Mathematik II für MB, WI/MB und andere

Prof. Dr. Wilhelm Stannat

Inhalt:

1. Folgen und Reihen von Funktionen
2. Kurven im $\mathbb{R}^n$
3. Funktionen in mehreren Variablen
4. Differentialrechnung von Funktionen mehrerer Veränderlicher
5. Implizite Funktionen und Extrema mit Nebenbedingungen

Das vorliegende Skript ist eine Zusammenfassung der Kapitel 4 und 5 der Vorlesung Mathematik II für MB, WI/MB und andere, die im SS 2007 an der TU Darmstadt gehalten wurde.

Korrekturen bitte per Email an `stannat@mathematik.tu-darmstadt.de`

4 Differentialrechnung von Funktionen mehrerer Veränderlicher

4.1 Partielle Ableitungen

Es sei $D \subset \mathbb{R}^n$, $f : D \to \mathbb{R}$ eine Funktion und $X = [x_1, \ldots, x_n]^T \in D$ ein innerer Punkt. Für kleine h ist

$$h \mapsto f(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n)$$

wohldefiniert als Funktion von h.
Existiert dann der Grenzwert der Differenzenquotienten in der i-ten Koordinatenrichtung

$$\lim_{h\to 0} \frac{1}{h}(f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)) =: \frac{\partial f}{\partial x_i}(X)$$

so heißt f an der Stelle X **partiall differenzierbar nach** x_i, $1 \leq i \leq n$. Der Grenzwert $\frac{\partial f}{\partial x_i}(X)$ heißt **partielle Ableitung** von f nach x_i an der Stelle X.
f heißt an der Stelle X **partiell differenzierbar**, falls alle partiellen Ableitungen $\frac{\partial f}{\partial x_i}(X)$, $1 \leq i \leq n$, existieren. Der Vektor

$$\text{grad } f(X) := \left[\frac{\partial f}{\partial x_1}(X), \ldots, \frac{\partial f}{\partial x_n}(X)\right]^T$$

der partiellen Ableitungen von f in X heißt **Gradient von** f **in** X.

Bemerkung Bezeichnen wir mit $e_i = [0, \ldots, 0, 1, 0, \ldots, 0]^T$ den i-ten Einheitsvektor im $\mathbb{R}^n$, so können wir schreiben

$$\frac{\partial f}{\partial x_i}(X) = \lim_{h\to 0} \frac{1}{h}(f(X + he_i) - f(X))$$

Beispiele 4.1

(i) $f(x, y) = 2x^2 + y^2$ auf $\mathbb{R}^2$.

$$\begin{aligned}\frac{\partial f}{\partial x}(x, y) &= \lim_{h\to 0} \frac{1}{h}(f(x + h, y) - f(x, y)) = \lim_{h\to 0} \frac{1}{h}(2(x + h)^2 + y^2 - (2x^2 + y^2)) \\ &= \lim_{h\to 0} \frac{1}{h}(4xh + 2h^2) = 4x\end{aligned}$$

Entsprechend gilt $\frac{\partial f}{\partial y}(x, y) = 2y$

(ii) $f(x, y) = \sin(xy^2)$ auf $\mathbb{R}^2$

$$\frac{\partial f}{\partial x}(x, y) = y^2 \cos(xy^2) \qquad\qquad \frac{\partial f}{\partial y}(x, y) = 2yx \cos(xy^2)$$

Bei partieller Ableitung nach x behandelt man also y wie eine Konstante und bei partieller Ableitung nach y behandelt man x wie eine Konstante.

Alternative Bezeichnungen für partielle Ableitungen: $f_{x_i}, f_x, f_y, f_z, f_t, \ldots$

Partielle Ableitungen höherer Ordnung

Ist f partiell differenzierbar nach x_i, so definiert

$$X \mapsto \frac{\partial f}{\partial x_i}(X)$$

wieder eine Funktion in n Veränderlichen, die man wieder gegebenenfalls nach x_j partiell differenzieren kann:

$$\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(X) =: \frac{\partial^2 f}{\partial x_j \partial x_i}(X) \quad \text{für} \quad 1 \leq i, j \leq n$$

Man erhält auf diese Weise die **partiellen Ableitungen zweiter Ordnung** von f.

Beispiele 4.2 (i) $f(x, y) = 2x^2 + y^2$, also

$$\frac{\partial f}{\partial x}(x, y) = 4x \qquad \frac{\partial f}{\partial y}(x, y) = 2y$$

und damit

$$\frac{\partial^2 f}{\partial x \partial y}(x, y) = 0 \qquad \frac{\partial^2 f}{\partial y \partial x}(x, y) = 0 \qquad \frac{\partial^2 f}{\partial x^2}(x, y) = 4 \qquad \frac{\partial^2 f}{\partial y^2}(x, y) = 2$$

(ii) $f(x, y) = \sin(xy^2)$, also

$$\frac{\partial^2 f}{\partial x^2}(x, y) = -y^4 \sin(xy^2)$$

Analog zu den partiellen Ableitungen zweiter Ordnung bildet man partielle Ableitungen höherer Ordnung:

$$\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(x), \frac{\partial^3 f}{\partial x_i^2 \partial x_j}(x), \ldots$$

Bemerkungen

(i) Bei Funktionen f in einer reellen Veränderlichen folgt aus der Differenzierbarkeit von f die Stetigkeit von f. Das ist für Funktionen in mehr als einer Veränderlichen im allgemeinen nicht richtig.

