset \Omega$ und $\cdot$ $\frac{\partial f}{\partial x_k}(x)$ existiert.

Dann ist die Hilfsfunktion

$$\varphi_k(t) = f(x + te_k) = f(x_1, \dots, x_{k-1}, x_k + t, x_{k+1}, \dots, x_n)$$

auf dem Intervall $(-\varepsilon, \varepsilon)$ definiert und es gilt

$$\begin{aligned}\varphi_k'(0) &= \lim_{t \to 0} \frac{1}{t}(\varphi_k(t) - \varphi_k(0)) \\ &= \lim_{t \to 0} \frac{1}{t}(f(x_1, \dots, x_{k-1}, x_k + t, x_{k+1}, \dots, x_n) - f(x)) \\ &= \lim_{t \to 0} \frac{1}{t}(f(x + te_k) - f(x)) = \frac{\partial f}{\partial x_k}(x).\end{aligned}$$

Die partielle Ableitung von f nach x_k ist also nichts anderes als die gewöhnliche Ableitung von f nach x_k, wenn alle anderen Variablen $x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n$ festgehalten werden. Damit stehen uns zur Berechnung partieller Ableitungen alle Ableitungsregeln (Produkt-, Quotienten-, Kettenregel) zur Verfügung.

Bsp.: (1) Ein Polynom in mehreren Variablen:

$$P: \mathbb{R}^4 \to \mathbb{R}, \quad x = (x_1, \dots, x_4)^T \mapsto P(x) = x_1 x_2 + x_3^2 x_4 .$$

Hierfür ist

$$\frac{\partial P}{\partial x_1}(x) = x_2, \quad \frac{\partial P}{\partial x_2}(x) = x_1$$

$$\frac{\partial P}{\partial x_3}(x) = 2x_3 x_4 \quad \text{und} \quad \frac{\partial P}{\partial x_4}(x) = x_3^2 .$$

(2) Die Variablen müssen nicht immer $x, y, (z)$ oder $x_1, \dots, x_n$ heißen, z. B. ist ebenfalls möglich

$$f : (0,\infty) \times (0,\infty) \to \mathbb{R}, \quad (r,s)^T \mapsto f(r,s) = r^s.$$

Hierfür haben wir

$$\frac{\partial f}{\partial r}(r,s) = \frac{\partial}{\partial r} \exp(s \cdot \ln(r)) = \exp(s \ln(r)) \frac{s}{r} = s \cdot r^{s-1}$$

$$\frac{\partial f}{\partial s}(r,s) = \frac{\partial}{\partial s} \exp(s \cdot \ln(r)) = \ln(r) \cdot \exp(s \ln(r)) = \ln(r) r^s.$$

(3) Die Cobb-Douglas-Funktion

$$f : (0,\infty)^n \to (0,\infty), \quad x \mapsto f(x) = x_1^{s_1} \cdots x_n^{s_n} = \prod_{k=1}^{n} x_k^{s_k}$$

tritt in den Wirtschaftswissenschaften z. B. als Produktions- oder Ertragsfunktion auf. (Die Exponenten $s_k > 0$ sind hierbei Parameter, keine Variablen.) Hierfür ist

$$\frac{\partial f}{\partial x_k}(x) = \frac{\partial}{\partial x_k}\left(\prod_{j=1}^{n} x_j^{s_j}\right) = x_1^{s_1} \cdots x_{k-1}^{s_{k-1}} \left(\frac{\partial}{\partial x_k} x_k^{s_k}\right) x_{k+1}^{s_{k+1}} \cdots x_n^{s_n} = \frac{s_k}{x_k} \cdot \prod_{j=1}^{n} x_j^{s_j} = \frac{s_k}{x_k} f(x).$$

Def.: Es sei $f : \Omega \to \mathbb{R}$ partiell differenzierbar. Dann heißt der Zeilenvektor

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x)\right)$$

der Gradient von f im Punkt $x \in \Omega$.

