Sei im folgenden $A \subseteq \mathbb{R}^n$ und $\mathbf{x}^0 \in A$.

Partielle Ableitungen von $f : \mathbb{R}^n \to \mathbb{K}$

Sei $f : A \to \mathbb{K}$ und $j \in \{1, \ldots, n\}$. Existiert ein $\varepsilon > 0$, s.d. $\forall h \in (-\varepsilon, \varepsilon) : \mathbf{x}^0 + h\mathbf{e}_j \in A$, so nennen wir f in $\mathbf{x}^0$ partiell differenzierbar nach x_j, wenn der Grenzwert $\lim_{h\to 0} \frac{f(\mathbf{x}^0 + h\mathbf{e}_j) - f(\mathbf{x}^0)}{h}$ existiert.
Den Grenzwert nennen wir dann j-te partielle Ableitung von f in $\mathbf{x}^0$ und wird bezeichnet durch: $\partial_j f(\mathbf{x}^0)$, $\partial_{x_j} f(\mathbf{x}^0)$, $\frac{\partial f(\mathbf{x}^0)}{\partial x_j}$, $f_{x_j}(\mathbf{x}^0)$, $D_j f(\mathbf{x}^0)$
Randnotiz: Für $\mathbf{x}^0 \in \partial A$ definieren wir $\partial_j f(\mathbf{x}^0)$ so wie oben, bloss mit $h \in [0, \varepsilon)$ bzw. $h \in (\varepsilon, 0]$.

Existieren alle partiellen Ableitungen von f in $\mathbf{x}^0$, so nennen wir den (Zeilen)vektor $\nabla f(\mathbf{x}^0) = (\partial_1 f(\mathbf{x}^0), \ldots, \partial_n f(\mathbf{x}^0))$ Gradient von f in $\mathbf{x}^0$.

Exisitert $\partial_j f(\mathbf{x}^0)$ für alle $\mathbf{x}^0 \in A$, so nennen wir die Funktion $\partial_j f : A \to \mathbb{K}, \mathbf{x} \mapsto \partial_j f(\mathbf{x})$ die j-te partielle Ableitung von f.

Partielle Ableitung berechnen

Um die $\partial_j f(\mathbf{x}^0)$ zu berechnen, falls diese existiert, tun wir so, als seien alle x_i mit $i \neq j$ konstant und berechnen die Ableitung der Funktion
$\phi(x) := f(x_1^0, \ldots, x_{j-1}^0, x, x_{j+1}^0, \ldots, x_n^0)$ in $x = x_j^0$ wie im 1-dimensionalen Fall.
Beispiel:
Sei $f : \mathbb{R}^2 \to \mathbb{R}, f(x_1, x_2) = x_1^2 x_2 e^{2x_1 + x_2} - 3x_2 + x_1$ und $\mathbf{x}^0 = (1, 0)$. Dann ist
$\partial_1 f(x_1, x_2) = 2x_1 x_2 e^{2x_1+x_2} + 2x_1^2 x_2 e^{2x_1+x_2} + 1$
$\partial_2 f(x_1, x_2) = x_1^2 e^{2x_1+x_2} + x_1^2 x_2 e^{2x_1+x_2} - 3$
$\nabla f(\mathbf{x}^0) = (1, e^2 - 3)$

Partielle Ableitungen von $f : \mathbb{R}^n \to \mathbb{K}^m$

Sei $f : A \to \mathbb{K}^m$. Existieren alle partiellen Ableitungen $\partial_j f_l(\mathbf{x}^0)$ für $j \in \{1, \ldots, n\}, l \in \{1, \ldots, m\}$, dann nennen wir die $m \times n$-Matrix

$$\begin{bmatrix} \partial_1 f_1(\mathbf{x}^0) & \ldots & \partial_n f_1(\mathbf{x}^0) \\ \vdots & & \vdots \\ \partial_1 f_m(\mathbf{x}^0) & \ldots & \partial_n f_m(\mathbf{x}^0) \end{bmatrix} = \begin{bmatrix} \nabla f_1(\mathbf{x}^0) \\ \vdots \\ \nabla f_m(\mathbf{x}^0) \end{bmatrix}$$

die Jacobimatrix von f in $\mathbf{x}^0$ und bezeichnen diese mit $J_f(\mathbf{x}^0)$. Für $m = n$ bezeichnen wir $\det(J_f(\mathbf{x}^0))$ als die Jacobideterminante von f in $\mathbf{x}^0$.