Beispiel

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{für} \quad [x, y]^T \neq [0, 0]^T \\ 0 & \text{für} \quad [x, y]^T = [0, 0]^T \end{cases}$$

ist nicht stetig in 0 (siehe Beispiel 3.1(iii)). Jedoch ist

$$\frac{\partial f}{\partial x}(0, 0) = \frac{\partial f}{\partial y}(0, 0) = 0$$

(ii) Im allgemeinen ist die Reihenfolge der partiellen Ableitungen nicht vertauschbar, d.h. im allgemeinen ist z.B.
$$\frac{\partial^2 f}{\partial x \partial y} \neq \frac{\partial^2 f}{\partial y \partial x}$$

Beispiel
$$f(x,y) = \begin{cases} xy\frac{x^2-y^2}{x^2+y^2} & \text{für } [x,y]^T \neq [0,0]^T \\ 0 & \text{sonst} \end{cases}$$

Für $[x,y]^T \neq 0$ gilt:
$$\begin{aligned} f_x(x,y) &= y\frac{x^2-y^2}{x^2+y^2} + xy\left(\frac{2x}{x^2+y^2} - \frac{2x(x^2-y^2)}{(x^2+y^2)^2}\right) \\ &= y\left(\frac{x^2-y^2}{x^2+y^2} + \frac{4x^2y^2}{(x^2+y^2)^2}\right). \end{aligned}$$

Analog gilt
$$f_y(x,y) = x\left(\frac{x^2-y^2}{x^2+y^2} - \frac{4x^2y^2}{(x^2+y^2)^2}\right)$$

und $f_x(0,0) = f_y(0,0) = 0$. Folglich
$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = \lim_{h\to 0}\frac{1}{h}\left(\frac{\partial f}{\partial x}(0,h) - \frac{\partial f}{\partial x}(0,0)\right) = \lim_{h\to 0}\frac{1}{h}\left(h\cdot(-1)\right) = -1$$

aber
$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{h\to 0}\frac{1}{h}\left(\frac{\partial f}{\partial y}(h,0) - \frac{\partial f}{\partial y}(0,0)\right) = \lim_{h\to 0}\frac{1}{h}(h\cdot 1) = 1$$

Es gilt jedoch allgemein: Sind die partiellen Ableitungen k-ter Ordnung **stetig**, so ist die Reihenfolge der partiellen Ableitungen bis zur k-ten Ordnung vertauschbar.

(iii) Für die partiellen Ableitungen von Summen und Produkten gelten dieselben **Rechenregeln** wie im Falle $n = 1$: Sind $f, g : D \mapsto \mathbb{R}$ an der Stelle X_0 partiell nach x_i differenzierbar, so sind auch die folgenden Funktionen an der Stelle X_0 partiell differenzierbar nach x_i:

(a) **Linearität** $\alpha f + \beta g$ für α, $\beta \in \mathbb{R}$ mit
$$\frac{\partial(\alpha f + \beta g)}{\partial x_i}(X_0) = \alpha\frac{\partial f}{\partial x_i}(X_0) + \beta\frac{\partial g}{\partial x_i}(X_0)$$

(b) **Produktregel** $f \cdot g$
$$\frac{\partial(f\cdot g)}{\partial x_i}(X_0) = \frac{\partial f}{\partial x_i}(X_0)\cdot g(X_0) + f(X_0)\cdot\frac{\partial g}{\partial x_i}(X_0)$$

(c) **Quotientenregel** Ist $g(X) \neq 0$ für $X \in D$, so ist auch $\frac{f}{g}$ an der Stelle X_0 partiell differenzierbar nach x_i mit
$$\frac{\partial}{\partial x_i}\left(\frac{f}{g}\right)(X_0) = \frac{\frac{\partial f}{\partial x_i}(X_0)\cdot g(X_0) - f(X_0)\cdot\frac{\partial g}{\partial x_i}(X_0)}{g^2(X_0)}$$

Weiterhin gilt:

Satz (Kettenregel): Es sei $D \subset \mathbb{R}^n$ offen, $f : D \to \mathbb{R}, Y \mapsto f(Y) = f(y_1, \ldots, y_n)$ stetig partiell differenzierbar. Weiterhin sei $U \subset \mathbb{R}^m$ $(m \geq 1)$ offen,

$$g_i : U \to \mathbb{R} \quad \text{stetig partiell differenzierbar für} \quad 1 \leq i \leq n$$

und es gelte

$$[g_1(X), \ldots, g_n(X)]^T \in D \quad \text{für alle} \quad X \in U\,.$$

Dann ist die Verkettung

$$\begin{aligned} F\ &: U \to \mathbb{R} \\ & X \mapsto f\left(g_1(X), \ldots, g_n(X)\right) \end{aligned}$$

stetig partiell differenzierbar und für die partiellen Ableitungen gilt:

$$\frac{\partial F}{\partial x_i}(X) = \sum_{j=1}^{n} \underbrace{\frac{\partial f}{\partial y_j}(g_1(X), \ldots, g_n(X))}_{\text{äußere Abl.}} \cdot \underbrace{\frac{\partial g_j}{\partial x_i}(X)}_{\text{innere Abl.}}$$

Beispiele

(i) **Polarkoordinaten im $\mathbb{R}^2$**

Für einen Punkt $P = \begin{bmatrix} x \\ y \end{bmatrix}$ bezeichne

$$\begin{aligned} r(x,y) &= \left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\| = \sqrt{x^2 + y^2} \qquad \text{(Länge des zugehörigen Ortsvektors)} \\ \varphi(x,y) &= \arctan\left(\frac{x}{y}\right) \qquad \text{(Winkel des Ortsvektors zu } P \text{ mit der x-Achse)} \end{aligned}$$

die Polarkoordinaten von P.

Für einen Punkt $P = \begin{bmatrix} x \\ y \end{bmatrix}$ mit Polarkoordinaten (r, φ) gilt also

$$\begin{aligned} x &= r \cdot \cos\varphi = g_1(r,\varphi) \\ y &= r \cdot \sin\varphi = g_2(r,\varphi) \end{aligned}$$

für $(r, \varphi) \in]0, \infty[\times [0, 2\pi[$.

Ist eine Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben, so beschreibt die Vekettung

$$F(r, \varphi) = f(r\cos\varphi, r\sin\varphi)$$

also dieselbe Funktion in Polarkoordinaten.

