Analysis II

Apl. Prof. Dr. Axel Grünrock

Heinrich-Heine-Universität Düsseldorf

Sommersemester 2020

2.2 Totale Differenzierbarkeit

Die aus der Analysis I bekannte Schlussweise

$$f \text{ differenzierbar} \quad \Rightarrow \quad f \text{ stetig}$$

ist nicht verallgemeinerbar auf Funktionen mehrerer Veränderlicher, wenn wir von f nur die partielle Differenzierbarkeit verlangen. Wir müssen einen etwas schärferen Differenzierbarkeitsbegriff für solche Funktionen einführen, damit die obige Schlussweise wieder korrekt wird. Dies führt auf den Begriff der *totalen Differenzierbarkeit*. Um diesen zu erläutern, gehen wir noch einmal zurück zu den Funktionen einer reellen Veränderlichen:

Lemma 1

Es sei $I \subset \mathbb{R}$ ein Intervall, $x_0 \in I$ und $f : I \to \mathbb{R}$ eine Funktion. Dann gilt: f ist differenzierbar in x_0 genau dann, wenn $c \in \mathbb{R}$, $\varepsilon > 0$ und eine Funktion $\varphi : (x_0 - \varepsilon, x_0 + \varepsilon) \to \mathbb{R}$ mit

$$\lim_{h \to 0} \frac{\varphi(h)}{h} = 0$$

existieren, so dass für alle $|h| < \varepsilon$

$$f(x_0 + h) = f(x_0) + ch + \varphi(h). \tag{1}$$

Beweis des Lemmas

Aus (1) folgt

$$\frac{1}{h}\big(f(x_0+h)-f(x_0)\big) = c + \frac{\varphi(h)}{h} \to 0 \qquad (h \to 0),$$

und das ist die Differenzierbarkeit von f in x_0. Umgekehrt sei die Differenzierbarkeit von f in x_0 vorausgesetzt. Dann definieren wir

$$\varphi(h) := f(x_0+h) - f(x_0) - f'(x_0)h,$$

so dass (1) mit $c = f'(x_0)$ gilt. Ferner ist

$$\frac{\varphi(h)}{h} = \frac{1}{h}\big(f(x_0+h)-f(x_0)\big) - f'(x_0) \to 0 \qquad (h \to 0),$$

also ist auch diese Eigenschaft von φ erfüllt.

$\square$

Lineare Approximierbarkeit

Die in Lemma 1 formulierte äquivalente Eigenschaft bezeichnet man als “Approximierbarkeit durch eine (affin-) lineare Abbildung” oder kurz als “lineare Approximierbarkeit”. Im Gegensatz zur üblichen Forderung nach der Existenz des Grenzwerts

$$\lim_{h \to 0} \frac{1}{h}\big(f(x_0 + h) - f(x_0)\big),$$

bei der man für h keinen Vektor einsetzen kann, ist sie problemlos verallgemeinerbar auf Funktionen mehrerer Variablen:

Definition der totalen Differenzierbarkeit

Es sei $\Omega \subset \mathbb{R}^n$ offen und $f : \Omega \to \mathbb{R}^m$ eine Funktion. f heißt *total differenzierbar* (oder einfach kurz: differenzierbar) in $x_0 \in \Omega$, falls eine lineare Abbildung $A : \mathbb{R}^n \to \mathbb{R}^m$ existiert, so dass

$$f(x_0 + h) = f(x_0) + Ah + \varphi(h) \tag{2}$$

gilt. Hierbei ist $\varphi : \mathbb{R}^n \supset B_\varepsilon(0) \to \mathbb{R}^m$ mit der Eigenschaft

$$\lim_{h \to 0} \frac{\varphi(h)}{|h|} = 0.$$

Bemerkungen

(1) Ausgeschrieben als Vektoren $f = (f_1, ..., f_m)^\top$, $\varphi = (\varphi_1, ..., \varphi_m)^\top$ bzw. als Matrix $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ lautet die Gleichung (2)

$$f_i(x_0 + h) = f_i(x_0) + \sum_{j=1}^{n} a_{ij} h_j + \varphi_i(h), \qquad 1 \leq i \leq m.$$

Also ist f in x_0 (total) differenzierbar genau dann, wenn dies für alle Komponentenfunktionen von f der Fall ist.

