

Differentialrechnung in mehreren Veränderlichen

Es gibt zwei verschiedene Differenzierbarkeitsbegriffe für Funktionen $\mathbb{R}^n \rightarrow \mathbb{R}^m$:

- die partielle Differenzierbarkeit oder allgemeiner die Differenzierbarkeit in einer Richtung
- die totale Differenzierbarkeit

Während die partielle Differenzierbarkeit leichter zu verstehen und zu handhaben ist, liefert erst die totale Differenzierbarkeit den Begriff, der in der Theorie benötigt wird. Dennoch sind bei der Betrachtung totaler Ableitungen auch partielle Ableitungen von Nutzen. Wir stellen die Begriffe zunächst im Einzelnen vor und beschreiben dann ihren Zusammenhang.

Partielle Differenzierbarkeit/Differenzierbarkeit in einer Richtung

Wir betrachten zunächst nur $\mathbb{R}$ -wertige Funktionen. Sei also $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $x = (x_1, \dots, x_n) \mapsto f(x) = f(x_1, \dots, x_n)$ eine Funktion. Da der Funktionswert in einem Punkt $x \in \mathbb{R}^n$ nicht mehr nur von einer, sondern von n reellen Zahlen abhängt, ist der aus $\mathbb{R}$ bekannte Differenzierbarkeitsbegriff nicht anwendbar. Dieses Problem können wir beheben, indem wir für ein $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ und ein $j \in \{1, \dots, n\}$ die Koordinaten $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n$ festhalten und an x_j „wackeln“. Wir erhalten also die Funktion

$$f_{x,j} : \mathbb{R} \rightarrow \mathbb{R}, \quad t \mapsto f(x_1, \dots, x_{j-1}, x_j + t, x_{j+1}, \dots, x_n)$$

und damit folgende

DEFINITION:

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt in $x \in \mathbb{R}^n$ *partiell nach der j -ten Variablen differenzierbar*, falls $f_{x,j}$ in $0 \in \mathbb{R}$ differenzierbar ist. In diesem Fall heißt

$$(\partial_j f)(x) := f'_{x,j}(0) \quad \text{andere Bezeichnungen:} \quad (\partial_{x_j} f)(x), \quad \frac{\partial f}{\partial x_j}(x), \quad (D_j f)(x), \quad (D_{x_j} f)(x), \quad f_{x_j}(x)$$

die j -te partielle Ableitung von f in x . f heißt *partiell nach der j -ten Variablen differenzierbar*, falls f in allen $x \in \mathbb{R}^n$ partiell nach der j -ten Variablen differenzierbar ist. In diesem Fall heißt

$$\partial_j f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad x \mapsto (\partial_j f)(x)$$

die j -te partielle Ableitung von f . f heißt *partiell differenzierbar*, falls f für alle $j \in \{1, \dots, n\}$ partiell nach der j -ten Variablen differenzierbar ist. In diesem Fall heißt der (Zeilen-)Vektor

$$(\text{grad } f)(x) = ((\partial_j f)(x))_{j=1, \dots, n} = ((\partial_1 f)(x), \dots, (\partial_n f)(x)) \in \mathbb{R}^{1 \times n} \simeq \mathbb{R}^n$$

der *Gradient* von f in x .

BEISPIEL:

$$f : \mathbb{R}^3 \rightarrow \mathbb{R}, \quad (x_1, x_2, x_3) \mapsto f(x_1, x_2, x_3) = x_1 x_2^2 \sin x_3$$

ist partiell differenzierbar mit

$$(\text{grad } f)(x_1, x_2, x_3) = (x_2^2 \sin x_3, \quad 2x_1 x_2 \sin x_3, \quad x_1 x_2^2 \cos x_3)$$

Wir bemerken, dass mit der Kurve

$$\gamma_{x,j} : \mathbb{R} \rightarrow \mathbb{R}^n, \quad t \mapsto (x_1, \dots, x_{j-1}, x_j + t, x_{j+1}, \dots, x_n) = x + t e_j$$

gerade gilt: $f_{x,j} = f \circ \gamma_{x,j}$. Wir haben also einfach nur f auf die Gerade $\gamma_{x,j}(\mathbb{R}) = x + \mathbb{R} e_j$ eingeschränkt und diese mit $\mathbb{R}$ identifiziert, um eine auf $\mathbb{R}$ definierte Funktion zu erhalten, auf die wir den bekannten Ableitungsbegriff anwenden konnten. Lässt man neben den kanonischen Einheitsvektoren $e_1, \dots, e_n$ auch andere Richtungen zu, so ergibt sich der allgemeinere Begriff der Richtungsableitung.