Def.: Eine partiell differenzierbare Funktion $f : \Omega \to \mathbb{R}$ heißt total differenzierbar in $x \in \Omega$, wenn

$$\lim_{h \to 0} \frac{1}{|h|} \left(f(x+h) - f(x) - \nabla f(x) \cdot h \right) = 0.$$

f heißt total differenzierbar in Ω, wenn f in jedem $x \in \Omega$ total differenzierbar ist.

Erläuterungen: (1) h steht hier für eine Folge oder Schar von Vektoren $h \in \mathbb{R}^n$, die in der Euklidischen Norm $|h| = \left(\sum_{k=1}^{n} h_k^2\right)^{\frac{1}{2}}$ gegen $0 \in \mathbb{R}^n$ strebt. Der Summand mit dem Gradienten lautet ausgeschrieben $\nabla f(x) \cdot h = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(x)\, h_k$. Es handelt sich also um das Skalarprodukt von $\nabla f(x)$ mit dem Vektor h.

(2) Bei festem $x \in \Omega$ ist die Abbildung

$$h \mapsto f(x) + \nabla f(x) \cdot h$$

eine affin-lineare Abbildung. Totale Differenzierbarkeit bedeutet also die Approximierbarkeit von f durch eine affin-lineare Funktion.

(3) Es gibt partiell differenzierbare Funktionen, die nicht total differenzierbar sind. Hingegen ist jede stetig partiell differenzierbare Funktion auch total differenzierbar. Im Folgenden werden wir es stets mit stetig partiell differenzierbaren Funktionen zu tun haben.

Weitere Aspekte:

(1) Richtungsableitungen: In der Definition der partiellen Ableitungen muss man sich nicht auf die kanonischen Einheitsvektoren beschränken, sondern kann stattdessen einen beliebigen Vektor $\xi \in \mathbb{R}^n$ der Länge $|\xi| = 1$ wählen. Das Ergebnis ist die <u>Richtungsableitung</u>

$$\frac{\partial f}{\partial \xi}(x) := \lim_{t \to 0} \frac{1}{t}\left(f(x+t\xi) - f(x)\right) = \frac{d}{dt} f(x+t\xi)\Big|_{t=0}$$

von f nach ξ im Punkt $x \in \Omega$.

- Die Richtungsableitung ist ein absolutes Maß für die Änderung der Funktion f in Richtung ξ.
- Die partiellen Ableitungen $\frac{\partial f}{\partial x_k}$ sind spezielle Richtungsableitungen, nämlich die jeweiligen in Richtung der kanonischen Basisvektoren e_k.
- Wenn f total differenzierbar ist, gilt

$$\frac{\partial f}{\partial \xi}(x) = \nabla f(x) \cdot \xi = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(x)\, \xi_k \,,$$

was oft leichter zu berechnen ist.

- Es ist $\nabla f(x) \cdot \xi = |\nabla f(x)| \cdot |\xi| \cdot \cos(\alpha) = |\nabla f(x)| \cos(\alpha)$, wobei α der von $\nabla f(x)$ und ξ eingeschlossene Winkel ist. Dieser Ausdruck wird maximal, wenn $\cos(\alpha) = 1$ ist, d.h. wenn $\xi = \frac{\nabla f(x)^T}{|\nabla f(x)|}$ ist. Der Gradient $\nabla f(x)$ weist also in die Richtung des maximalen Anstiegs der Funktion f, ausgehend vom Punkt $x \in \Omega$.