Beispiel $f(x,y) = e^{-(x^2+y^2)}$

$$F(r,\varphi) = e^{-((r\cos\varphi)^2 + (r\sin\varphi)^2)} = e^{-r^2} \quad (!)$$

Mit Hilfe der Kettenregel lassen sich die partiellen Ableitungen von F aus den partiellen Ableitungen von f berechnen. Aus $F(r,\varphi) = f(g_1(r,\varphi), g_2(r,\varphi))$ folgt

$$\begin{aligned} \frac{\partial F}{\partial r}(r,\varphi) &= \frac{\partial f}{\partial x}(r\cdot\cos\varphi, r\sin\varphi)\cos\varphi + \frac{\partial f}{\partial y}(r\cos\varphi, r\sin\varphi)\cdot\sin\varphi \\ \frac{\partial F}{\partial \varphi}(r,\varphi) &= \frac{\partial f}{\partial x}(r\cos\varphi, r\sin\varphi)(-r\sin\varphi) + \frac{\partial f}{\partial y}(r\cos\varphi, r\sin\varphi)(r\cos\varphi) \end{aligned} \tag{4.1}$$

In obigem Beispiel etwa:

$$\begin{aligned} \frac{\partial F}{\partial r} &= -2r\cos\varphi\cdot e^{-r^2}\cdot\cos\varphi - 2r\sin\varphi\cdot e^{-r^2}\cdot\sin\varphi \\ &= -2re^{-r^2} \qquad \text{(hätte man auch } \textbf{direkt} \text{ sehen können!)} \end{aligned}$$

(ii) **Ableitung entlang von Kurven** Ist $\gamma : [a,b] \to D$ stetig differenzierbare Kurve und $f : D \to \mathbb{R}$ stetig partiell differenzierbar, so ist die Funktion

$$f\circ\gamma \,:\, [a,b] \mapsto f(\gamma(t)) = f(\gamma_1(t),\ldots,\gamma_n(t))$$

stetig differenzierbar nach t und es gilt

$$(f\circ\gamma)'(t) = \frac{d}{dt}\left(f\circ\gamma\right)(t) = \sum_{i=1}^{n} \underbrace{\underbrace{\frac{\partial f}{\partial y_i}(\gamma(t))}_{\text{äußere}}\cdot\underbrace{\gamma_i'(t)}_{\text{innere}}}_{\text{Ableitung}}$$

Beispiel $f(x,y) = e^{-(x^2+y^2)}$ und $\gamma(t) = \begin{bmatrix} r\cos t \\ r\sin t \end{bmatrix}$

$$\begin{aligned} \Rightarrow \frac{d}{dt}(f\circ\gamma)(t) &= -2\gamma_1(t)e^{-(\gamma_1(t)^2+\gamma_2(t)^2)}\cdot\gamma_1'(t) - 2\gamma_2(t)e^{-(\gamma_1(t)^2+\gamma_2(t)^2)}\cdot\gamma_2'(t) \\ &= \left(2r^2\cos(t)\sin(t) - 2r^2\sin(t)\cos(t)\right)e^{-r^2} = 0 \end{aligned}$$

Setzt man speziell für $X_0 \in D$

$$\gamma(t) = X_0 + t\cdot\vec{v}$$

für einen Vektor $\vec{v}\in\mathbb{R}^n$, so ist $\gamma(t)\in D$ für t in einer Umgebung der 0, also etwa für $t \in]-\varepsilon,\varepsilon[$.

Für die Ableitung von f entlang γ erhalten wir in $t=0$ wegen $\gamma'(t) = \vec{v}$

$$\begin{aligned} (f\circ\gamma)'(0) &= \lim_{h\to 0}\frac{1}{h}\left(f(\gamma(h)) - f(\gamma(0))\right) = \lim_{h\to 0}\frac{1}{h}\left(f(X_0+h\vec{v}) - f(X_0)\right) \\ &= \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\left(X_0\right)\cdot v_i = \text{ grad } f(X_0)^T\cdot v \end{aligned}$$

Definition Ist $f : D \to \mathbb{R}$ in einer Umgebung von $X_0 \in D$ stetig partiell differenzierbar und $\vec{v} \in \mathbb{R}^n$ ein Richtungsvektor, so heißt

$$\partial_{\vec{v}} f(X_0) = \lim_{h \to 0} \frac{1}{h} \left(f(X_0 + h\vec{v}) - f(X_0) \right) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0) \cdot v_i = \text{ grad } f(X_0)^T \cdot \vec{v}$$

die **Ableitung von f in Richtung** $\vec{v}$ (und im Punkt X_0).

Speziell für $\vec{v} = \vec{e_i}$ erhalten wir:

$$\partial_{\vec{v}} f(X_0) = \frac{\partial f}{\partial x_i}(X_0)$$

- Die Richtungsableitung in Richtung der i-ten Einheitsvektoren stimmt also mit der partiellen Ableitung nach x_i überein.
- **Geometrische Interpretation** des Gradienten: Aus der Cauchy-Schwarzschen Ungleichung folgt für $\|\vec{v}\| = 1$

  $$\lim_{h \to 0} \frac{1}{h} \left(f(X_0 + h\vec{v}) - f(X_0) \right) = \partial_{\vec{v}} f(X_0) = \text{ grad } f(X_0)^T \vec{v} \leq \| \text{ grad } f(X_0)\|$$

  und falls grad $f(X_0) \neq 0$, so gilt die Gleichheit genau dann, wenn

  $$\vec{v} = \frac{\text{grad} f(X_0)}{\|\text{grad} f(X_0)\|}$$

  Der **Gradient von** f in X_0 **zeigt** also stets **in** die **Richtung des steilsten Anstiegs** der Funktion!

Beispiele

(i) $f(x, y) = x^2 + y^2$, also grad $f(x, y) = 2 \begin{bmatrix} x \\ y \end{bmatrix}$

Für die Richtungsableitungen erhält man $\partial_{\vec{v}} f(x, y) = 2(xv_1 + yv_2)$.

