## 4 . Taylorentwicklung; Differential; Gradient; Hesse-Matrix

Die Taylorentwicklung für eine Funktion f(x) ist bis zur 2. Ordnung gegeben durch:

$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2$ + Rest; Entwicklungsstelle $x_0$ .

Wir setzen die Existenz aller erforderlichen Ableitungen natürlich voraus.

$T(x) = f(x_0) + f'(x_0)(x - x_0)$ ist das Taylorpolynom 1. Ordnung. Die graphische Darstellung dieses Polynoms, also der Funktion $y = T(x)$, ist die Tangente an den Graphen von $y = f(x)$ im Punkt $\tilde{P}_0 = (x_0, y = f(x_0))$.

$df = f'(x_0)(x - x_0)$ ist das Differential (an der Stelle $x_0$).

Es gilt in der Nähe von $x_0$:

(a) $f(x) \approx T(x)$ und

(b) $\Delta f = f(x) - f(x_0) \approx df$ .

**Entsprechendes gilt für eine Funktion f(x, y)**, Entwicklungsstelle $(x_0, y_0)$:
ihre Taylorentwicklung bis zur 2. Ordnung ist gegeben durch

$$f(x,y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

$$+ \frac{1}{2!}\{ f_{xx}(x_0, y_0)(x - x_0)^2 + 2 f_{xy}(x_0, y_0)(x - x_0)(y - y_0) + f_{yy}(x_0, y_0)(y - y_0)^2 \}$$

$$+ \text{Rest} .$$

Es treten jetzt natürlich partielle Ableitungen auf, deren **Existenz und Stetigkeit** vorausgesetzt wird. Im { } - Term gilt dann die Gleichheit der gemischt-partiellen Ableitungen: $f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0)$, daher der Faktor 2. Wir wollen diesen Term, der auch einige Quadrate enthält, mit QF (für: **quadratische Form**) benennen.

Die Taylorentwicklung kann kompakt geschrieben werden, wenn man die folgenden Abkürzungen einführt: $\vec{r} = (x, y)$, $\vec{a} = (x_0, y_0)$, $\Delta x = x - x_0$, $\Delta y = y - y_0$ .

Damit erhält man:

$$f(\vec{r}) \approx f(\vec{a}) + f_x(\vec{a})\,\Delta x + f_y(\vec{a})\,\Delta y + \frac{1}{2!}\{ f_{xx}(\vec{a})\,(\Delta x)^2 + 2 f_{xy}(\vec{a})\,\Delta x\,\Delta y + f_{yy}(\vec{a})\,(\Delta y)^2 \} .$$

**<u>Lineare Approximation</u>**

$TE(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$ ist das Taylorpolynom 1.Ordnung für die Funktion f(x, y) .

Die graphische Darstellung dieses Polynoms, also der Funktion z = TE(x, y), ist die **Tangentialebene** an den Graphen von f(x, y) im Punkt $\tilde{P}_0 = (x_0, y_0, z_0 = f(x_0, y_0))$ .

$df = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$ heißt <u>vollständiges</u> oder **<u>totales</u> Differential** <u>an der Stelle</u> $(x_0, y_0)$**.** Die einzelnen Summanden in df werden auch als <u>partielle</u> Differentiale bezeichnet. $\Delta x = x - x_0$ und $\Delta y = y - y_0$ heißen Zuwächse (auch wenn sie negativ sind) oder <u>unabhängige</u> Differentiale und werden oft als dx und dy geschrieben. Es handelt sich bei diesen unabhängigen Differentialen um endliche Größen!

Es gilt in der Nähe von $\vec{a} = (x_0, y_0)$ :

(a) $f(x, y) \approx TE(x, y)$ ;

Das Taylorpolynom erster Ordnung ist die **lineare Approximation** für f(x, y): die Funktion f(x, y) wird näherungsweise durch die (einfachere) <u>lineare</u> Funktion TE(x, y) ersetzt.

