Ableitungsbegriff und Taylorentwicklung von (vektorwertigen) Funktionen (in mehreren Veränderlichen)

Im folgenden betrachten wir Funktionen, die Punkte (Vektoren) auf Punkte (Vektoren) abbilden, beschreiben den Ableitungsbegriff und definieren in den ersten beiden Fällen eine Taylorentwicklung.
Die Definitionen sollen der gedanklichen Strukturbildung von VWL-Studenten helfen und sind demzufolge an manchen Stellen etwas ungenau.

$f:\mathbb{R}\to\mathbb{R}$ $x\mapsto y:=f(x)$ Ein Punkt x wird auf einen Wert $y:=f(x)$ abgebildet.	$f:\mathbb{R}^n\to\mathbb{R}$ $\boldsymbol{x}\mapsto y:=f(\boldsymbol{x})$ Ein Vektor $\boldsymbol{x}$ wird auf einen Wert $y:=f(\boldsymbol{x})$ abgebildet, z. B. zwei Koordinaten werden einer Höhe (= Landkarten), drei Koordinaten werden einer Farbe (= Temperaturkarten) zugeordnet.	$\boldsymbol{f}:\mathbb{R}^n\to\mathbb{R}^m$ $\boldsymbol{x}\mapsto \boldsymbol{y}:=\boldsymbol{f}(\boldsymbol{x})$ Ein Vektor $\boldsymbol{x}$ wird auf einen Vektor $\boldsymbol{y}:=\boldsymbol{f}(\boldsymbol{x})$ abgebildet, z. B. Strömungen im Schwimmbad.
Die Ableitung von f an einem Punkt x_0 ist durch folgenden Grenzwert (falls er existiert) definiert: $$Df(x_0):=\frac{df(x)}{dx}\,\Big\vert_{x=x_0}:=f'(x_0):=\lim\frac{f(x_0+h)-f(x_0)}{h}$$ Die Ableitung an diesem Punkt ist ein Skalar, den man graphisch als Tangentensteigung an x_0 illustrieren kann. Man kann die Funktion zweimal ableiten (falls beide Grenzwerte existieren) und erhält die zweite Ableitung an einem Punkt, welche wieder ein Skalar ist. **Wichtig:** Im $\mathbb{R}^1\to\mathbb{R}^1$-Fall korrespondiert eine Veränderung der x-Variable mit **einer** Veränderung der y-Variable. Diese Änderungsrate ist gerade die Ableitung. Im $\mathbb{R}^n\to\mathbb{R}$ gibt es n-Veränderungen der $\boldsymbol{x}$-Variable und dementsprechend n Ableitungen, im $\mathbb{R}^n\to\mathbb{R}^m$ gibt es für n Veränderungen der $\boldsymbol{x}$ Variable gerade m Veränderungen der $\boldsymbol{y}$ Variable und dementsprechend eine $m\times n$-Matrix als Ableitung.	Motiviert durch die erste Ableitung als Steigung, definiert man im Fall von mehreren Veränderlichen $\boldsymbol{x}$ die partielle Ableitung (bzw. interpretiert als Steigung in Richtung x_i): $$\frac{\partial f(x)}{\partial x_i}\,\Big\vert_{\boldsymbol{x}_0}:=\partial_i f(\boldsymbol{x}_0):=f_{x_i}(\boldsymbol{x}_0):=\lim\frac{f(x_1,...,x_i+h,...x_n)-f(\boldsymbol{x})}{h}$$ Man bekommt also für jede Variable eine partielle Ableitung, welche man als Gradient zusammenfasst: $$\nabla f(\boldsymbol{x}_0):=\left(\frac{\partial f(\boldsymbol{x}_0)}{x_1},\quad\cdots\quad,\frac{\partial f(\boldsymbol{x}_0)}{x_n}\right)=\left(\partial_1 f(\boldsymbol{x}_0),\quad\cdots\quad,\partial_n f(\boldsymbol{x}_0)\right)$$ Dies ist ein $1\times n$-Vektor (je nach Literatur auch ein $n\times 1$-Vektor) und kann als Richtungsvektor des maximalen Anstiegs interpretiert werden. Während im einfachen Fall der Ableitungswert an einem Punkt ein Skalar war, ist es hier ein Vektor. Will man nun ausgehend von einem Punkt $\boldsymbol{x}_0$ die Steigung in die Richtung $\boldsymbol{vz}$ bestimmen, so berechnet man einfach $\nabla f(\boldsymbol{x}_0)\cdot\boldsymbol{vz}$ und erhält wieder ein Skalar, die Tangentensteigung in diese Richtung. Analog zum ersten Fall möchte man höhere Ableitungen bilden, dabei benötigt man aber zwischen allen möglichen Steigungen Krümmungen und definiert deshalb die $n\times n$-Hesse-Matrix von f an der Stelle $\boldsymbol{x}_0$: $$\mathrm{H}_f(\boldsymbol{x}_0):=\left[\left(\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i\partial x_j}\right)_{i,j=1,\dots,n}\right]_{\boldsymbol{x}=\boldsymbol{x}_0}:=\begin{pmatrix}\frac{\partial^2 f(\boldsymbol{x})}{\partial x_1\partial x_1} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1\partial