Bemerkungen

(1) Ausgeschrieben als Vektoren $f = (f_1, ..., f_m)^\top$, $\varphi = (\varphi_1, ..., \varphi_m)^\top$ bzw. als Matrix $A = (a_{ij})_{1 \le i \le m, 1 \le j \le n}$ lautet die Gleichung (2)

$$f_i(x_0 + h) = f_i(x_0) + \sum_{j=1}^{n} a_{ij} h_j + \varphi_i(h), \qquad 1 \le i \le m.$$

Also ist f in x_0 (total) differenzierbar genau dann, wenn dies für alle Komponentenfunktionen von f der Fall ist.

(2) f heißt (total) differenzierbar in Ω, wenn f in jedem $x_0 \in \Omega$ (total) differenzierbar ist.

Beispiel

$$f : \mathbb{R}^n \to \mathbb{R}, \quad x \mapsto f(x) := \langle x, Mx \rangle = \sum_{i,j=1}^{n} m_{ij} x_i x_j$$

mit einer Matrix $M \in M_n(\mathbb{K})$. Dann ist

$$f(x+h) = \langle x+h, M(x+h) \rangle$$

$$= \langle x, Mx \rangle + \langle h, Mx \rangle + \langle x, Mh \rangle + \langle h, Mh \rangle$$

$$= f(x) + Ah + \varphi(h),$$

wobei $A = A(x) = Mx + M^\top x = (M + M^\top)x$ und $\varphi(h) = \langle h, Mh \rangle$

$$\text{mit} \qquad \lim_{h \to 0} \frac{\varphi(h)}{|h|} = \lim_{h \to 0} \frac{1}{|h|} \langle h, Mh \rangle = 0,$$

letzteres, da $\frac{1}{|h|} |\langle h, Mh \rangle| \leq \|M\| |h|$. Die approximierende lineare Abbildung A hängt also von x ab!

Totale Differenzierbarkeit impliziert Stetigkeit

Lemma 2: Ist $f : \mathbb{R}^n \supset \Omega \to \mathbb{R}^m$ (total) differenzierbar in $x_0 \in \Omega$, so ist f in x_0 stetig.

Beweis:

$$|f(x_0+h)-f(x_0)| = |Ah+\varphi(h)| \leq \big(\|A\|+\frac{\varphi(h)}{|h|}\big)|h| \to 0 \quad (h \to 0).$$

$\square$

Totale Differenzierbarkeit impliziert Stetigkeit

Lemma 2: Ist $f : \mathbb{R}^n \supset \Omega \to \mathbb{R}^m$ (total) differenzierbar in $x_0 \in \Omega$, so ist f in x_0 stetig.

Beweis:

$$|f(x_0+h)-f(x_0)| = |Ah+\varphi(h)| \leq \Big(\|A\|+\frac{\varphi(h)}{|h|}\Big)|h| \to 0 \quad (h \to 0).$$

$\square$

Als nächstes klären wir den Zusammenhang zwischen totaler und partieller Differenzierbarkeit:

Satz 1

Es sei $\Omega \subset \mathbb{R}^n$ offen und $f : \Omega \to \mathbb{R}^m$ total differenzierbar in $x_0 \in \Omega$, es gelte also

$$f(x_0 + h) - f(x_0) = Ah + \varphi(h)$$

mit einer Matrix $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ und einer Funktion $\varphi : \mathbb{R}^n \supset B_\varepsilon(0) \to \mathbb{R}^m$, für die $\lim\limits_{h \to 0} \dfrac{\varphi(h)}{|h|} = 0$ ist.