(b) $\Delta f = f(x, y) - f(x_0, y_0) \approx f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) = df$ ;

Wenn man mit den Argumenten der Funktion von $(x_0, y_0)$ nach (x, y) geht, ändert sich der Funktionswert um $\Delta f = f(x, y) - f(x_0, y_0)$ . <u>Die Veränderung des Funktionswertes kann näherungsweise (**aber einfacher**) durch den linearen Ausdruck df berechnet werden</u>. Das Differential df ist die <u>lineare Approximation</u> für $\Delta f$ . Dies wird z.B. bei der **Fehlerrechnung** benutzt.

Mit den obigen Abkürzungen schreibt sich das Differential etwas knapper als:

$df = f_x(\vec{a})\,\Delta x + f_y(\vec{a})\,\Delta y$ . Es zeigt noch einmal deutlich, welche unabhängige Variable „den größten Einfluss" hat: wenn $\Delta x = \Delta y$ , dann bestimmt die partielle Ableitung, welches partielle Differential (betragsmäßig) größer ist.

Das Differential hängt von der Entwicklungsstelle $(x_0, y_0)$ ab. Genau genommen müsste man also schreiben: $df(x_0, y_0)$ , oder, wenn die Stelle mit $P_0$ bzw. $\vec{a}$ bezeichnet wird, $df(P_0)$ bzw. $df(\vec{a})$.
Noch präziser ist die Notation: $df(x_0, y_0, \Delta x, \Delta y)$. Denn erst wenn man für alle vier Größen Zahlenwerte einsetzt, wird auch das Differential eine Zahl. Lässt man (ähnlich wie bei den partiellen Ableitungen) den Index weg, erhält man das Differential an einer beliebigen Stelle: $df(x, y, \Delta x, \Delta y)$ . $\Delta x$ und $\Delta y$ sind dann beliebige Zuwächse (besser: Veränderungen) der unabhängigen Variablen.

## Gradient

Das vollständige Differential kann noch kompakter geschrieben werden, wenn man die folgenden Vektoren einführt: den **Gradienten** $\nabla f = (f_x, f_y)$ und den (Verschiebungs-) Vektor $\Delta\vec{r} = (\Delta x, \Delta y)$. Dann gilt: $df = \nabla f(\vec{a}) \cdot \Delta\vec{r}$ .
Statt $\Delta\vec{r}$ wird auch oft $d\vec{r} = (dx, dy)$ geschrieben (s.o.).

Die Auswertung des Skalarproduktes ergibt: $df = |\nabla f(\vec{a})|\,|\Delta\vec{r}|\cos(\varphi)$ .
Man sieht: das Differential ist - bei festem Betrag des Verschiebungsvektors - dann am größten, wenn man den Verschiebungsvektor parallel zum Gradienten legt ( $\varphi = 0$ ). Der Gradient zeigt also in diejenige Richtung, in der die Funktion am stärksten wächst: **der Gradient zeigt in die Richtung des steilsten Anstiegs.**
Daraus ergibt sich eine Anwendung bei **Optimierungsfragen** („in welche Richtung muss ich gehen, damit sich meine Zielfunktion optimal entwickelt?").

Dabei ist zu beachten, dass der Gradient und der Verschiebungsvektor zweikomponentige Vektoren sind: sie liegen also in der x-y-Ebene. Die Richtung des steilsten Anstiegs ist also eine Richtung in der Ebene der unabhängigen Variablen.

Bewegt man sich längs einer Höhenlinie, ändert sich f(x, y) nicht.
A und B sollen zwei Punkte auf derselben Höhenlinie sein: $\Delta f$ = f(B) – f(A) = 0.
Für das Differential wird dann gelten: df $\approx$ 0.

