

# Thema 12—Differentialrechnung, Partielle Ableitungen, Differenzierbarkeit, Taylor-Formel, Lokale Extrema

In diesem Kapitel befassen wir uns mit der Ableitung von Funktionen  $f : \mathbf{R}^m \rightarrow \mathbf{R}^n$ . Allein die Schreibweise liefert Probleme in diesem Fall. Betrachten wir z.B. ein Polynom in zwei Variablen, etwa

$$p(x, y) = 13 + 5xy + 3xy^3 - x^7y$$

oder in drei Variablen:

$$7 + 2z + 3xy + yz + xz - 17x^2yz.$$

Wie soll man aber das *allgemeine* Polynom in zwei Variablen schreiben? Versuche wie

$$a + bx + cy + dx^2 + exy + fy^2 + \dots$$

sind offensichtlich unbrauchbar. Besser ist

$$a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2 + \dots$$

oder

$$\sum_{k,l} a_{kl}x^k y^l.$$

Falls aber die Zahl der unabhängigen Variablen groß ist, dann ist diese Schreibweise auch unbrauchbar. Wir verwenden eine, die von L. Schwartz eingeführt wurde, wobei er den Begriff eines Multi-Index verwendete.

**Schreibweise.** Wir schreiben  $x = (x_1, \dots, x_n)$  für ein Element von  $\mathbf{R}^n$ . Ein **Multi-Index** ist ein  $n$ -Tupel  $p = (p_1, \dots, p_n)$ , wobei jedes  $p_i \in \mathbf{N}_0$ .  $x^p$  bedeutet dann  $x_1^{p_1} \dots x_n^{p_n}$ . Wir schreiben auch

$$\begin{aligned} p! &= p_1! \dots p_n! \\ |p| &= p_1 + \dots + p_n \\ \binom{p}{k} &= \frac{p!}{k!(p-k)!} = \binom{p_1}{k_1} \dots \binom{p_n}{k_n} \\ p \leq q &\iff p_i \leq q_i \text{ für jedes } i. \end{aligned}$$

Das allgemeine Polynom läßt sich also wie folgt schreiben:

$$\sum_{|p| \leq k} a_p x^p.$$

Als ein Beispiel der Eleganz dieser Schreibweise formulieren wir den binomischen Lehrsatz wie folgt:

Seien  $x, y \in \mathbf{R}^n$ ,  $p \in (\mathbf{N}_0)^n$  ein Multi-Index. Dann gilt

$$(x + y)^p = \sum_{0 \leq k \leq p} \binom{p}{k} x^k y^{p-k}.$$

Versuche, diesen Lehrsatz mit einer anderen Schreibweise zu formulieren, werden den Leser bald von der Unentbehrlichkeit der Multi-Indizes überzeugen.

BEISPIEL. (Multinominalsatz.)  $(x_1 + \dots + x_n)^k = \sum_{|p|=k} \frac{k!}{p!} x^p$ .  
(Induktionsbeweis.)

Um Stetigkeit von Funktionen von  $U(\subseteq \mathbf{R}^m)$  in  $\mathbf{R}^n$  zu definieren, verwenden wir den Abstands begriff in  $\mathbf{R}^m$ . Wir setzen

$$\begin{aligned} |x| &= (\xi_1^2 + \dots + \xi_n^2)^{\frac{1}{2}} \quad (x = (\xi_1, \dots, \xi_n)) \\ d(x, y) &= |x - y| = \sqrt{(\xi_1 - \eta_1)^2 + \dots + (\xi_n - \eta_n)^2}. \end{aligned}$$

$f$  heißt **stetig** im Punkt  $x$ , falls für jedes  $\epsilon > 0$  existiert ein  $\delta > 0$ , sodaß aus  $|y - x| < \delta$  folgt  $|f(x) - f(y)| \leq \epsilon$ .

## Beispiele von Funktionen mehrerer Variablen

### I. Lineare Funktionale:

$$\begin{aligned} m = 2, n = 1 : f(x, y) &= ax + by + c \\ m = 3, n = 1 : f(x, y) &= ax + by + cz + d. \end{aligned}$$

$$\text{allg.: } f(x_1, \dots, x_m) = a_1x_1 + a_2x_2 + \dots + a_mx_m + d.$$

### II. Quadratische Funktionale:

$$\begin{aligned} m = 2 : f(x, y) &= ax^2 + 2bxy + cy^2 \quad (\text{insbesondere: } x^2 + y^2, x^2 - y^2, xy, x^2) \\ m = 3 : f(x, y) &= ax^2 + by^2 + cz^2 + dyz + ezx + fxy. \end{aligned}$$

$$\text{allg.: } f(x_1, \dots, x_m) = \sum_{i,j=1}^m a_{ij}x_ix_j.$$

### III. Funktionen, die als Kombinationen von elementaren Funktionen gebildet werden:

$$\begin{aligned} f(x, y) &= \left( \sqrt{x^2 + y^2}, \frac{1}{\sqrt{x^2 + y^2}}, \sin(x + y), \sin(x^2 + xy + y^2) \right) \\ f(x, y, z) &= (\ln(x^2 + y^2 + z^2), e^{x+y} \sin(|y + z|)) \end{aligned}$$

usw.