DEFINITION:

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt in $x \in \mathbb{R}^n$ *in der Richtung $v \in \mathbb{R}^n$ differenzierbar*, falls die Funktion

$$f_{x,v} : \mathbb{R} \rightarrow \mathbb{R}, \quad t \mapsto f(x + tv)$$

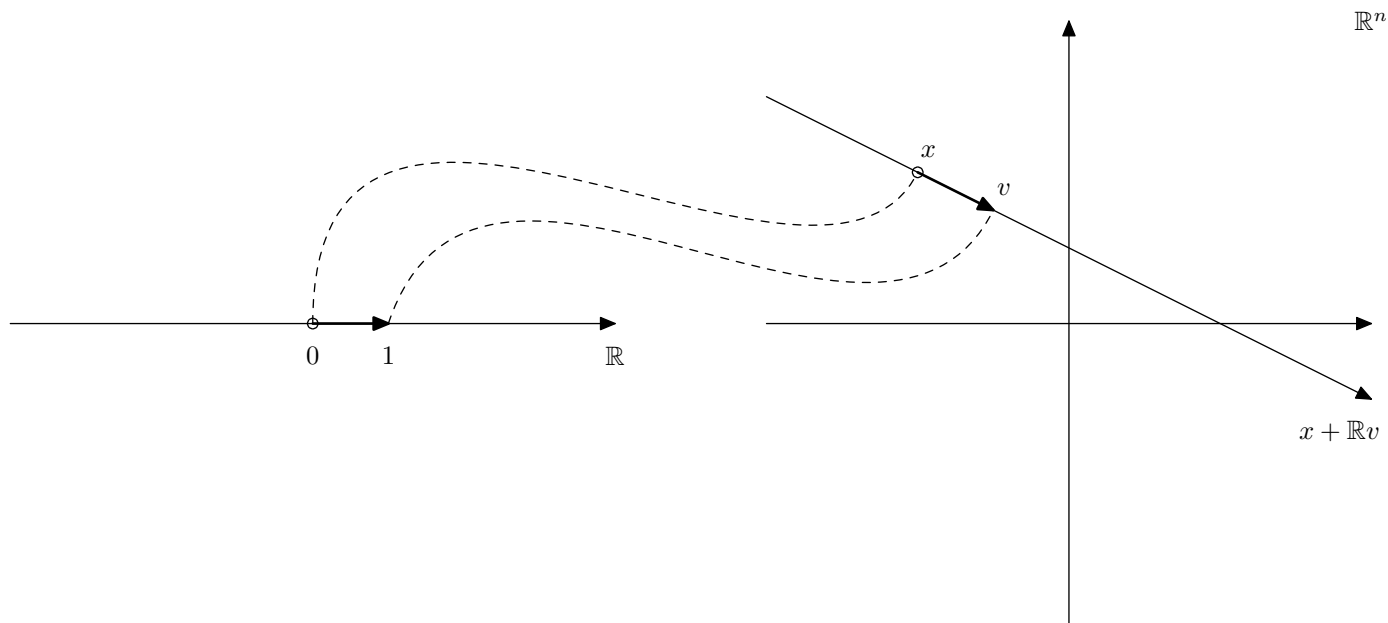
in $0 \in \mathbb{R}$ differenzierbar ist, d. h. falls der Grenzwert

$$\lim_{\substack{t \rightarrow 0 \\ t \neq 0}} \frac{f(x+tv) - f(x)}{t} \in \mathbb{R}$$

existiert. In diesem Fall heit

$$(\partial_v f)(x) := f'_{x,v}(0) = \lim_{\substack{t \rightarrow 0 \\ t \neq 0}} \frac{f(x+tv) - f(x)}{t} \quad \text{andere Bezeichnungen: } (D_v f)(x), \quad f_v(x)$$

die *Richtungsableitung von f in x in der Richtung v* . Fr $j \in \{1, \dots, n\}$ gilt $(\partial_j f)(x) = (\partial_{e_j} f)(x)$. Eine partielle Ableitung ist also eine Richtungsableitung in Richtung eines kanonischen Einheitsvektors.