Dann ist f in x_0 partiell differenzierbar und für alle $1 \leq i \leq m$ und $1 \leq j \leq n$ gilt

$$a_{ij} = \frac{\partial f_i}{\partial x_j}(x_0).$$

Beweis von Satz 1

Mit $h = te_j$ (die e_j sind die kanonischen Basisvektoren) haben wir für die Komponenten

$$f_i(x_0 + te_j) = f_i(x_0) + t(Ae_j)_i + \varphi_i(te_j)$$

beziehungsweise

$$\frac{1}{t}\big(f_i(x_0 + te_j) - f_i(x_0)\big) = \sum_{k=1}^{n} a_{ik}\delta_{kj} + \frac{\varphi_i(tej)}{t} = a_{ij} + \frac{\varphi_i(tej)}{t}.$$

Für $t \to 0$ verschwindet der letzte Summand, und also existiert der Grenzwert

$$\lim_{t\to 0} \frac{1}{t}\big(f_i(x_0 + te_j) - f_i(x_0)\big) = \frac{\partial f_i}{\partial x_j}(x_0) = a_{ij}.$$

$\square$

Bemerkung und Bezeichnungen

Die Matrix A ist also durch die Funktion f eindeutig festgelegt. Man nennt sie die “Jacobimatrix” oder auch “Funktionalmatrix” von f. Schreibweisen sind $A = Df(x_0) = J_f(x_0)$ beziehungsweise (am genauesten)

$$A = \left(\frac{\partial f_i}{\partial x_j}(x_0)\right)_{\substack{1 \le i \le m \\ 1 \le j \le n}}.$$

Ist f reellwertig, so besteht Df nur aus einer einzigen Zeile, nämlich dem Gradienten. Für vektorwertige Funktionen gilt: Die Zeilen der Jacobi-Matrix sind die Gradienten der Komponentenfunktionen. Ist also $m \ge 2$ und $f : \Omega \to \mathbb{R}^m$, so gilt

$$Df(x_0) = \begin{pmatrix} - & \nabla f_1(x_0) & - \\ . & . & . \\ . & . & . \\ - & \nabla f_m(x_0) & - \end{pmatrix}$$

Bemerkungen und Bezeichnungen (Fortsetzung)

Wir können die Jacobi-Matrix auch spaltenweise auffassen: Die Spaltenvektoren der Matrix Df sind gerade die partiellen Ableitungen der vektorwertigen Funktion f:

$$Df(x_0) = \begin{pmatrix} \Big| & \cdot & \cdot & \Big| \\ \Big| & \cdot & \cdot & \Big| \\ \frac{\partial f}{\partial x_1}(x_0) & \cdot & \cdot & \frac{\partial f}{\partial x_n}(x_0) \\ \Big| & \cdot & \cdot & \Big| \\ \Big| & \cdot & \cdot & \Big| \end{pmatrix}.$$

Beispiel zur Berechnung einer Jacobi-Matrix

Es sei

$$f : \mathbb{R}^3 \to \mathbb{R}^2,\ x = (x_1, x_2, x_3)^\top \mapsto f(x) = (f_1(x), f_2(x))^\top$$

mit $f_1(x) = x_1^2 + x_2^2 + x_3^2$ und $f_2(x) = x_1 x_2 x_3$. Hierfür ist

$$Df(x) = \left(\frac{\partial f_i}{\partial x_j}(x)\right)_{\substack{1\le i\le 2\\ 1\le j\le 3}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & .. & \frac{\partial f_1}{\partial x_3}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & .. & \frac{\partial f_2}{\partial x_3}(x) \end{pmatrix}$$

$$= \begin{pmatrix} 2x_1 & 2x_2 & 2x_3 \\ x_2x_3 & x_1x_3 & x_1x_2 \end{pmatrix}.$$

Das Beispiel zeigt: Die Abhängigkeit der Matrix Df von x ist in der Regel nichtlinear. Wenn in der Definition der totalen Differenzierbarkeit von einer "linearen Abbildung" die Rede ist, so ist die Zuordnung $h \mapsto Df(x) \cdot h$ (bei festem x) gemeint. Auf dieses Beispiel soll an späterer Stelle noch einmal zurückgegriffen werden.