DEFINITION. Sei  $f : U \rightarrow \mathbf{R}$ , wobei  $U$  offen in  $\mathbf{R}^n$  ist. Dann ist  $f$  im Punkt  $x_0$  **partiell-differenzierbar bzgl. der  $i$ -ten Koordinatenrichtung**, falls

$$D_i f(x) = \lim_{t \rightarrow 0} \frac{f(x + te_i) - f(x)}{t}$$

existiert. ( $e_i$  ist das  $i$ -te Basiselement  $(0, \dots, 0, 1, 0, \dots, 0)$ , wobei die 1 in der  $i$ -ten Stelle ist). Der Limes ist dann die  **$i$ -te partielle Ableitung** von  $f$  im Punkt  $x$ . Falls jede partielle Ableitung  $D_i f$  in jedem Punkt aus  $U$  existiert, dann ist  $f$  **partiell differenzierbar** und die Funktionen

$$x \mapsto D_i f(x)$$

sind die **partiellen Ableitungen** von  $f$ . Falls diese Ableitungen stetig sind, dann heißt  $f$  **stetig partiell differenzierbar**. (Andere Schreibweise  $\frac{\partial f}{\partial x_i}$  für  $D_i f$ .) Für eine Funktion

$f$  von zwei bzw. drei Variablen also  $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$  bzw.  $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$ .

Falls  $f = (f_1, \dots, f_m) : U \rightarrow \mathbf{R}^m$  so ist, daß jedes  $f_i$  (stetig) partiell differenzierbar ist, dann sagen wir, daß  $f$  (stetig) partiell differenzierbar ist.

DEFINITION. Sei  $f : U \rightarrow \mathbf{R}$  partiell differenzierbar. Der Vektor

$$\text{grad } f(x) = (D_1 f(x), \dots, D_n f(x))$$

heißt der **Gradient** von  $f$  in  $x$ . (In Dimensionen 2 bzw. 3,  $\text{grad } f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$  bzw.  $\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$ .)

**Richtungsableitung:** Falls  $f : U \rightarrow \mathbf{R}$  stetig partiell-differenzierbar ist und  $v$  ein Vektor aus  $\mathbf{R}^n$ , dann ist

$$D_v f(x) = (\text{grad } f(x)|v)$$

die **Richtungsableitung** von  $f$  in Richtung  $v$ . Die Bezeichnung stammt von der Tatsache, daß  $D_v f(x)$  die Ableitung der Funktion

$$t \mapsto f(x + tv)$$

einer Variable an der Stelle 0 ist. Das folgt aus dem nächsten Satz:

**Satz 1** Sei  $c$  eine glatte Kurve in  $U$  und  $f : U \rightarrow \mathbf{R}$  eine stetig partiell differenzierbare Funktion. Dann gilt:

$$\frac{d}{dt} f(c(t)) = (c'(t)|\text{grad } f(c(t))).$$

Dies folgt aus der Kettenregel (siehe unten).

Daraus sieht man (mit Hilfe der Cauchy-Schwarz Ungleichung), daß der Gradientenvektor  $\text{grad } f(x)$  folgende geometrische Bedeutung hat: Er ist die Richtung des *steilsten Zuwachses* der Funktion  $f$ .

**Satz 2** Sei  $f : U \rightarrow \mathbf{R}$  stetig partiell differenzierbar, wobei  $U$  eine Umgebung von  $x$  ist. Dann gilt für  $h \in \mathbf{R}^n$  mit  $|h| < \epsilon$ , wobei  $U(x, \epsilon) \subset U$ ,

$$f(x + h) = f(x) + (\text{grad } f(x)|h) + \rho(h),$$

wobei  $\lim_{h \rightarrow 0} \frac{\rho(h)}{|h|} \rightarrow 0$ .

Wir führen jetzt einen abstrakten Differentiationsbegriff ein, der eine direkte Übertragung der entsprechenden Definition für den eindimensionalen Fall ist. Unten werden wir die Beziehung zwischen dieser Definition und partieller Differenzierbarkeit untersuchen.

DEFINITION. Sei  $U \subset \mathbf{R}^n$  offen und  $f : U \rightarrow \mathbf{R}^m$  eine Abbildung.  $f$  heißt in  $x \in U$  (total) **differenzierbar**, falls es eine lineare Abbildung

$$A : \mathbf{R}^n \rightarrow \mathbf{R}^m$$

gibt, so daß

$$f(x + h) = f(x) + Ah + |h|\varphi(h) \text{ mit } \lim_{h \rightarrow 0} \varphi(h) = 0.$$

Wir schreiben  $(Df)_x$  für  $A$ —die **Ableitung** von  $f$  an der Stelle  $x$ . Falls  $(Df)_x$  existiert für jedes  $x \in U$ , dann heißt  $f$  **auf  $U$  differenzierbar**. Die Abbildung  $Df : x \mapsto (Df)_x$  (von  $U$  in  $M_{m,n}$ , die Familie aller  $m \times n$  Matrizen) ist dann die **Ableitung** von  $f$ . Falls  $Df$  stetig ist, dann ist  $f$  **stetig differenzierbar**.

BEMERKUNG. Ist  $U \subset \mathbf{R}^n$  offen und  $f : U \rightarrow \mathbf{R}^n$ , dann ist  $f$  genau dann differenzierbar, wenn alle  $f_i : U \rightarrow \mathbf{R}$  ( $i = 1, \dots, m$ ) differenzierbar sind, wobei  $f = (f_1, \dots, f_n)$ .