BEMERKUNGEN:

- Diese Begriffe lassen sich leicht auf Funktionen $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ verallgemeinern, indem man die m Komponenten $f_1, \dots, f_m : \mathbb{R}^n \longrightarrow \mathbb{R}$ von f getrennt behandelt. Eine partielle Ableitung bzw. eine Richtungsableitung ist dann ein Vektor in $\mathbb{R}^m$. Statt dem Gradienten erhlt man die sogenannte Jacobi-Matrix:

$$J_f(x) = ((\partial_j f_i)(x))_{\substack{i=1,\dots,m \\ j=1,\dots,n}} = \begin{pmatrix} (\partial_1 f_1)(x) & \cdots & (\partial_n f_1)(x) \\ \vdots & \ddots & \vdots \\ (\partial_1 f_m)(x) & \cdots & (\partial_n f_m)(x) \end{pmatrix} = \begin{pmatrix} (\text{grad } f_1)(x) \\ \vdots \\ (\text{grad } f_m)(x) \end{pmatrix} \in \mathbb{R}^{m \times n}$$

- Ist f nur auf einer offenen Menge $U \subset \mathbb{R}^n$ definiert, so werden die Begriffe analog eingefhrt.

Totale Differenzierbarkeit

Wir erinnern uns daran, dass Differenzierbarkeit in $\mathbb{R}$ eine „gute“ Approximierbarkeit durch eine affin-lineare Funktion in der Nhe des betrachteten Punkts bedeutet. Tatschlich ist eine Funktion $f : \mathbb{R} \longrightarrow \mathbb{R}$ genau dann differenzierbar in $x \in \mathbb{R}$, wenn ein $a \in \mathbb{R}$ existiert, so dass fr alle $h \in \mathbb{R}$ gilt:

$$f(x+h) = f(x) + ah + \phi(h) \quad \text{mit} \quad \phi : \mathbb{R} \longrightarrow \mathbb{R}, \quad \lim_{\substack{h \rightarrow 0 \\ h \neq 0}} \frac{\phi(h)}{|h|} = 0$$

Die Anschauung dabei ist, dass die affin-lineare Funktion $g : x+h \mapsto f(x) + ah$ die Funktion f nahe x derart gut approximiert, dass der Fehler $f(x+h) - g(x+h) = f(x+h) - f(x) - ah = \phi(h)$ fr $h \rightarrow 0$ sogar dann noch gegen 0 geht, wenn man ihn zunchst durch $|h|$ dividiert. (Man sagt „ $\phi(h)$ ist von der Ordnung klein-o von $|h|$ “ und schreibt $\phi(h) = o(|h|)$.) Das a ist durch f und x eindeutig bestimmt. Es gilt nmlich $a = f'(x)$.

Diese Idee lsst sich auf Funktionen $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ verallgemeinern, indem man die lineare Abbildung $\mathbb{R} \longrightarrow \mathbb{R}, h \mapsto ah$, durch eine lineare Abbildung $A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ und den Betrag $|\cdot|$ durch die Norm $\|\cdot\|_2$ ersetzt.

DEFINITION:

Eine Funktion $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ heißt *(total) differenzierbar in $x \in \mathbb{R}^n$* , falls eine lineare Abbildung $A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ existiert, so dass für alle $h \in \mathbb{R}^n$ gilt:

$$f(x+h) = f(x) + A(h) + \phi(h) \quad \text{mit} \quad \phi : \mathbb{R}^n \longrightarrow \mathbb{R}^m, \quad \lim_{\substack{h \rightarrow 0 \\ h \neq 0}} \frac{\phi(h)}{\|h\|_2} = 0$$

Die lineare Abbildung A ist durch f und x eindeutig bestimmt und heißt *(totale) Ableitung von f in x* . Sie wird mit $(Df)(x) \in \text{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ bezeichnet. Oft identifiziert man die lineare Abbildung $(Df)(x)$ mit ihrer Matrixdarstellung bzgl. der kanonischen Basen von $\mathbb{R}^n$ und $\mathbb{R}^m$ und schreibt $(Df)(x) \in \mathbb{R}^{m \times n}$. f heißt *(total) differenzierbar*, falls f in allen $x \in \mathbb{R}^n$ total differenzierbar ist. In diesem Fall heißt die Abbildung

$$Df : \mathbb{R}^n \longrightarrow \text{Hom}(\mathbb{R}^n, \mathbb{R}^m) \simeq \mathbb{R}^{m \times n}, \quad x \longmapsto (Df)(x)$$

(totale) Ableitung von f .