(2) Partielle und Richtungselastizitäten: Sei $\Omega \subset (0,\infty)^n$ und $f : \Omega \to (0,\infty)$ partiell differenzierbar. Dann definiert man als dimensionsloses Maß für die relative Änderung von f in Richtung der Koordinatenachsen die partiellen Elastizitäten

$$\varepsilon_{f,x_k}(x) := \frac{x_k}{f(x)} \cdot \frac{\partial f}{\partial x_k}(x) .$$

Hieraus bildet man den Elastizitätsgradienten

$$\vec{\varepsilon}_f(x) := (\varepsilon_{f,x_1}(x), \ldots, \varepsilon_{f,x_n}(x))$$

und, für $\xi \in \mathbb{R}^n$ mit $|\xi| = 1$, die Richtungselastizität

$$\varepsilon_{f,\xi}(x) = \vec{\varepsilon}_f(x) \cdot \xi = \sum_{k=1}^{n} \varepsilon_{f,x_k}(x) \cdot \xi_k .$$

Bsp.: Cobb-Douglas-Funktion $f : (0,\infty)^n \to (0,\infty),\ x \mapsto \prod_{k=1}^{n} x_k^{s_k}$.

Bereits festgestellt haben wir, dass

$$\frac{\partial f}{\partial x_k}(x) = \frac{s_k}{x_k} \prod_{j=1}^{n} x_j^{s_j} = \frac{s_k}{x_k} \cdot f(x) .$$

Hieraus folgen $\varepsilon_{f,x_k}(x) = s_k$, $\vec{\varepsilon}_f(x) = (s_1, \dots, s_n)$ und

$\varepsilon_{f,\xi}(x) = \sum_{k=1}^{n} s_k \xi_k$. Die Cobb-Douglas-Funktion erweist sich in alle Richtungen als <u>isoelastisch</u>, d.h. für alle $\xi \in \mathbb{R}^n$ ist $\varepsilon_{f,\xi}$ unabhängig von $x \in (0,\infty)^n$.

(3) Der Mittelwertsatz: Es seien $f : \Omega \to \mathbb{R}$ stetig partiell differenzierbar, $x, x+h \in \Omega$, so dass auch die Verbindungsstrecke $[x, x+h] := \{x + \lambda h : 0 \le \lambda \le 1\} \subset \Omega$. Dann existiert ein $\xi \in [x, x+h]$, so dass

$$f(x+h) = f(x) + \nabla f(\xi) \cdot h .$$

Im Gegensatz zur Definition der totalen Differenzierbarkeit hat man hier auch für $h \neq 0$ eine Identität. Andererseits ist zu beachten, dass ξ von h abhängt. Mit Hilfe des MWS können wir z.B. folgende Aussagen erhalten:

Sei $\Omega \subset \mathbb{R}^n$ zusammenhängend und $f : \Omega \to \mathbb{R}$ stetig partiell differenzierbar. Dann gelten:

(i) Ist $\nabla f(x) = 0 \ \forall x \in \Omega$, so ist f konstant.

(ii) Ist $\nabla f(x)$ ein konstanter Vektor, so ist f eine affin-lineare Abbildung.

Höhere Ableitungen und Taylor-Formel, die Hesse-Matrix

Man kann die Genauigkeit der Approximation im Mittelwertsatz verbessern, indem man zu $f(x) + \nabla f(x) \cdot h$ noch quadratische Terme $a_{ij} h_i h_j$ hinzunimmt. Zur Bestimmung der Koeffizienten a_{ij} benötigt man allerdings partielle Ableitungen zweiter Ordnung von f:

Wenn $f: \Omega \to \mathbb{R}$ partiell differenzierbar ist und auch die partiellen Ableitungen $\frac{\partial f}{\partial x_k}: \Omega \to \mathbb{R}$ nach den Variablen x_j partiell differenzierbar sind, so kann man die 2. partiellen Ableitungen

$$\frac{\partial^2 f}{\partial x_j \partial x_k} := \frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_k} : \Omega \longrightarrow \mathbb{R}$$

von f nach den Variablen x_k und x_j bilden.

Hierfür gilt der Satz von Schwarz: Ist f zweimal partiell differenzierbar nach x_j und x_k und sind die zweiten partiellen Ableitungen

$$\frac{\partial^2 f}{\partial x_j \partial x_k} : \Omega \to \mathbb{R} \qquad \text{und} \qquad \frac{\partial^2 f}{\partial x_k \partial x_j} : \Omega \to \mathbb{R}$$

stetig, so stimmen sie überein. (Das wird bei allen für uns relevanten Funktionen der Fall sein.)