Im folgenden Satz untersuchen wir die Beziehung zwischen Differenzierbarkeit und der Existenz von partiellen Ableitungen.

**Satz 3** Sei  $U \subset \mathbf{R}^n$  offen und  $f : U \rightarrow \mathbf{R}^m$ .

i)  $f$  sei differenzierbar in  $x \in U$ . Dann gilt: Alle  $f_i$  sind in  $x$  partiell differenzierbar mit

$$\frac{\partial f_i}{\partial x_j}(x) = a_{ij} \quad (i = 1, \dots, m; j = 1, \dots, n),$$

wobei  $[a_{ij}]$  die Matrix von  $(Df)_x$  ist;

ii) Seien die  $f_i$  stetig partiell differenzierbar in einer Umgebung von  $x$  und

$$a_{ij} := \frac{\partial f_i}{\partial x_j}(x) \text{ f\"ur } i = 1, \dots, m \text{ und } j = 1, \dots, n.$$

Dann ist  $f$  in  $x$  differenzierbar und es gilt

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

wobei

$$A = (a_{ij}) \text{ und } \lim_{h \rightarrow 0} \varphi(h) = 0.$$

**Bezeichnung.** Sei  $f : U \rightarrow \mathbf{R}^m$  differenzierbar in  $x$ . Dann heit die  $(m \times n)$ -Matrix

$$(Df)(x) = J_f(x) = \left( \frac{\partial f_i}{\partial x_j}(x) \right)$$

die **Funktionalmatrix** oder **Jacobi-Matrix** von  $f$  in  $x$ .

**Satz 4 (Kettenregel):** Seien  $U \subset \mathbf{R}^n$  und  $V \subset \mathbf{R}^m$  offen und  $g : U \rightarrow \mathbf{R}^m$ ,  $f : V \rightarrow \mathbf{R}^k$  Abbildungen mit  $g(U) \subset V$ . Ist  $g$  in  $x \in U$  differenzierbar und  $f$  in  $y := g(x) \in V$  differenzierbar, dann ist die zusammengesetzte Abbildung

$$f \circ g : U \rightarrow \mathbf{R}^k$$

in  $x$  differenzierbar, und es gilt:

$$D(f \circ g)_x = (Df)_y \circ (Dg)_x \text{ bzw. } J_{f \circ g}(x) = J_f(g(x))J_g(x).$$

In Koordinaten schaut das so aus:

**Satz 5** Seien  $V \subset \mathbf{R}^m$ ,  $U \subset \mathbf{R}^n$  offen und  $f : V \rightarrow \mathbf{R}$ ,  $g : U \rightarrow \mathbf{R}^m$  differenzierbare Abbildungen mit  $g(U) \subset V$ . Dann ist

$$h := f \circ g : U \rightarrow \mathbf{R}$$

partiell differenzierbar und es gilt fr  $i = 1, 2, \dots, n$ :

$$\frac{\partial h}{\partial x_i}(x_1, \dots, x_n) = \sum_{j=1}^m \frac{\partial f}{\partial y_j}(g_1(x), \dots, g_m(x)) \frac{\partial g_j}{\partial x_i}(x_1, \dots, x_n).$$

BEWEIS.

$$J_{f \circ g}(x) = J_h(x) = \left( \frac{\partial h}{\partial x_1}(x), \dots, \frac{\partial h}{\partial x_n}(x) \right) = \text{grad } h(x).$$

$$J_f(g(x)) = \left( \frac{\partial f}{\partial y_1}(g(x)), \dots, \frac{\partial f}{\partial y_m}(g(y)) \right) = \text{grad } f(g(x)).$$

$$J_g(x) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(x) & \dots & \frac{\partial g_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_1}(x) & \dots & \frac{\partial g_m}{\partial x_n}(x) \end{bmatrix}$$

Die Behauptung ergibt sich aus  $J_{f \circ g}(x) = J_f(g(x))J_g(x)$ . ■

BEISPIEL. Sei  $f$  die Funktion  $(x, y) \mapsto (3 \sin xy, e^{x^2+y^2}, 2x-6)$ ,  $g$  die Funktion  $(u, v, w) \mapsto u^2 + v^2 + w^2$ .

$$J_f(x, y) = \begin{bmatrix} 3y \cos xy & 3x \cos xy \\ 2xe^{x^2+y^2} & 2ye^{x^2+y^2} \\ 2 & 0 \end{bmatrix}.$$

$$J_g(u, v, w) = \begin{bmatrix} 2u \\ 2v \\ 2w \end{bmatrix}.$$

$$J_{g \circ f}(x, y) = J_g(f(x, y))J_f(x, y)$$

$$= \begin{bmatrix} 2(3 \sin xy) \\ 2(e^{x^2+y^2}) \\ 2(2x-6) \end{bmatrix} \begin{bmatrix} 3y \cos xy & 3x \cos xy \\ 2xe^{x^2+y^2} & 2ye^{x^2+y^2} \\ 2 & 0 \end{bmatrix}.$$

BEISPIEL. Die wichtigsten Anwendungen der Kettenregel sind die Koordinatentransformationen: Z.B.