Der Verschiebungsvektor $\Delta\vec{r}$ geht von A nach B. Er liegt auf der Sekante, die durch diese beiden Punkte geht. $\Delta f$ bleibt Null, wenn B → A (Grenzprozess). Im Grenzfall gilt dann auch df **=** 0 . Der Verschiebungsvektor $\Delta\vec{r}$ liegt dann tangential zur Höhenlinie. Da das Differential 0 ist, muss der Gradient senkrecht auf (dem tangentialen) $\Delta\vec{r}$ stehen: nur dann hat das Skalarprodukt df = $\nabla f(\vec{a}) \cdot \Delta\vec{r}$ den Wert 0!
Kurz: **der Gradient** $\nabla f(\vec{a})$ **steht senkrecht „auf der Höhenlinie“** durch $\vec{a}$ .

Mit Hilfe des Gradienten kann auch die Ableitung an der Stelle $\vec{a}$ in eine beliebige Richtung berechnet werden. Die Richtung sei durch den entsprechenden Einheitsvektor $\vec{e}$ gegeben. Diese sog. **Richtungsableitung** in Richtung von $\vec{e}$ ist gegeben durch das Skalarprodukt $\nabla f(\vec{a}) \cdot \vec{e}$ . Sie ist – genauso wie das Differential – dann am größten, wenn $\vec{e}$ in Richtung des Gradienten, also in Richtung des steilsten Anstiegs geht.

Die partiellen Ableitungen sind spezielle Richtungsableitungen, nämlich in Richtung der kartesischen Einheitsvektoren $\vec{e}_x$ bzw. $\vec{e}_y$ .

## Hesse-Matrix

Bis jetzt haben wir nicht zwischen Zeilen- und Spaltenvektoren unterschieden.
Da im Folgenden das Produkt „Matrix · Vektor“ auftritt, muss dies aber aus formalen Gründen geschehen (s. Matrizenrechnung).

Der „quadratische“ Term QF kann ebenfalls kompakt geschrieben werden:
QF = $\frac{1}{2!}$ $(\Delta\vec{r})^{\mathbf{t}} \cdot H(\vec{a}) \cdot (\Delta\vec{r})$. H ist die sog. **Hesse-Matrix**: H = $\begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix}$ .

$(\Delta\vec{r})^{\mathbf{t}}$ ist ein **Zeilen**vektor. Er geht aus dem **Spalten**vektor $\Delta\vec{r}$ durch Transponieren hervor.

In $H(\vec{a})$ werden die zweiten partiellen Ableitungen an der Stelle $\vec{a} = (x_0, y_0)$ ausgewertet. Man kann die Hesse-Matrix H natürlich auch an einer beliebigen Stelle $P = (x, y)$ bestimmen.

**Bem.**

1) Für skalarwertige Funktionen mit mehr als zwei Veränderlichen gelten analoge Formeln.

2) Das Differential spielt auch bei wirtschaftswissenschaftlichen Untersuchungen eine Rolle (s. z.B. Tietze, Einführung in die angewandte Wirtschaftsmathematik).

3) Wenn TE(x, y) die Funktion f(x, y) in der Nähe von $\vec{a} = (x_0, y_0)$ linear approximiert, dann heißt f(x, y) **total differenzierbar**. Oben haben wir das garantiert, indem wir die Existenz und Stetigkeit der partiellen Ableitungen vorausgesetzt haben. Es gibt Funktionen, wo die partiellen Ableitungen und damit TE(x, y) existieren, aber TE(x, y) nicht die Funktion approximiert. TE(x, y) kann dann auch nicht als Tangentialebene bezeichnet werden.

4) Für eine vektorwertige Funktion wird der Gradient durch die sog. **Jacobi-Matrix** J ersetzt. Für eine Funktion $\vec{f}(x, y) = ( f_1(x,y) , f_2(x,y) )$, also mit zwei unabhängigen Variablen und zwei (Vektor-) Komponenten sieht J so aus: $\begin{pmatrix} f_{1,x} & f_{1,y} \\ f_{2,x} & f_{2,y} \end{pmatrix}$ .

Der erste Index bezeichnet die Vektorkomponente, der zweite die Variable, nach der abgeleitet wird.