Insbesondere wird $\partial_{\vec{v}} f(x, y)$ maximal für $\vec{v} = \frac{1}{\sqrt{x^2+y^2}} \begin{bmatrix} x \\ y \end{bmatrix}$ und 0 für $\vec{v} \perp \begin{bmatrix} x \\ y \end{bmatrix}$

(ii) $f(x, y) = x^2 - y^2$, also grad$f(x, y) = 2 \begin{bmatrix} x \\ -y \end{bmatrix}$

Für die Richtungsableitungen erhält man $\partial_{\vec{v}} f(x, y) = 2(xv_1, -yv_2)$

Insbesondere ist $\partial_{\vec{v}} f(x, y) = 0$ für $\vec{v} = \begin{bmatrix} x \\ y \end{bmatrix}$

4.2 Die totale Ableitung

Ist $f : [a, b] \to \mathbb{R}$ differenzierbar in $x_0 \in]a, b[$, so beschreibt die Tangentengleichung

$$g : x \mapsto f(x_0) + f'(x_0)(x - x_0)$$

eine **lineare Näherung** von f in x_0.
Die Güte der Approximation kann man dabei wie folgt beschreiben:

$$f(x)-g(x) = f(x)-(f(x_0)+f'(x_0)(x-x_0)) = \left(\underbrace{\frac{f(x)-f(x_0)}{x-x_0} - f'(x_0)}_{\to 0 \text{ für } x\to x_0}\right)(x-x_0) = o\left(|x - x_0|\right)$$

Notation (Landau Symbol)

Es sei $D \subset \mathbb{R}^n$ und $X_0 \in D$. Für zwei Funktionen $f, g : D \to \mathbb{R}$ schreiben wir

$$f(X) = g(X) + o\left(\|X - X_0\|^k\right)$$

falls

$$\lim_{X\to X_0} \frac{f(X) - g(X)}{\|X - X_0\|^k} = 0\,.$$

Das Landausche Symbol $o(\|X - X_0\|^k)$ besagt also, dass der bei der Approximation von f durch g in der Nähe um X_0 gemachte Fehler von einer kleineren Ordnung als $\|X - X_0\|^k$ ist.

Wir wollen den Gedanken der linearen Approximation von Funktionen einer reellen Veränderlichen auf Funktionen in mehreren Veränderlichen übertragen. Statt Näherungen von f durch eine Geradengleichung

$$g(x) = f(x_0) + f'(x_0)(x - x_0)$$

betrachten wir die Näherung von f durch Abbildungen der Form $f(X_0) + \vec{a}^T \cdot (X - X_0)$ für einen Vektor $\vec{a} \in \mathbb{R}^n$.

Definition Es sei $D \subset \mathbb{R}^n$, $X_0 \in D$ innerer Punkt und $f : D \to \mathbb{R}$ eine Funktion. f heißt in X_0 **total differenzierbar** (oder **linear approximierbar**), falls ein Vektor $\vec{a} \in \mathbb{R}^n$ existiert, so dass

$$f(X) = f(X_0) + \vec{a}^T \cdot (X - X_0) + o\left(\|X - X_0\|\right)$$

für X in einer Umgebung von X_0 gilt.

Ist f in $X_0 \in D$ total differenzierbar, so gilt

- f ist in X_0 stetig
- f ist in X_0 partiell differenzierbar und es gilt

$$\vec{a} = \text{ grad } f(X_0) = \left[\frac{\partial f}{\partial x_1}(X_0), \ldots, \frac{\partial f}{\partial x_n}(X_0)\right]^T$$

  In den Komponenten des Vektors $\vec{a}$ stehen also die partiellen Ableitungen von f in X_0. Insbesondere ist der Vektor $\vec{a}$ eindeutig bestimmt.

Aus der partiellen Differenzierbarkeit folgt im allgemeinen nicht die totale Differenzierbarkeit. Jedoch gilt, falls $D \subset \mathbb{R}^n$ offen:

$f : D \to \mathbb{R}$ stetig partiell differenzierbar

$\Longrightarrow \qquad f$ total differenzierbar (in allen Punkten $X_0 \in D$)

Beispiele

(i) $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = 2 - x^2 - y^2$, also grad$f(x,y) = -2[x,y]^T$

Da f stetig partiell differenzierbar, ist f insbesondere total differenzierbar und die lineare Näherung von f im Punkte $[x_0, y_0]^T \in \mathbb{R}^2$ ist

$$f(x,y) = 2 - x_0^2 - y_0^2 - 2x_0(x - x_0) - 2y_0(y - y_0) + o\left(\left\| \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \right\|\right)$$

etwa in $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix}$ gilt

$$f(x,y) = 0 - 2\sqrt{2}\left(x - \sqrt{2}\right) + o\left(\sqrt{\left(x - \sqrt{2}\right)^2 + y^2}\right)$$

4.3 Die Taylorformel für Funktionen in n Variablen

Die Taylorformel für die Approximation einer differenzierbaren Funktion

$$f \ :]a,b[\to \mathbb{R}$$

durch Polynome der Form

$$f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(m)}(x_0)}{m!}(x - x_0)^m$$

lässt sich auf differenzierbare Funktionen in mehreren Variablen verallgemeinern:

Satz Es sei $D \subset \mathbb{R}^n$ offen, $f : D \to \mathbb{R}$ $(m+1)$-mal stetig partiell differenzierbar, $X_0 \in D, \vec{v} \in \mathbb{R}^n$, so dass die Verbindungsstrecke

$$\{X_0 + t\vec{v} \ : \ t \in [0,1]\}$$

zwischen X_0 und $X_0 + \vec{v}$ ganz in D verläuft. Dann gilt die **Taylorformel**

$$f(X_0 + \vec{v}) = f(X_0) + \partial_{\vec{v}} f(X_0) + \frac{1}{2!}\partial_{\vec{v}}^2 f(X_0) + \ldots + \frac{1}{m!}\partial_{\vec{v}}^m f(X_0) + R_{m+1}(X_0, \vec{v})$$

mit dem Restglied

$$R_{m+1}(X_0, \vec{v}) = \frac{1}{(m+1)!}\partial_{\vec{v}}^{m+1} f(X_0 + \xi\vec{v})$$