Eigenschaften von partiellen Ableitungen

Die j-te partielle Ableitung ist eine lineare Funktion. Somit sind auch der Gradient und die Jacobimatrix lineare Funktionen.
Das heißt $\forall \lambda \in \mathbb{K}, \forall l \in \{1, \ldots, m\}, k \in \{1, \ldots, n\}$:
$\partial_j(\lambda f + g)_l(\mathbf{x}^0) = \lambda \cdot \partial_j(f)_l(\mathbf{x}^0) + \partial_j(g)_l(\mathbf{x}^0)$
$\nabla(\lambda f + g)_l(\mathbf{x}^0) = \lambda \cdot \nabla f_l(\mathbf{x}^0) + \nabla g_l(\mathbf{x}^0)$
$J_{\lambda f + g}(\mathbf{x}^0) = \lambda \cdot J_f(\mathbf{x}^0) + J_g(\mathbf{x}^0)$

Jacobimatrix Beispiel

Die Funktion $f : \mathbb{R}^3 \to \mathbb{R}^2$ sei gegeben durch

$$f(x_1, x_2, x_3) = \begin{pmatrix} x_1^2 + x_2^2 \\ x_3^2 + x_3 \cdot \sin(x_2) \\ x_2 x_3 \end{pmatrix}$$

$$\partial_1 f(x_1, x_2, x_3) = \begin{pmatrix} 2x_1 \\ 0 \\ 0 \end{pmatrix}$$

$$\partial_2 f(x_1, x_2, x_3) = \begin{pmatrix} 2x_2^2 \\ x_3 \cdot \cos(x_2) \\ x_3 \end{pmatrix}$$

$$\partial_3 f(x_1, x_2, x_3) = \begin{pmatrix} 0 \\ 2x_3 + \sin(x_2) \\ x_2 \end{pmatrix}$$

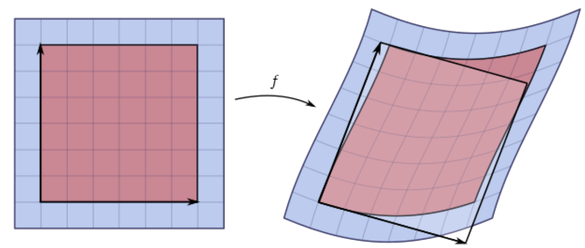
$$J_f(x_1, x_2, x_3) = \begin{pmatrix} 2x_1 & 2x_2^2 & 0 \\ 0 & x_3 \cdot \cos(x_2) & 2x_3 + \sin(x_2) \\ 0 & x_3 & x_2 \end{pmatrix}$$

$$J_f(1, 0, 1) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 1 & 0 \end{pmatrix}, \det(J_f(1, 0, 1)) = -4.$$

Partielle Ableitungen höherer Ordnung

Sei $f : A \to \mathbb{K}^m$ und $k \in \mathbb{N}$.
Für jedes $p = (p_1, \ldots, p_k) \in \{1, \ldots, n\}^k$ nennen wir
$\partial_p f(\mathbf{x}^0) := \frac{\partial^k f(\mathbf{x}^0)}{\partial x_{p_1} \ldots \partial x_{p_k}} := \partial_{p_1} \partial_{p_2} \ldots \partial_{p_k} f(\mathbf{x}^0)$
die partielle Ableitung k-ter Ordnung von f in $\mathbf{x}^0$ nach p (wenn existent). D.h. wir berechnen zuerst die p_k-te partielle Ableitung von f, dann davon die p_{k-1}-te partielle Ableitung, usw. und zum Schluss die p_1-te partielle Ableitung.
f selbst wird auch als partielle Ableitung 0-ter Ordnung bezeichnet.