in  $\mathbf{R}^2$ :

$$\phi_1 : (r, \theta) \mapsto (r \cos \theta, r \sin \theta) \text{ (Polarkoordinaten)}$$

in  $\mathbf{R}^3$ :

$$\phi_2 : (r, \theta, \zeta) \mapsto (r \cos \theta, r \sin \theta, \zeta) \text{ (Zylinder-Koordinaten)}$$

$$\phi_3 : (r, \theta, \phi) \mapsto (r \cos \theta \sin \phi, r \sin \theta \sin \phi, r \cos \phi) \text{ (sphärische Koordinaten)}.$$

Es gilt etwa:

$$J_{\phi_3}(r, \theta, \phi) = \begin{bmatrix} \cos \theta \sin \phi & -r \sin \theta \sin \phi & r \cos \theta \cos \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \phi & 0 & -r \sin \phi \end{bmatrix}.$$

**Partielle Ableitungen höherer Ordnung:** Sei  $U \subset \mathbf{R}^n$  offen. Dann wird durch vollständige Induktion definiert: Eine Funktion  $f : U \rightarrow \mathbf{R}$  heißt  **$k$ -mal stetig partiell differenzierbar**, wenn  $f$   $(k-1)$ -mal stetig partiell differenzierbar ist und alle  $(k-1)$ -ten partiellen Ableitungen

$$D_{i_{k-1}} \dots D_{i_2} D_{i_1} f : U \rightarrow \mathbf{R} \quad (1 \leq i_\ell \leq n \text{ für } \ell = 1, \dots, k-1)$$

stetig partiell differenzierbar sind. Wir können dann die partiellen Ableitungen der Ordnung  $k$  definieren. Dass diese Definition sinnvoll ist, folgt aus dem Satz.

**Satz 6** Sei  $U \subset \mathbf{R}^n$  offen und  $f : U \rightarrow \mathbf{R}$  zweimal stetig partiell differenzierbar. Dann gilt für jedes  $x \in U$  und für alle  $i, j = 1, \dots, n$

$$D_j D_i f(x) = D_i D_j f(x).$$

**Korollar 7** Sei  $U \subset \mathbf{R}^n$  offen und  $f : U \rightarrow \mathbf{R}$   $k$ -mal stetig partiell differenzierbar. Dann ist

$$D_{i_k} \dots D_{i_2} D_{i_1} f = D_{i_{\pi(k)}} \dots D_{i_{\pi(2)}} D_{i_{\pi(1)}} f$$

für jede Permutation  $\pi$  der Zahlen  $1, \dots, k$  und  $1 \leq i_\ell \leq n$  für  $\ell = 1, \dots, k$ .

Der Beweis erfolgt durch vollständige Induktion über  $k$  unter Verwendung der Tatsache, daß sich jede Permutation aus Vertauschungen benachbarter Glieder zusammensetzen läßt.

Damit können wir  $D^p f$  eindeutig definieren, falls  $p$  ein Multi-Index ist und  $f$  hinreichend glatt.

**Satz 8** Sei  $U \subset \mathbf{R}^n$  offen,  $f : U \rightarrow \mathbf{R}$   $k$ -mal stetig partiell differenzierbar,  $x \in U, h \in \mathbf{R}^n$ . Dann ist die Funktion

$$t \mapsto g(t) := f(x + th)$$

in einem Intervall  $]-\epsilon, \epsilon[ \subset \mathbf{R}$  ( $\epsilon > 0$ ) definiert und  $k$ -mal stetig differenzierbar, und es gilt

$$\frac{d^k g}{dt^k}(t) = \sum_{|\alpha|=k} \frac{k!}{\alpha!} D^\alpha f(x + th) h^\alpha.$$

**Satz 9 (Taylorsche Formel):** Sei  $U \subset \mathbf{R}^n$  offen,  $x \in U$  und  $h \in \mathbf{R}^n$ , so daß  $x + th \in U$  für alle  $t \in [0, 1]$ . Sei  $f : U \rightarrow \mathbf{R}$  eine  $(k + 1)$ -mal stetig differenzierbare Funktion. Dann existiert ein  $\theta \in [0, 1]$ , so daß

$$f(x + h) = \sum_{|\alpha| \leq k} \frac{D^\alpha f(x)}{\alpha!} h^\alpha + \sum_{|\alpha|=k+1} \frac{D^\alpha f(x + \theta h)}{\alpha!} h^\alpha.$$

BEWEIS. Die Funktion  $g : [0, 1] \rightarrow \mathbf{R}$ , wobei  $t \mapsto f(x + th)$  ist  $(k + 1)$ -mal stetig differenzierbar. Nach der Taylor-Formel für Funktionen einer reellen Veränderlichen, existiert ein  $\theta \in [0, 1]$ , so daß

$$g(1) = \sum_{i=0}^k \frac{g^{(i)}(0)}{i!} 1^i + \frac{g^{(k+1)}(\theta)}{(k+1)!} 1^{k+1}.$$

Es gilt für  $i = 0, 1, \dots, k$

$$\frac{g^{(i)}(0)}{i!} = \sum_{|\alpha|=i} \frac{D^\alpha f(x)}{\alpha!} h^\alpha$$

und

$$\frac{g^{(k+1)}(\theta)}{(k+1)!} = \sum_{|\alpha|=k+1} \frac{D^\alpha f(x + \theta h)}{\alpha!} h^\alpha,$$

woraus die Behauptung folgt. ■

**Korollar 10** Sei  $U \subset \mathbf{R}^n$  offen,  $x \in U, h \in \mathbf{R}^n$  mit  $x + th \in U$  für alle  $t \in [0, 1]$ . Sei  $f : U \rightarrow \mathbf{R}$   $k$ -mal stetig differenzierbar. Dann ist

$$f(x + h) = \sum_{|\alpha| \leq k} \frac{D^\alpha f(x)}{\alpha!} h^\alpha + \varphi(h) |h|^k,$$

wobei  $\lim_{h \rightarrow 0} \varphi(h) = 0$ .