Bsp.: Cobb-Douglas-Funktion $f(x) = \prod_{k=1}^{n} x_k^{s_k}$ (alle $x_k > 0$)

Bereits gesehen: $\frac{\partial f}{\partial x_k}(x) = \frac{s_k}{x_k} \cdot \prod_{j=1}^{n} x_j^{s_j}$.

Zur Berechnung der zweiten partiellen Ableitungen unter-

Schreiben wir zwei Fälle:

(i) $j \neq k$: $\frac{\partial^2 f}{\partial x_j \partial x_k}(x) = \frac{\partial}{\partial x_j}\left(\frac{s_k}{x_k} \prod_{i=1}^{n} x_i^{s_i}\right) = \frac{s_k}{x_k} \cdot \frac{\partial}{\partial x_j} \overbrace{\prod_{i=1}^{n} x_i^{s_i}}^{f(x)}$

$$\underset{\substack{\text{~~bewert~~}\\ \text{berechnet}}}{=} \frac{s_k}{x_k} \cdot \frac{s_j}{x_j} \prod_{i=1}^{n} x_i^{s_i} = \frac{s_k s_j}{x_k x_j} f(x).$$

(ii) $j = k$: $\frac{\partial^2 f}{\partial x_k^2}(x) = \frac{\partial}{\partial x_k}\left(\frac{s_k}{x_k} \cdot \prod_{i=1}^{n} x_i^{s_i}\right) = -\frac{s_k}{x_k^2} \cdot \prod_{i=1}^{n} x_i^{s_i} + \frac{s_k^2}{x_k^2} \prod_{i=1}^{n} x_i^{s_i}$

$$= \frac{1}{x_k^2} s_k(s_k - 1) \cdot \prod_{i=1}^{n} x_i^{s_i} = \frac{s_k(s_k-1)}{x_k^2} f(x).$$

Es erweist sich als nützlich, die zweiten partiellen Ableitungen zu einer quadratischen Matrix zusammenzufassen.

<u>Def.</u>: Es sei $f: \Omega \to \mathbb{R}$ in $x_0 \in \Omega$ zweimal partiell nach allen Variablen $x_1, \dots, x_n$ differenzierbar. Dann heißt

$$\operatorname{Hess} f(x_0) := \left(\frac{\partial^2 f}{\partial x_j \partial x_k}(x_0)\right)_{1 \le j,k \le n}$$

die <u>Hesse-Matrix</u> von f in x_0.

Bem.: (1) Ausgeschrieben:

$$\left(\frac{\partial^2 f}{\partial x_j \partial x_k}(x_0)\right)_{1 \le j,k \le n} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x_0) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x_0) \end{pmatrix}$$

(2) Sind alle 2. partiellen Ableitungen stetig, so ist die Hesse-Matrix nach dem Satz von Schwarz symmetrisch. Dabei heißt eine Matrix $A = (a_{ij})_{1 \le i,j \le n}$ symmetrisch, wenn für alle $i,j \in \{1,\dots,n\}$ gilt, dass $a_{ij} = a_{ji}$.

In Präzisierung des Mittelwertsatzes lässt sich dann die folgende Approximationsaussage zeigen:

<u>Taylor-Formel</u> mit Termen 2. Ordnung: Es seien $f:\Omega\to\mathbb{R}$ zweimal stetig partiell differenzierbar, $x, x+h\in\Omega$, so dass auch $[x,x+h]\subset\Omega$. Dann existiert ein $\xi\in[x,x+h]$, so dass

$$f(x+h) = f(x) + \nabla f(x)\cdot h + \frac{1}{2}h^T\cdot \mathrm{Hess}\, f(\xi)\, h$$

Im nächsten Abschnitt werden wir dann notwendige und hinreichende Bedingungen an $\nabla f(x)$ und $\mathrm{Hess}\, f(x)$ angeben dafür, dass f in $x\in\Omega$ ein lokales Extremum besitzt.