Idee der Jacobideterminante



Eine nicht-lineare Abbildung $f : \mathbb{R}^2 \to \mathbb{R}^2$ transformiert ein Rechteck (links in rot) zu einer verzerrten Fläche (rechts in rot). Würden wir für das rote Rechteck die Funktionswerte mit Hilfe der Jacobimatrix linear approximieren, dann würden wir das rote Rechteck in das weiße durchsichtige Parallelprogramm überführen. Die Jacobideterminante ist das Verhältnis von der Fläche des ursprünglichen Rechtecks zum weißen durchsichtigen Parallelprogramm und dient, in der Umgebung von dem Punkt in dem wir sie berechnet haben, als Approximation für das Verhältnis von der Fläche des ursprünglichen Rechtecks zur verzerrten roten Fläche.

Satz von Schwarz

Sei $f : A \to \mathbb{K}$. Sind alle partiellen Ableitungen 2-ter Ordnung von f in $\mathbf{x}^0$ stetig, dann gilt:
$\forall i, j \in \{1, \ldots, n\} : \partial_i \partial_j f(\mathbf{x}^0) = \partial_j \partial_i f(\mathbf{x}^0)$

C^k-Funktionen

Sei $f : A \to \mathbb{K}^m$ und $k \in \mathbb{N}_0$. Wir nennen f eine C^k-Funktion, wenn für alle Koordinatenfunktionen f_l alle partiellen Ableitungen der Ordnung 0 bis k von f_l in A existieren und stetig sind.
Die Menge aller $\mathbb{K}^m$-wertigen C^k-Funktionen auf A bezeichnen wir mit $C^k(A, \mathbb{K}^m)$.
Haben alle f_l stetige partielle Ableitungen beliebiger Ordnung, so nennen wir f eine C^∞-Funktion oder glatte Funktion.

Totales Differential/ Totale Ableitung

Sei A offen und $f : A \to \mathbb{R}^m$ und $k \in \mathbb{N}_0$. Wir nennen f differenzierbar in $\mathbf{x}^0$, wenn eine lineare Abbildung $L_{\mathbf{x}^0} : \mathbb{R}^n \to \mathbb{R}^m$ existiert, so dass
$\lim\limits_{\mathbf{h}\to 0} \frac{f(\mathbf{x}^0+\mathbf{h})-f(\mathbf{x}^0)-L_{\mathbf{x}^0}(\mathbf{h})}{\|\mathbf{h}\|} = \mathbf{0}$
Ist f differenzierbar in $\mathbf{x}^0$, so nennen wir L das totales Differential oder totale Ableitung von f in $\mathbf{x}^0$ und wir schreiben dann $Df(\mathbf{x}^0)$ statt L.