BEWEIS. Es existiert ein  $\theta \in [0, 1]$ , so daß

$$f(x + h) = \sum_{|\alpha| \leq k} \frac{D^\alpha f(x)}{\alpha!} h^\alpha + \sum_{|\alpha|=k} r_\alpha(h) h^\alpha$$

mit  $r_\alpha(h) = \frac{D^\alpha f(x + \theta h) - D^\alpha f(x)}{\alpha!}$ . Wegen der Stetigkeit von  $D^\alpha f$  ist  $\lim_{h \rightarrow 0} r_\alpha(h) = 0$ .

Mit

$$\varphi(h) := \sum_{|\alpha|=k} r_\alpha(h) \frac{h^\alpha}{|h|^k}$$

folgt die Behauptung, da

$$\frac{|h^\alpha|}{|h|^k} = \frac{|h_1^{\alpha_1} \dots h_n^{\alpha_n}|}{|h|^{\alpha_1} \dots |h|^{\alpha_n}} \leq 1 \text{ für } |\alpha| = k.$$

■

BEISPIEL. Für  $f = f(x, y)$  hat die Taylorreihe die Gestalt

$$\sum_{m,n} \frac{1}{m!n!} \frac{\partial^{m+n} f}{\partial^m x \partial^n y} \Big|_{x=0} x^m y^n$$

z.B.

$$\sin(xy) = \sum_{m=0}^{\infty} (-1)^{m+1} \frac{x^{2m+1} y^{2m+1}}{(2m+1)!}.$$

DEFINITION. Sei  $U \subset \mathbf{R}^n$  offen und  $f : U \rightarrow \mathbf{R}$  eine Funktion. Ein Punkt  $x \in U$  heißt **lokales Maximum** (bzw. **lokales Minimum**) von  $f$ , falls eine Umgebung  $V \subset U$  von  $x$  existiert, so daß  $f(x) \geq f(y)$  (bzw.  $f(x) \leq f(y)$ ) für alle  $y \in V$ .

Tritt in dieser Definition der Fall  $f(x) = f(y)$  nur für  $x = y$  ein, so spricht man von einem **isolierten** oder **strengen** lokalen Maximum bzw. Minimum. Ein **lokales Extremum** ist ein lokales Maximum oder Minimum.

**Satz 11** Sei  $U \subset \mathbf{R}^n$  offen und  $f : U \rightarrow \mathbf{R}$  eine partiell differenzierbare Funktion. Besitzt  $f$  in  $x \in U$  ein lokales Extremum, so gilt:

$$\text{grad } f(x) = 0.$$

BEWEIS. Für  $k = 1, \dots, n$  sei

$$g_k(t) := f(x + te_k)$$

$g_k$  ist definiert auf einem Intervall  $[-\epsilon, \epsilon] \subset \mathbf{R}$  mit  $\epsilon > 0$  und dort differenzierbar.

Hat  $f$  in  $x$  ein lokales Extremum, so hat  $g_k$  in 0 ein lokales Extremum und es ist  $g'_k(0) = 0$ .

Wegen

$$g'_k(0) = (\text{grad } f(x)|_{e_k}) = \frac{\partial f}{\partial x_k}(x)$$

ist

$$\text{grad } f(x) = \left( \frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right) = 0, \text{ q.e.d.}$$

■

DEFINITION. Sei  $U \subset \mathbf{R}^n$  offen und  $f : U \rightarrow \mathbf{R}$  eine zweimal stetig partiell differenzierbare Funktion. Unter der **Hesseschen Matrix** von  $f$  im Punkt  $x \in U$  versteht man die  $n \times n$ -Matrix

$$\text{Hess}(f)(x) := \left[ \frac{\partial^2 f}{\partial x_i \partial x_k} \right]$$

$\text{Hess}(f)(x)$  ist symmetrisch, da

$$\frac{\partial^2 f}{\partial x_i \partial x_k}(x) = \frac{\partial^2 f}{\partial x_k \partial x_i}(x).$$

DEFINITION. Eine symmetrische Matrix  $A \in M_n$  heißt **positiv (negativ) definit**, wenn alle Eigenwerte von  $A$  positiv (negativ) sind.

Die Matrix heißt **indefinit**, wenn sie mindestens einen positiven und einen negativen Eigenwert besitzt. (Bekanntlich sind alle Eigenwerte einer reellen symmetrischen Matrix reell.)