für ein $\xi \in [0,1]$

Mit der Substitution $\vec{v} = X - X_0$ erhält man aus

$$f(X_0) + \partial_{\vec{v}} f(X_0) + \ldots + \frac{1}{m!}\partial_{\vec{v}}^m f(X_0)$$

ein Polynom $p(X)$ vom Grade m in den Variablen $x_1, \dots, x_n$. Wie im Falle $n = 1$ heißt p **m-tes Taylorpolynom von f im Entwicklungspunkt** X_0. Mit diesem Polynom gilt dann

$$f(X) = p(X) + o\left(\|X - X_0\|^m\right)$$

Die wichtigen Spezialfälle $m = 1, 2$

$m = 1$: **Lineare Approximation** (f zweimal stetig partiell differenzierbar)

$$\begin{aligned} f(X_0 + \vec{v}) &= f(X_0) + \partial_{\vec{v}} f(X_0) + R_2(X_0, \vec{v}) \\ &= f(X_0) + \text{ grad } f(X_0)^T \cdot \vec{v} + R_2(X_0, \vec{v}) \end{aligned}$$

mit $R_2(X_0, \vec{v}) = \frac{1}{2!}\partial_{\vec{v}}^2 f(X_0 + \xi\vec{v})$ für ein $\xi \in [0, 1]$.
Mit der Substitution $\vec{v} = X - X_0$ erhält man hieraus

$$f(X) = \underbrace{f(X_0) + \text{ grad } f(X_0)^T \cdot (X - X_0)}_{p(X)} + o\left(\|X - X_0\|\right)$$

$m = 2$: **Quadratische Approximation** (f dreimal stetig partiell differenzierbar)

$$\begin{aligned} f(X_0 + \vec{v}) &= f(X_0) + \partial_{\vec{v}} f(X_0) + \partial_{\vec{v}}^2 f(X_0) + R_3(X_0, \vec{v}) \\ &= f(X_0) + \text{ grad } f(X_0)^T \cdot \vec{v} + \frac{1}{2}\vec{v}^T \cdot H_f(X_0)\vec{v} + R_3(X_0, \vec{v}) \end{aligned} \tag{4.2}$$

mit $R_3(X_0, \vec{v}) = \frac{1}{3!}\partial_{\vec{v}}^3 f(X_0 + \xi\vec{v})$ für ein $\xi \in [0, 1]$.

Hierbei ist

$$H_f(X_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(X_0) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(X_0) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(X_0) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(X_0) & \frac{\partial^2 f}{\partial x_2^2}(X_0) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(X_0) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(X_0) & \frac{\partial^2 f}{\partial x_n \partial x_2}(X_0) & \dots & \frac{\partial^2 f}{\partial x_n^2}(X_0) \end{bmatrix} = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(X_0)\right]$$

die $n \times n$-Matrix der partiellen Ableitungen zweiter Ordnung von f in X_0.
Die Matrix $H_f(X_0)$ heißt **Hesse-Matrix von f in** X_0. Da die partiellen Ableitungen vertauscht werden können

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(X_0) = \frac{\partial^2 f}{\partial x_j \partial x_i}(X_0) \qquad \text{für} \quad 1 \le i, j \le n$$

ist $H_f(X_0)$ eine **symmetrische Matrix**, d.h. es gilt $H_f(X_0)^T = H_f(X_0)$.
Mit der Substitution $\vec{v} = X - X_0$ erhält man aus (4.2)

$$f(X) = \underbrace{f(X_0) + \text{ grad } f(X_0)^T \cdot (X - X_0) + \frac{1}{2}(X - X_0)^T \cdot H_f(X_0)(X - X_0)}_{=p(X)} + o\left(\|X - X_0\|^2\right)$$

Beispiele

(i) $f(x,y) = e^{xy}$ ist 3-mal stetig partiell differenzierbar

$$\operatorname{grad} f(x_0,y_0) = e^{x_0y_0}\begin{bmatrix} y_0 \\ x_0 \end{bmatrix}$$
$$H_f(x_0,y_0) = e^{x_0y_0}\begin{bmatrix} y_0^2 & 1+x_0y_0 \\ 1+x_0y_0 & x_0^2 \end{bmatrix}$$

Also ist

$$\begin{aligned} & e^{x_0y_0} + e^{x_0y_0}\begin{bmatrix} y_0 \\ x_0 \end{bmatrix}^T \begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix} + \frac{1}{2}e^{x_0y_0}\begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix}^T \begin{bmatrix} y_0^2 & 1+x_0y_0 \\ 1+x_0y_0 & x_0^2 \end{bmatrix}\begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix} \\ &= e^{x_0y_0} + e^{x_0y_0}\left(y_0(x-x_0) + x_0(y-y_0)\right) \\ &\quad + \frac{1}{2}e^{x_0y_0}\left(y_0^2(x-x_0)^2 + 2(1+x_0y_0)(x-x_0)(y-y_0) + x_0^2(y-y_0)^2\right) \end{aligned}$$

die quadratische Näherung für f im Punkte $\begin{bmatrix} y_0 \\ x_0 \end{bmatrix}$

(ii) $f(x,y) = x^2 + 4xy + 8y^2 + 3x + 5y + 1$

$$\operatorname{grad} f(x_0,y_0) = \begin{bmatrix} 2x+4y+3 \\ 4x+16y+5 \end{bmatrix}$$
$$H_f(x_0,y_0) = \begin{bmatrix} 2 & 4 \\ 4 & 16 \end{bmatrix}$$

also ist die quadratische Näherung in $[x_0,y_0]^T$gegeben durch

$$\begin{aligned} & (x_0^2 + 4x_0y_0 + 8y_0^2 + 3x_0 + 5y_0 + 1) \\ &\qquad + ((2x_0+4y_0+3)(x-x_0) + (4x_0+16y_0+5)(y-y_0)) \\ &\qquad + \frac{1}{2}\left(2(x-x_0)^2 + 8(x-x_0)(y-y_0) + 16(y-y_0)^2\right) \\ &= \cdots = 1 + 3x + 5y + x^2 + 4xy + 8y^2 = f(x,y) \qquad (!) \end{aligned}$$

d.h., wie im Falle $n = 1$ stimmt das Taylorpolynom zweiter Ordnung mit f überein, wenn f ein Polynom vom Grad ≤ 2 ist. Analoges gilt für Polynome höheren Grades.