Idee/Intuition Totales Differential

Im eindimensionalen haben wir die Ableitung bestimmt, indem wir zwischen $f(x^0)$ und $f(x^0+h)$ eine Linie/Gerade (lineare Funktion) zeichnen und davon die Steigung bestimmen. Die Ableitung ist dann diese Steigung für $h \to 0$. Im mehrdimensionalen reicht eine Gerade dafür nicht mehr aus, da es nun unendlich viele Punkte um $\mathbf{x}^0$ gibt, die Abstand $\|\mathbf{h}\|$ zu $\mathbf{x}^0$ haben. Das heißt wir brauchen eine Verallgemeinerung von "Linie"/"Gerade". Diese wäre im mehrdimensionalen eine lineare Abbildung L für die gilt, dass für $\mathbf{h} \approx \mathbf{0}$ gilt: $f(\mathbf{x}^0+h) \approx f(\mathbf{x}^0) + L(\mathbf{h})$. Von der linearen Abbildung L können wir analog zu der Gerade im eindimensionalen die "Steigung"/Ableitung in alle Richtungen bestimmen. Wenn wir also den Grenzwert dazu bilden und dieser existiert, dann nennen wir diesen die totale Ableitung von f in $\mathbf{x}^0$.
Geometrisch bedeutet differenzierbar, so wie im eindimensionalen, dass f keine Sprünge und keine Knicke hat. Bei einer mehrdimensionalen Funktion kann man sich "Knick"als Falte oder Zacken vorstellen. Die Existenz der partiellen Ableitungen bedeutet nur, dass es keine Knicke in Richtung der Achsen gibt. Dennoch kann es in eine andere Richtung betrachtet Knicke geben. Deswegen folgt aus der Existenz der partiellen Ableitungen in $\mathbf{x}^0$ **nicht**, dass f in $\mathbf{x}^0$ differenzierbar ist.

Totales Differential Satz

Sei $f : A \to \mathbb{K}^m$ differenzierbar in $\mathbf{x}^0$. Dann ist f stetig in $\mathbf{x}^0$, alle $\partial_k f_l(\mathbf{x}^0)$ existieren und es gilt $Df(\mathbf{x}^0) = J_f(\mathbf{x}^0)$. Im Fall von $m = 1$ gilt dann also $Df(\mathbf{x}^0) = \nabla f(\mathbf{x}^0)$

Totales Differential Satz

Sei $f : A \to \mathbb{K}^m$. Existieren alle $\partial_j f_l$ auf ganz A und sind diese stetig in $\mathbf{x}^0$, so ist f differenzierbar in $\mathbf{x}^0$.

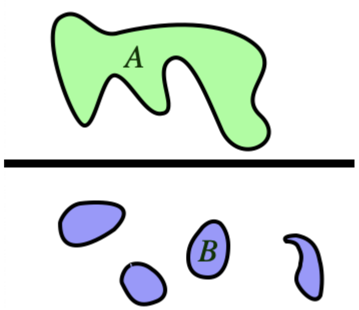
Kettenregel

Sei $A_f \subseteq \mathbb{R}^n$ offen und $f : A_f \to \mathbb{R}^m$.
Sei $f(A_f) \subseteq A_g \subseteq \mathbb{R}^m$ offen und $g : A_g \to \mathbb{K}^p$. Ist f differenzierbar in $\mathbf{x}^0$ und g differenzierbar in $f(\mathbf{x}^0)$, so ist $g \circ f : A_f \to \mathbb{K}^p$ differenzierbar in $\mathbf{x}^0$ und es gilt die Kettenregel:
$D(g \circ f)(\mathbf{x}^0) = Dg(f\mathbf{x}^0)) \circ Df(\mathbf{x}^0)$
Insbesondere gilt, wenn f und g differenzierbar sind, so ist auch $g \circ f$ differenzierbar.

Weg-zusammenhängend

Wir nennen A weg-zusammenhängend, wenn für alle $\mathbf{x}, \mathbf{y} \in A$ eine stetige Funktion $p : [0,1] \to A$ existiert mit $p(0) = x$ und $p(1) = y$.
Ist p auch noch differenzierbar, dann nennen wir A differenzierbar weg-zusammenhängend.

Wegzusammenhängend



Die Menge A ist weg-zusammenhängend, weil du alle Punkte aus der Menge mit einem stetigen Pfad innerhalb der Menge verbinden kannst. Die Menge B ist nicht weg-zusammenhängend, da es keinen stetigen Pfad von einem Punkt aus der ersten "Bubble"zu einem Punkt in der letzten "Bubble"gibt.

Konstante Funktion

Ist A differenzierbar weg-zusammenhängend und $f : A \to \mathbb{K}$ differenzierbar mit $\nabla f \equiv 0$, dann ist f konstant.