Wir verwenden folgende Tatsache (vgl. Vorlesung "Lineare Algebra"): Sei  $A \in M_n$  eine symmetrische, positiv definite Matrix mit den Eigenwerten  $\lambda_1, \dots, \lambda_n$  und  $\lambda := \min(\lambda_1, \dots, \lambda_n)$ . Dann gilt für alle  $x \in \mathbf{R}^n$

$$(Ax|x) \geq \lambda|x|^2.$$

**Satz 12** Sei  $U \subset \mathbf{R}^n$  offen,  $f \in U \rightarrow \mathbf{R}$  zweimal stetig partiell differenzierbar. Sei  $x \in U$  und  $\text{grad } f(x) = 0$ .

- i) Ist  $\text{Hess}(f)(x)$  positiv definit, so hat  $f$  in  $x$  ein isoliertes lokales Minimum.
- ii) Ist  $\text{Hess}(f)(x)$  negativ definit, so hat  $f$  in  $x$  ein isoliertes lokales Maximum.
- iii) Ist  $\text{Hess}(f)(x)$  indefinit, so hat  $f$  in  $x$  weder ein lokales Maximum noch ein lokales Minimum.

BEWEIS. Die Taylorentwicklung von  $f$  um  $x$  liefert

$$f(x+h) = f(x) + \sum_{|\alpha|=2} \frac{D^\alpha f(x)}{\alpha!} h^\alpha + \varphi(h)|h|^2,$$

wobei  $\lim_{h \rightarrow 0} \varphi(h) = 0$ .

Es gilt:

$$Q(h) := \sum_{|\alpha|=2} \frac{D^\alpha f(x) h^\alpha}{\alpha!} = \frac{1}{2} \sum_{i,j=1}^n D_i D_j f(x) h_i h_j = \frac{1}{2} \sum_{i,j=1}^n \alpha_{ij} h_i h_j = \frac{1}{2} (Ah|h),$$

mit

$$h = (h_1, \dots, h_n) \text{ und } (\alpha_{ij}) = A = \text{Hess}(f)(x).$$

Also ist

$$f(x+h) - f(x) = Q(h) + \varphi(h)|h|^2.$$



i) Ist  $A$  positiv definit und  $\lambda$  der kleinste Eigenwert von  $A$ , dann ist

$$Q(h) \geq \frac{1}{2}\lambda|h|^2.$$

Es gibt ein  $\epsilon > 0$ , so daß  $|\varphi(h)| \leq \frac{\lambda}{4}$  für  $|h| < \epsilon$ .

$$f(x+h) - f(x) = Q(h) + \varphi(h)|h|^2 \geq \frac{1}{4}\lambda|h|^2 > 0$$

für alle  $h$  mit  $0 < |h| < \epsilon$ , d.h.  $x$  ist isoliertes lokales Minimum von  $f$ .

ii) Ist die Hessesche Matrix negativ definit, so geht man zur Funktion  $-f$  über und hat dadurch Fall ii) auf den Fall i) zurückgeführt.

iii) Sei  $A$  indefinit. Wir müssen zeigen, daß in jeder Umgebung von  $x$  Punkte  $x'$  existieren mit  $f(x') < f(x)$ , als auch Punkte  $x''$  mit  $f(x'') > f(x)$ .

Sei  $v$  ein Eigenvektor von  $A$  zu einem Eigenwert  $\lambda < 0$ . Wir können annehmen, daß  $|v| = 1$ . Für  $h := tv, t > 0$ , gilt dann

$$(Ah|h) = t^2(Av|v) = t^2\lambda = \lambda|h|^2,$$

also

$$f(x+h) - f(x) = \frac{1}{2}\lambda|h|^2 + \varphi(h)|h|^2 \geq \frac{1}{4}\lambda|h|^2 < 0,$$

falls nur  $t$  genügend klein ist.

Ist  $w$  ein Eigenvektor von  $A$  zu einem Eigenwert  $\mu > 0$  mit  $|w| = 1$ , so gilt für  $h := tw, t > 0$ , analog wie oben

$$f(x+h) - f(x) = \frac{1}{2}\mu|h|^2 + \varphi(h)|h|^2 \leq \frac{1}{4}\mu|h|^2 < 0$$

für genügend kleines  $t$ . ■

**BEISPIELE.** Wir betrachten einige typische Beispiele von Funktionen  $f : \mathbf{R}^2 \rightarrow \mathbf{R}$

a)  $f(x, y) = a + x^2 + y^2$ . Die Funktion  $f$  hat im Nullpunkt  $(0, 0) \in \mathbf{R}^2$  ein isoliertes lokales Minimum, da

$$\frac{1}{2} \text{Hess} (f) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Der Graph von  $f$ ,

$$\Gamma_f = \{(x, y, z) \in \mathbf{R}^3 : z = a + x^2 + y^2\}$$

ist ein nach oben geöffnetes Paraboloid (wenn man sich die  $z$ -Achse als nach oben gerichtet denkt).

b)  $f(x, y) = a - x^2 - y^2$ . Hier liegt im Nullpunkt ein isoliertes lokales Maximum vor, es gilt

$$\frac{1}{2} \text{Hess} (f) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Der Graph von  $f$ ,

$$\Gamma_f = \{(x, y, z) \in \mathbf{R}^3 : z = a - x^2 - y^2\}$$

ist ein nach unten geöffnetes Paraboloid.

c)  $f(x, y) = a + x^2 - y^2$ . Der Gradient von  $f$  im Nullpunkt verschwindet, und es gilt

$$\frac{1}{2} \text{Hess} (f) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Die Hessesche Matrix ist also indefinit, es liegt weder ein lokales Maximum noch Minimum vor.