BEISPIELE:

- Ist $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ linear, so ist f total differenzierbar und es gilt $(Df)(x) = f$ für alle $x \in \mathbb{R}^n$, denn für alle x und $h \in \mathbb{R}^n$ gilt:

$$f(x+h) = f(x) + f(h) = f(x) + f(h) + 0 \quad \text{mit} \quad \lim_{\substack{h \rightarrow 0 \\ h \neq 0}} \frac{0}{\|h\|_2} = 0$$

Man wählt also $A := f$ und $\phi(h) := 0$ für alle $h \in \mathbb{R}^n$.

- Sei $\langle \cdot, \cdot \rangle$ das Standardskalarprodukt in $\mathbb{R}^n$ und $B \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix. Dann heißt

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad x \longmapsto \langle Bx, x \rangle$$

eine *quadratische Form*. f ist total differenzierbar und es gilt $(Df)(x) = 2 \langle Bx, \cdot \rangle$, d. h. $(Df)(x)(h) = 2 \langle Bx, h \rangle$ für alle $x, h \in \mathbb{R}^n$. Beweis: Sei $x \in \mathbb{R}^n$ fest. Setze $A(h) := 2 \langle Bx, h \rangle$ für alle $h \in \mathbb{R}^n$. Dann gilt:

$$f(x+h) = \langle B(x+h), x+h \rangle = \langle Bx+Bh, x+h \rangle = \langle Bx, x \rangle + \langle Bx, h \rangle + \langle Bh, x \rangle + \langle Bh, h \rangle$$

Nun ist $\langle Bh, x \rangle = \langle h, B^T x \rangle = \langle h, Bx \rangle = \langle Bx, h \rangle$ wegen der Symmetrie von B und $\langle \cdot, \cdot \rangle$. Es folgt:

$$f(x+h) = \langle Bx, x \rangle + 2 \langle Bx, h \rangle + \langle Bh, h \rangle = f(x) + A(h) + \phi(h) \quad \text{mit} \quad \phi(h) = \langle Bh, h \rangle$$

Mit der Ungleichung von Cauchy-Schwarz gilt:

$$|\phi(h)| = |\langle Bh, h \rangle| \leq \|Bh\|_2 \|h\|_2 \leq \|B\| \|h\|_2^2$$

wobei $\|B\|$ die Matrixnorm (Operatornorm) von B ist. Es folgt:

$$0 \leq \left| \frac{\phi(h)}{\|h\|_2} \right| = \frac{|\phi(h)|}{\|h\|_2} \leq \frac{\|B\| \|h\|_2^2}{\|h\|_2} = \|B\| \|h\|_2 \xrightarrow{h \rightarrow 0} 0 \quad \implies \quad \lim_{\substack{h \rightarrow 0 \\ h \neq 0}} \frac{\phi(h)}{\|h\|_2} = 0$$

BEMERKUNG: Dies ist offenbar eine Verallgemeinerung der Aussage $\frac{d}{dx}(bx^2) = 2bx$ für festes $b \in \mathbb{R}$.

BEMERKUNGEN:

- Ist f nur auf einer offenen Menge $U \subset \mathbb{R}^n$ definiert, so wird der Begriff analog eingeführt.
- Der Begriff der totalen Differenzierbarkeit lässt sich auf naheliegende Weise auf Abbildungen $f : V \longrightarrow W$ zwischen beliebigen (insbesondere nicht notwendigerweise endlich-dimensionalen) normierten Räumen V und W verallgemeinern. Dann fordert man üblicherweise zusätzlich, dass die lineare Abbildung $A : V \longrightarrow W$ stetig ist. Im endlich-dimensionalen Fall, insbesondere bei $V = \mathbb{R}^n$ und $W = \mathbb{R}^m$, ist dies automatisch gegeben.

Zusammenhang der beiden Begriffe

Es gilt:

- Die partielle Differenzierbarkeit bzw. die Differenzierbarkeit in einer Richtung liefert nicht die affin-lineare Approximierbarkeit, die man von einem Differenzierbarkeitsbegriff erwarten würde. Tatsächlich muss eine Funktion, deren sämtliche Richtungsableitungen in einem Punkt existieren, in diesem Punkt nicht einmal stetig sein. Dafür lassen sich partielle Ableitungen bzw. Richtungsableitungen meist leicht berechnen.

- Die totale Differenzierbarkeit ist gerade als affin-lineare Approximierbarkeit definiert und stellt daher einen „guten“ Differenzierbarkeitsbegriff dar. Tatsächlich ist eine in einem Punkt total differenzierbare Funktion in diesem Punkt insbesondere stetig. Dafür ist die totale Ableitung schwieriger zu handhaben.