Anwendung: Bestimmung von Extremalstellen

Wie im Falle einer Funktion einer reellen Variablen definieren wir:

Eine Funktion $f : D \to \mathbb{R}$ hat in $x_0 \in D$ ein **lokales Maximum (Minimum)**, falls ein $\varepsilon > 0$ existiert mit

$$f(X_0) \geq f(X) \quad (\text{bzw.} \quad f(X_0) \leq f(X))$$

für alle $X \in U_\varepsilon(X_0)$.

Ist $f(X_0) \geq f(X)$ (bzw. $f(X_0) \leq f(X)$) für alle $X \in D$, so heißt X_0 **absolutes Maximum (Minimum)**.

Notwendiges Kriterium

Wie im Falle einer Funktion einer reellen Variablen gilt: Ist $f : D \to \mathbb{R}$ partiell differenzierbar und X_0 lokales Extremum (also X_0 eine lokales Maximum oder lokales Minimum), so ist $\operatorname{grad} f(X_0) = 0$. Man nennt die Nullstellen des Gradienten von f, also die Stellen X_0 mit $\operatorname{grad} f(X_0) = 0$ **kritische Stellen**.

Möchte man also (lokale) Maxima (bzw. (lokale) Minima) einer Funktion bestimmen, so sucht man zunächst die Menge der kritischen Stellen. Diese kritischen Stellen sind die Kandidaten für lokale Extrema. Man versucht dann mit den folgenden hinreichenden Kriterien für lokale Maxima und Minima zu entscheiden, ob es sich bei einer gegebenen kritischen Stelle um ein lokales Maximum oder um eine lokales Minimum handelt. Zur Bestimmung aller lokalen Extrema muss man schließlich gegebenenfalls auch noch das Verhalten der Funktion am Rand des Definitionsbereiches untersuchen.

Für hinreichende Kriterien benötigen wir wie im Falle von Funktionen einer reellen Veränderlichen die höheren Ableitungen von f. Bevor wir ein hinreichendes Kriterium formulieren können, benötigen wir noch Informationen über positiv definite Matrizen.

Einschub: Positiv definite Matrizen

Definition Es sei $A = [a_{ij}]$ eine **symmetrische** $n \times n$-Matrix (also $A^T = A$, d.h. $a_{ij} = a_{ji}$ für $1 \leq i, j \leq n$). A heißt **positiv definit**, falls gilt

$$\vec{x}^T A \vec{x} > 0 \quad \text{für alle} \quad \vec{x} \in \mathbb{R}^n, \vec{x} \neq 0\,.$$

A heißt **negativ definit**, falls $-A$ positiv definit ist, und **indefinit**, falls $\vec{x}, \vec{y} \in \mathbb{R}^n$ existieren mit

$$\vec{x}^T A \vec{x} > 0 \quad \text{und} \quad \vec{y}^T A \vec{y} < 0$$

Beispiele

(i) $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ist positiv definit, denn

$$\begin{bmatrix} x \\ y \end{bmatrix}^T A \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + y^2 = \left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\|^2 > 0 \quad \text{für} \quad \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2, \begin{bmatrix} x \\ y \end{bmatrix} \neq 0$$

(ii) $A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ ist indefinit, denn

$$\vec{e_1}^T \underbrace{A\vec{e_1}}_{=-\vec{e_1}} = -\left\|\vec{e_1}\right\|^2 = -1$$

und

$$\vec{e_2}^T \underbrace{A\vec{e_2}}_{=-\vec{e_2}} = -\left\|\vec{e_2}\right\|^2 = +1$$

Bemerkung 4.3

(i) Wir haben in Mathematik I für MB gesehen, dass für eine $n \times n$-Matrix A alle Eigenwerte von A reell sind. Weiter gilt dann:

$$\begin{aligned} A \text{ positiv definit} &\quad\Leftrightarrow\quad \text{alle Eigenwerte von } A \text{ positiv} \\ A \text{ negativ definit} &\quad\Leftrightarrow\quad \text{alle Eigenwerte von } A \text{ negativ} \\ A \text{ indefinit} &\quad\Leftrightarrow\quad A \text{ besitzt mindestens einen positiven} \\ &\qquad\quad \text{und mindestens einen negativen Eigenwert} \end{aligned}$$

(ii) Das folgende Kriterium charakterisiert positive Definitheit einer Matrix mit Hilfe von Determinanten. Es gilt:

A ist genau dann positiv definit, wenn für $k = 1, 2, \ldots, n$ gilt

$$\det \begin{bmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \ldots & a_{kk} \end{bmatrix} > 0$$

(iii) Für 2×2-Matrizen ist dieses Kriterium besonders einfach anzuwenden: die Matrix

$$A = \begin{bmatrix} a & b \\ b & d \end{bmatrix} \qquad \text{ist}$$

ist

$$\begin{aligned} \text{positiv definit} &\quad\Leftrightarrow\quad a > 0 \quad \text{und } \det A = ad - b^2 > 0 \\ \text{negativ definit} &\quad\Leftrightarrow\quad a < 0 \quad \text{und } \det A = ad - b^2 > 0 \\ \text{indefinit} &\quad\Leftrightarrow\quad \det A = ad - b^2 < 0 \end{aligned}$$

Kehren wir zurück zu hinreichenden Kriteria für lokale Extrema einer Funktion in mehreren reellen Veränderlichen. Mit Hilfe der Aussagen über positiv definite Matrizen erhält man nun:

Satz Es sei $f : D \to \mathbb{R}$ zweimal stetig partiell differenzierbar und $X_0 \in D$ mit grad $f(X_0) = 0$. Es sei $H_f(X_0) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(X_0)\right]$ die Hesse-Matrix von f in X_0. Dann gilt:

- $H_f(X_0)$ positiv definit $\Rightarrow$ X_0 ist lokales Minimum
- $H_f(X_0)$ negativ definit $\Rightarrow$ X_0 ist lokales Maximum
- $H_f(X_0)$ indefinit $\Rightarrow$ X_0 ist kein lokales Extremum (sondern Sattelpunkt)

Spezialfall $n = 2$

Im Spezialfall $n = 2$ erhält man aufgrund von Bemerkung 4.3 (iii) insbesondere für $f(x, y)$:

$$H_f(x_0, y_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(x_0, y_0) & \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) \\ \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) & \frac{\partial^2 f}{\partial^2 y}(x_0, y_0) \end{bmatrix}$$

und ist

$$\Delta := \det H_f(x_0,y_0) = \frac{\partial^2 f}{\partial^2 x}(x_0,y_0)\cdot\frac{\partial^2 f}{\partial^2 y}(x_0,y_0) - \left(\frac{\partial^2 f}{\partial x\partial y}(x_0,y_0)\right)^2$$

so gilt:

- $\Delta > 0$ und $\frac{\partial^2 f}{\partial^2 x}(x_0,y_0) > 0 \qquad \Rightarrow \qquad \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ ist lokales Minimum
- $\Delta > 0$ und $\frac{\partial^2 f}{\partial^2 x}(x_0,y_0) < 0 \qquad \Rightarrow \qquad \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ ist lokales Maximum
- $\Delta < 0 \qquad \Rightarrow \qquad \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ ist kein lokales Extremum (sondern Sattelpunkt).

Beispiele Es sei $f(x,y) = -x^4 - y^4 + 2x^2 + 2y^2$. Dann gilt

$$\text{grad } f(x,y) = 4\begin{bmatrix} -x^3 + x \\ -y^3 + y \end{bmatrix}$$

und dieser besitzt neun Nullstellen:
$A_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$,
$A_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $A_3 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$, $A_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $A_5 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$,
$A_6 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $A_7 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $A_8 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, $A_9 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$
Für die Hesse-Matrix

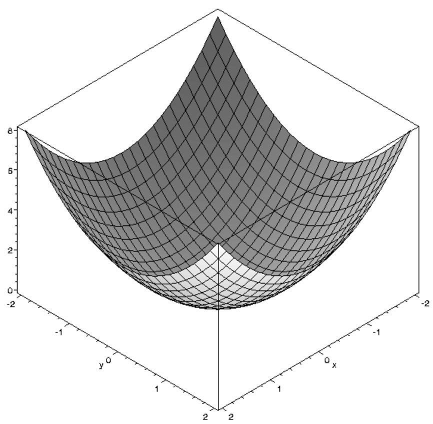
$$H_f(x,y) = 4\begin{bmatrix} -3x^2+1 & 0 \\ 0 & -3y^2+1 \end{bmatrix}$$

gilt in diesen kritischen Punkten:

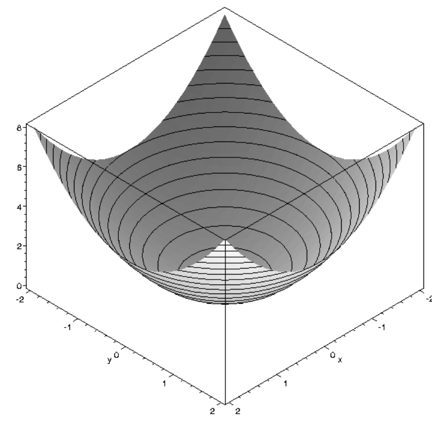
$H_f(A_1) = 4\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ positiv definit $\quad\Rightarrow\quad A_1$ lokales Minimum

$H_f(A_2) = 4\begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}$ indefinit $\quad\Rightarrow\quad A_2$ Sattelpunkt (ebenso A_3, A_4, A_5)

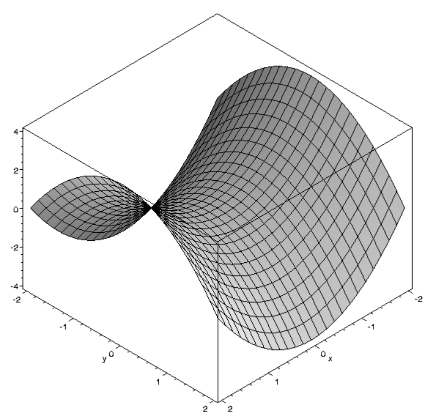
$H_f(A_6) = 4\begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$ negativ definit $\quad\Rightarrow\quad A_1$ lokales Maximum (ebenso A_7, A_8, A_9)



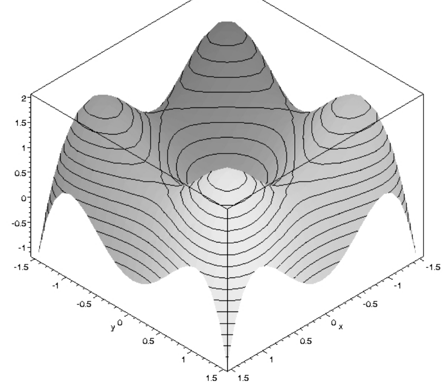
$(i)\quad f(x,y)=x^2+y^2$ mit Schnittkurven



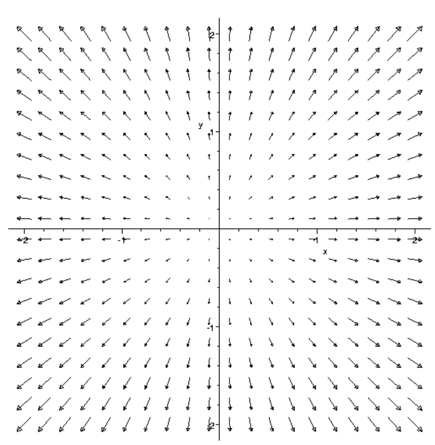
$(ii)\quad f(x,y)=x^2+y^2$ mit Höhenlinien