Der Graph von  $f$ ,

$$\Gamma_f = \{(x, y, z) \in \mathbf{R}^3 : z = a + x^2 - y^2\},$$

ist eine sog. **Sattelfläche**. Längs der Achse  $x = 0$  fallen die Werte von  $f$  vom Nullpunkt aus ab, längs der Achse  $y = 0$  nehmen die Werte von  $f$  vom Nullpunkt aus zu.

d) Ist die Hessesche Matrix in einer Nullstelle des Gradienten semi-definit, so lassen sich keine allgemeinen Aussagen machen, wie folgende Beispiele zeigen:

$$\begin{aligned} f_1(x, y) &= x^2 + y^4 \\ f_2(x, y) &= x^2 \\ f_3(x, y) &= x^2 + y^3. \end{aligned}$$

Für alle drei Funktionen verschwindet der Gradient im Nullpunkt und es gilt

$$\text{Hess} (f_k)(0) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \quad (k = 1, 2, 2),$$

die Hessesche Matrix ist also positiv semi-definit.

Die Funktion  $f_1$  hat im Nullpunkt ein isoliertes lokales Minimum. Die Funktion  $f_2$  hat im Nullpunkt ein lokales Minimum, das aber nicht isoliert ist (alle Punkte auf der Achse  $x = 0$  sind ebenfalls lokale Minima).

Die Funktion  $f_3$  hat im Nullpunkt weder ein lokales Minimum noch ein lokales Maximum.

BEISPIELE. Für  $f(x, y) = \sin(x^2 + y^2)$  gilt

$$\begin{aligned} D_1 f &= \frac{\partial f}{\partial x} = 2x \cos(x^2 + y^2), \quad D_2 f = \frac{\partial f}{\partial y} = 2y \cos(x^2 + y^2) \\ D_{12} f &= \frac{\partial^2 f}{\partial x \partial y} = -4xy \sin(x^2 + y^2). \end{aligned}$$

BEISPIEL. Für  $f = f(x, y)$  hat die Taylorreihe die Gestalt

$$\sum_{m,n} \frac{1}{m!n!} \left. \frac{\partial^{m+n} f}{\partial^m x \partial^n y} \right|_{x=0} x^m y^n$$

z.B.

$$\sin(xy) = \sum_{m=0}^{\infty} (-1)^{m+1} \frac{x^{2m+1} y^{2m+1}}{(2m+1)!}.$$

BEISPIELE. a) Für  $f(x, y) = \sin xy + e^{x^2-y^2}$  gilt:

$$\begin{aligned} D_1 f(x, y) &= y \cos xy + 2xe^{x^2-y^2} \\ D_2 f(x, y) &= x \cos xy - 2ye^{x^2-y^2} \\ D_{11} f(x, y) &= -y^2 \sin xy + 4x^3 e^{x^2-y^2} + 2e^{x^2-y^2} \\ D_{22} f(x, y) &= -x^2 \sin xy - 4y^2 e^{x^2-y^2} - 2e^{x^2-y^2} \\ D_{12} f(x, y) &= \cos xy - xy \sin xy - 5xye^{x^2-y^2} \\ D_{21} f(x, y) &= \cos xy - xy \sin xy - 5xye^{x^2-y^2} \end{aligned}$$

b) Berechne die Taylorreihe von  $e^{x+3y}$  an der Stelle  $(1, 1)$  bis zum 3. Grad.

$$\begin{aligned} f(1+h, 1+k) &= e^3 + e^3 h + 2e^3 h + \frac{e^3}{2!} h^2 + 2e^3 h k + \frac{4e^3}{2k^2} k^2 \\ &+ \frac{e^3}{3!} h^3 + \frac{2e^3}{2!} h^2 k + \frac{4e^3}{2!} h k^2 + \frac{8e^3}{3!} k^3 + R_4. \end{aligned}$$

BEISPIEL. Sei  $f$  die Funktion  $f(x, y) = (3 \sin xy, e^{x^2+y^2}, 2x-6)$ ,  $g$  die Funktion  $g(u, v, w) = u^2 + v^2 + w^2$ .

$$J_f(x, y) = \begin{bmatrix} 3y \cos xy & 3x \cos xy \\ 2xe^{x^2+y^2} & 2ye^{x^2+y^2} \\ 2 & 0 \end{bmatrix}$$

$$J_g(u, v, w) = \begin{bmatrix} 2u \\ 2v \\ 2w \end{bmatrix}$$

$$\begin{aligned} J_{g \circ f}(x, y) &= J_g(f(x, y)) J_f(x, y) \\ &= \begin{bmatrix} 2(3 \sin xy) \\ 2(e^{x^2+y^2}) \\ 2(2x-6) \end{bmatrix} \begin{bmatrix} 3y \cos xy & 3x \cos xy \\ 2xe^{x^2+y^2} & 2ye^{x^2+y^2} \\ 2 & 0 \end{bmatrix}. \end{aligned}$$

BEISPIEL. Die wichtigsten Anwendungen der Kettenregel sind die Koordinatentransformationen: Z.B. in  $\mathbf{R}^2$ :  $\phi_1 : (r, \theta) \mapsto (r \cos \theta, r \sin \theta)$  (Polarkoordinaten) in  $\mathbf{R}^3$ :  $\phi_2 : (r, \theta, \zeta) \mapsto (r \cos \theta, r \sin \theta, \zeta)$  (Zylinder Koordinaten)

$$\phi_3 : (r, \theta, \phi) \mapsto (r \cos \theta \sin \phi, r \sin \theta \sin \phi, r \cos \phi)$$

(sphärische Koordinaten).