Glücklicherweise gelten jedoch die folgenden wichtigen Zusammenhänge:

- (i) Ist $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ total differenzierbar in $x \in \mathbb{R}^n$, so ist f in x in allen Richtungen differenzierbar und es gilt:

$$(Df)(x)(v) = (\partial_v f)(x) \text{ für alle } v \in \mathbb{R}^n, \quad \text{insbesondere } (Df)(x)(e_j) = (\partial_j f)(x) \text{ für } j \in \{1, \dots, n\}$$

Man erhält also eine partielle Ableitung bzw. eine Richtungsableitung durch Einsetzen eines kanonischen Einheitsvektors bzw. einer Richtung in die totale Ableitung. Hieraus ergibt sich, dass die Matrix der linearen Abbildung $(Df)(x)$ bzgl. der kanonischen Basen von $\mathbb{R}^n$ und $\mathbb{R}^m$ gerade die Jacobi-Matrix $J_f(x)$ ist.

- (ii) Ist $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ in einer Umgebung von $x \in \mathbb{R}^n$ nach allen Variablen partiell differenzierbar und sind die partiellen Ableitungen stetig in x , so ist f in x total differenzierbar. Nach (i) gilt dann:

$$(Df)(x)(v) = J_f(x)v \quad \text{für alle } v \in \mathbb{R}^n$$

Wir können also die totale Differenzierbarkeit dadurch nachweisen, dass wir die Existenz und Stetigkeit der partiellen Ableitungen zeigen. Außerdem ergibt sich die Matrix der totalen Ableitung bzgl. der kanonischen Basen als die Jacobi-Matrix, die wiederum aus den partiellen Ableitungen besteht.

BEISPIEL:

Wir testen Aussage (i) an der oben definierten quadratischen Form f . Ist $B = (b_{kl})_{k,l=1,\dots,n}$, so gilt für $x = (x_1, \dots, x_n) \in \mathbb{R}^n$:

$$f(x) = \langle Bx, x \rangle = \left\langle \left(\sum_{l=1}^n b_{kl} x_l \right)_{k=1,\dots,n}, (x_k)_{k=1,\dots,n} \right\rangle = \sum_{k=1}^n \left(\sum_{l=1}^n b_{kl} x_l \right) x_k = \sum_{k,l=1}^n b_{kl} x_k x_l$$

Wir rechnen nach, dass f nach allen Variablen partiell differenzierbar ist mit

$$(\partial_j f)(x) = \frac{\partial}{\partial x_j} \left(\sum_{k,l=1}^n b_{kl} x_k x_l \right) = \sum_{\substack{k=1 \\ k \neq j}}^n b_{kj} x_k + \sum_{\substack{l=1 \\ l \neq j}}^n b_{jl} x_l + 2b_{jj} x_j \stackrel{B=B^T}{=} 2 \sum_{k=1}^n b_{jk} x_k$$

Damit gilt

$$J_f(x) = (\text{grad } f)(x) = ((\partial_j f)(x))_{j=1,\dots,n} = \left(2 \sum_{k=1}^n b_{jk} x_k \right)_{j=1,\dots,n} = 2(Bx)^T$$

Da die partiellen Ableitungen stetig sind, ist f total differenzierbar mit

$$(Df)(x)(v) = J_f(x)v = 2(Bx)^T v = 2 \langle Bx, v \rangle$$

Dies entspricht aber gerade obigem Ergebnis.

Differentialoperatoren

Wir führen schließlich noch einige Differentialoperatoren ein.

- *Nabla-Operator* für Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}$:

$$(\nabla f)(x) := (\text{grad } f)(x)$$

- *Divergenz* für Funktionen $f = (f_1, \dots, f_n) : \mathbb{R}^n \rightarrow \mathbb{R}^n$:

$$(\text{div } f)(x) = \sum_{j=1}^n (\partial_j f_j)(x) = \text{Spur } J_f(x)$$

- *Laplace-Operator* für Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}$:

$$(\Delta f)(x) = \sum_{j=1}^n (\partial_j (\partial_j f))(x) = (\text{div}(\nabla f))(x)$$

- *Rotation* für Funktionen $f = (f_1, f_2, f_3) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$:

$$(\text{rot } f)(x) = \begin{pmatrix} (\partial_2 f_3)(x) - (\partial_3 f_2)(x) \\ (\partial_3 f_1)(x) - (\partial_1 f_3)(x) \\ (\partial_1 f_2)(x) - (\partial_2 f_1)(x) \end{pmatrix}$$