Stetige partielle Differenzierbarkeit impliziert die totale Differenzierbarkeit

Kann man anhand der partiellen Ableitungen bereits feststellen, ob eine Funktion total differenzierbar ist? Hierzu gibt es ein zwar nicht scharfes, aber zumindest hinreichendes Kriterium:

Satz 2: Sei $\Omega \subset \mathbb{R}^n$ offen und $f : \Omega \to \mathbb{R}^m$ partiell differenzierbar. Alle partiellen Ableitungen seien in $x_0 \in \Omega$ stetig. Dann ist f in x_0 total differenzierbar.

Beweis, o.E. für $m = 1$

Wir setzen

$$\varphi(h) = f(x_0 + h) - f(x_0) - \nabla f(x_0)h$$

und zeigen, dass $\lim_{h \to 0} \frac{\varphi(h)}{|h|} = 0$. Dazu führen wir die Zwischenstellen

$$z_k = x_0 + \sum_{j=1}^{k} h_j e_j$$

ein, so dass $z_0 = x_0$, $z_n = x_0 + h$ und $z_k - z_{k-1} = h_k e_k$. (Man "wandert" hierbei also entlang der Koordinatenachsen von x_0 nach $x_0 + h$, wobei man bei den z_k einen Zwischenstopp einlegt und im rechten Winkel abbiegt.) Dann gilt ..

Beweis, Fortsetzung

$$f(x_0 + h) - f(x_0) = \sum_{k=1}^{n} f(z_k) - f(z_{k-1}) = .. = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\xi_k) h_k,$$

wobei der Mittelwertsatz für Funktionen einer Veränderlicher wiederholt angewendet wurde, was das Auftreten der Stellen ξ_k erklärt, welche auf der Verbindungsstrecke zwischen z_{k-1} und z_k liegen (einen weiteren Zwischenschritt der Rechnung finden Sie im Manuskript). Mit $\nabla f(x_0)h = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(x_0) h_k$ folgt aufgrund der Stetigkeit der partiellen Ableitungen

$$\frac{\varphi(h)}{|h|} = \sum_{k=1}^{n} \left(\frac{\partial f}{\partial x_k}(\xi_k) - \frac{\partial f}{\partial x_k}(x_0) \right) \frac{h_k}{|h|} \to 0 \quad (h \to 0).$$

$\square$

Zusammenfassung

Zwischen den Regularitätseigenschaften einer Funktion *mehrerer* Veränderlicher bestehen also die Implikationen

$$\begin{array}{c}\text{stetig partiell}\\ \text{differenzierbar}\end{array} \quad \underset{\text{(Satz 2)}}{\Rightarrow} \quad \begin{array}{c}\text{total}\\ \text{differenzierbar}\end{array} \begin{cases} \underset{\text{(Lemma 2)}}{\Rightarrow} & : \quad \text{stetig} \\ & : \\ \underset{\text{(Satz 1)}}{\Rightarrow} & : \quad \begin{array}{c}\text{partiell}\\ \text{diff.bar}\end{array}\end{cases}$$

Insbesondere ist jede stetig partiell differenzierbare Funktion stetig. Alle Umkehrungen im obigen Diagramm gelten im Allgemeinen *nicht*! Was den Satz 1 anbelangt, sei an das Bsp. (3) aus Abschnitt 2.1 erinnert: Die dort angegebene Funktion ist im Nullpunkt zwar partiell differenzierbar, aber unstetig und somit auch nicht total differenzierbar. Das Problem bei der umgekehrten Schlussrichtung in Satz 2 tritt bereits im eindimensionalen Fall auf, etwa für $g(x) = x^2 \cos\left(\frac{1}{x}\right)$ (in $x_0 = 0$ stetig durch $g(0) = 0$ interpoliert). Diese Funktion ist differenzierbar, aber die Ableitung ist unstetig.

Frage zum Abschluss

Gibt es eine Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, die stetig und partiell differenzierbar ist, aber nicht stetig partiell differenzierbar?

Tip: Versuchen Sie es mit

$$f(x,y) = \begin{cases} \frac{x^2 y}{x^2+y^2} & : \quad (x,y) \neq (0,0) \\ 0 & : \quad (x,y) = (0,0) \end{cases}$$