Es gilt:

$$J_{\phi_3}(r, \theta, \phi) = \begin{bmatrix} \cos \theta \sin \phi & -r \sin \theta \sin \phi & r \cos \theta \cos \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \tau \phi & 0 & -r \sin \phi \end{bmatrix}$$

$$\det J_{\phi_3}(r, \theta, \phi) = -r^2 \sin \phi$$

$$J_{\phi_3}^{-1} = \begin{bmatrix} -\frac{\sin \theta}{r \sin \phi} & \frac{\cos \theta}{r \sin \phi} & 0 \\ \frac{\cos \theta \cos \phi}{r} & \frac{\sin \theta \cos \phi}{r} & -\frac{\sin \phi}{r} \end{bmatrix}.$$

Sei jetzt  $f$  eine Funktion auf  $\mathbf{R}^3$ ,  $F = f \circ \phi_3$ . ( $f$  ist ein Skalarfeld “in kartesischen Koordinaten”,  $F$  das Feld “in sphärischen Koordinaten”). Z.B. Für

$$f(x, y, z) = \frac{1}{(x^2 + y^2 + z^2)^{\frac{\alpha}{2}}}$$

ist

$$F(r, \theta, \phi) = \frac{1}{r^\alpha}.$$

Es gilt:

$$\begin{aligned} J_{F(r, \theta, \phi)} &= \begin{bmatrix} D_1 F \\ D_2 F \\ D_3 F \end{bmatrix}_{(r, \theta, \phi)} = J_f(x, y, z) J_{\phi_3}(r, \theta, \phi) \\ &= \begin{bmatrix} D_1 f \\ D_2 f \\ D_3 f \end{bmatrix}_{(x, y, z)} = \begin{bmatrix} \cos \theta \sin \phi & -r \sin \theta \sin \phi & r \cos \theta \cos \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \phi & 0 & -r \sin \phi \end{bmatrix} \end{aligned}$$

d.h.

$$\begin{aligned} D_1 F &= \cos \theta \sin \phi (D_1 f) \circ \phi_3 + \sin \theta \sin \phi (D_2 f) \circ \phi_3 + \cos \phi (D_3 f) \circ \phi_3 \\ &\quad \left( \text{“} \frac{\partial f}{\partial r} = \cos \theta \sin \phi \frac{\partial f}{\partial x} + \sin \theta \sin \phi \frac{\partial f}{\partial y} + \cos \phi \frac{\partial f}{\partial z} \text{”} \right). \end{aligned}$$

$$D_2 F = -r \sin \theta \sin \phi (D_1 f) \circ \phi_3 + r \cos \theta \sin \phi (D_2 f) \circ \phi_3$$

$$D_3 F = r \cos \phi (D_1 f) \circ \phi_3 + r \sin \theta \cos \phi (D_2 f) \circ \phi_3 - \sin \phi (D_3 f) \circ \phi_3.$$

Um höhere Ableitungen auszurechnen, kann man diese Formeln noch einmal verwenden, z.B.

$$\begin{aligned} D_{11} F &= \cos \theta \sin \phi (\cos \theta \sin \phi (D_{11} f) \circ \phi_3 + \cos \theta \sin \phi (D_{12} f) \circ \phi_3 + \cos \phi (D_{13} f) \circ \phi_3) \\ &+ \sin \theta \sin \phi (\cos \theta \sin \phi (D_{11} f) \circ \phi_3 + \sin \theta \sin \phi (D_{22} f) \circ \phi_3) \\ &+ \cos \phi (\cos \theta \sin \phi (D_{13} f) \circ \phi_3 + \sin \theta \sin \phi (D_{23} f) \circ \phi_3 + \cos \phi (D_{33} f) \circ \phi_3). \end{aligned}$$

$$\begin{aligned} \left( \frac{\partial^2 f}{\partial r^2} &= \cos^2 \theta \sin^2 \phi \frac{\partial^2 f}{\partial x^2} + \sin^2 \theta \sin^2 \phi \frac{\partial^2 f}{\partial y^2} \right. \\ &+ \cos^2 \phi \frac{\partial^2 f}{\partial y^2} + 2 \cos \theta \sin \theta \sin^2 \phi \frac{\partial^2 f}{\partial x \partial y} \\ &+ 2 \cos \theta \sin \theta \sin^2 \phi \frac{\partial^2 f}{\partial x \partial y} \\ &+ 2 \cos \theta \sin \phi \cos \phi \frac{\partial^2 f}{\partial x \partial y} \\ &\left. + 2 \sin \theta \cos \phi \sin \phi \frac{\partial^2 f}{\partial y \partial z} \right) \end{aligned}$$