x_n}\\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2\partial x_1} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2\partial x_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n\partial x_1} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n\partial x_n}\end{pmatrix}_{\boldsymbol{x}=\boldsymbol{x}_0}$$	Während wir im $\mathbb{R}^1\to\mathbb{R}^1$ ein Skalar (einen 1×1-Vektor) als Ableitung definierten, im $\mathbb{R}^n\to\mathbb{R}^1$ einen Vektor, den $1\times n$-Gradienten definierten, wollen wir dies nun konsistent für den $\mathbb{R}^n\to\mathbb{R}^m$-Fall fortführen und eine $m\times n$-Matrix definieren: $\boldsymbol{f}(\boldsymbol{x}_0)$ ist an der Stelle $\boldsymbol{x}_0$ differenzierbar, falls es eine lineare Abbildung $L:\mathbb{R}^n\to\mathbb{R}^m$ gibt, so dass gilt: $$\lim_{\boldsymbol{h}\to\boldsymbol{0}}\frac{\boldsymbol{f}(\boldsymbol{x}_0+\boldsymbol{h})-\boldsymbol{f}(\boldsymbol{x}_0)-L\cdot\boldsymbol{h}}{\vert\vert\,\boldsymbol{h}\,\vert\vert}=0$$ Diese lineare Abbildung kann nach Basiswahl als Matrix dargestellt werden; in der Standardbasis ist es gerade die Jacobimatrix im Punkt $\boldsymbol{x}_0$: $$\boldsymbol{J_f}(\boldsymbol{x}_0):=\begin{pmatrix}\frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial f_m(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n}\end{pmatrix}_{\boldsymbol{x}=\boldsymbol{x}_0}$$ Dies ist eine $m\times n$ Matrix, im Fall $m=1$ gerade der Gradient.
Optimalitätsbedingungen für x_0 lok. Max. bzw. lok. Min.: Notwendig: $f'(x_0)=0$ Hinreichend: $f''(x_0)\underset{(<)}{>}0\vee f'''(x_0)\underset{(<)}{>}0\vee ...$: x_0 lok. Minimum (Maximum) Man sucht also zuerst die Punkte, an denen die Funktion weder steigt, noch fällt und untersucht die Krümmung an diesem Punkt. Ist die Krümmung konvex, d. h. $f''(x_0)>0$, so liegt hier ein lokales Minimum.	Optimalitätsbedingungen für $\boldsymbol{x}_0$ lok. Max. bzw. lok. Min.: Notwendig: $\nabla f(\boldsymbol{x}_0)=\boldsymbol{v}0\iff\partial_i f(\boldsymbol{x}_0)=0\quad\forall i=1,...,n$ Hinreichend: $\mathrm{H}_f(\boldsymbol{x}_0)$ positiv (negativ) definit: $\boldsymbol{x}_0$ lok. Minimum (Maximum) Wie im einfachen Fall gibt die zweite Ableitung die Auskunft über die Art des Extremas. Hier kann man aber nicht mehr von einer „positiven Matrix" sprechen, sondern muss die Definitheit, für die es mehrere Kriterien gibt, der Matrix testen.	Da $\mathbb{R}^m$ keine Ordnungsrelation mehr besitzt, d.h. man zwei Vektoren $\boldsymbol{vv},\boldsymbol{vw}\in\mathbb{R}^m$ nicht mehr vergleichen kann ($\boldsymbol{vv}\overset{?}{\leq}\boldsymbol{vw}$), macht es hier nicht viel Sinn von Maxima und Minima zu sprechen.
Taylorpolynom von f erzeugt an der Stelle x_0 vom Grad k : $$T_f^k(x;x_0):=\sum_{n=0}^{k}\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$ Ausgeschrieben für $k=2$: $T_f^k(x;x_0)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2=c+bx+ax^2$	Wir beschränken uns hier auf die für die Vorlesung notwendige Taylorapproximation zweiten Grades von f an der Stelle $\boldsymbol{x}_0$: $$T_f^2(\boldsymbol{x},\boldsymbol{x}_0):=f(\boldsymbol{x}_0)+\nabla f(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^T\cdot\mathrm{H}_f(\boldsymbol{x}_0)\cdot(\boldsymbol{x}-\boldsymbol{x}_0)$$ Wie im einfachen Fall ist dies eine Funktion in der Variablen $\boldsymbol{x}$. Analog zum quadratischen Polynom links ist dies eine quadratische Form.	Da eine Taylorentwicklung trotzdem von Interesse sein kann, definiert man die Taylorentwicklung 1. Grades von $\boldsymbol{f}$ an der Stelle $\boldsymbol{x}_0$ durch: $$\boldsymbol{T_f^1}(\boldsymbol{x};\boldsymbol{x}_0)=\boldsymbol{f}(\boldsymbol{x}_0)+\boldsymbol{J_f}(\boldsymbol{x}_0)\cdot(\boldsymbol{x}-\boldsymbol{x}_0)$$ Achtung bei den Dimensionen!