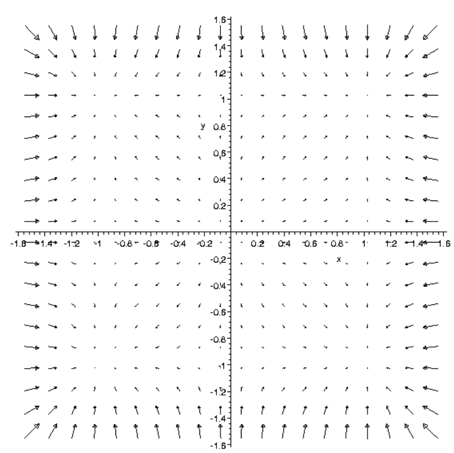
$(iii)\quad f(x,y)=x^2-y^2$



$(iv)\quad f(x,y)=-x^4-y^4+2\,x^2+2\,y^2$



(v) Gradientenfeld zu (i)



(vi) Gradientenfeld zu (iv)

5 Implizite Funktionen und Extrema mit Nebenbedingungen

5.1 Vektorfelder

Bisher haben wir nur reellwertige Abbildungen $f : D \to \mathbb{R}$ betrachtet. Nun wollen wir auch Abbildungen

$$F : D \to \mathbb{R}^m \qquad \text{für} \quad D \subset R^n\,, n, m \geq 1$$

betrachten. Eine solche Abbildung nennen wir **Vektorfeld**.

Schreiben wir $F = [F_1, \ldots, F_m]^T$, so erhalten wir zu F die **Komponentenfunktionen** des Vektorfeldes

$$F_i : D \to \mathbb{R}\,, \qquad 1 \leq i \leq m\,.$$

- **n=1** Ein Vektorfeld $F : D \subset \mathbb{R} \to \mathbb{R}^m$ ist nichts weiteres als eine Kurve im $\mathbb{R}^m$.
- **m=1** Ein Vektorfeld $F : D \subset \mathbb{R}^n \to \mathbb{R}$ ist nichts weiteres als eine reellwertige Funktion in n Veränderlichen.

Beispiel Strömung einer zähen Flüssigkeit durch ein Rohr mit Radius r

Es bezeichne $D \subset \mathbb{R}^3$ ein Rohr mit Radius r dessen Mittelpunkt durch die x-Achse gegeben ist. Dann beschreibt das Vektorfeld

$$\begin{aligned} v &: D \to \mathbb{R}^3 \\ [x, y, z]^T &\mapsto c[0, r^2 - y^2 - z^2, 0]^T \qquad y^2 + z^2 \leq r^2\,, c > 0\,. \end{aligned}$$

das Geschwindigkeitsfeld einer zähen Flüssigkeit durch D.

Definition Ein Vektorfeld $F : \mathbb{R}^n \to \mathbb{R}^m$ heißt

- **stetig (in X_0)**, falls alle Komponentenfunktionen F_i stetig (in X_0) sind,
- **partiell differenzierbar (in X_0)**, falls alle Komponentenfunktionen F_i partiell differenzierbar (in X_0) sind,
- **total differenzierbar in X_0** oder **linear approximierbar**, falls eine $m \times n$-Matrix A existiert, so dass

  $$F(X) = F(X_0) + A(X - X_0) + o\left(\|X - X_0\|\right)$$

  für X in einer Umgebung von X_0.

  Ist F total differenzierbar in X_0, so ist F stetig in X_0 und partiell differenzierbar in X_0 und es gilt

  $$A = \left[\frac{\partial F_i}{\partial x_j}(X_0)\right] \qquad (= J_F(X_0) \text{ siehe unten })\,.$$

Fassen wir alle partiellen Ableitungen eines Vektorfeldes F im Punkt X_0 in einer Matrix zusammen, so erhält man die **Funktionalmatrix von F in X_0**:

$$\begin{pmatrix} \frac{\partial F_1}{\partial x_1}(X_0) & \frac{\partial F_1}{\partial x_2}(X_0) & \dots & \frac{\partial F_1}{\partial x_n}(X_0) \\ \frac{\partial F_2}{\partial x_1}(X_0) & \frac{\partial F_2}{\partial x_2}(X_0) & \dots & \frac{\partial F_2}{\partial x_n}(X_0) \\ \vdots & \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1}(X_0) & \frac{\partial F_m}{\partial x_2}(X_0) & \dots & \frac{\partial F_m}{\partial x_n}(X_0) \end{pmatrix} = \left[\frac{\partial F_i}{\partial x_j}(X_0)\right]$$

$J_F(X_0)$ ist also eine $m \times n$-Matrix. In den Zeilen stehen die partiellen Ableitungen der Komponenten des Vektorfeldes. Ist $m = n$, so ist $J_F(X_0)$ also eine quadratische Matrix. Ihre Determinante $\det(J_F(X_0))$ heißt **Funktionaldeterminante von F in X_0**.

Beispiel Polarkoordinaten im R^2

$$\begin{aligned} G\ :&]0,\infty[\times[0,2\pi[\to \mathbb{R}^2 \\ & [r,\varphi]^T \mapsto [r\cos\varphi, r\sin\varphi]^T \end{aligned}$$

also $G_1(r,\varphi) = r\cos\varphi$, $G_2(r,\varphi) = r\sin\varphi$ und

$$J_G(r,\varphi) = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}$$

und $\det(J_G(r,\varphi)) = r\left((\cos\varphi)^2 + (\sin\varphi)^2\right) = r$.

Mit Hilfe der Funktionalmatrix J_G lässt sich die Kettenregel (4.1) nun wie folgt umschreiben: Ist $f\ :\ \mathbb{R}^2 \to \mathbb{R}$ stetig partiell differenzierbar, so ist

$$F = f \circ G\ :]0,\infty[\times[0,2\pi[\to \mathbb{R}$$

stetig partiell differenzierbar und es gilt

$$J_F(r,\varphi) = J_f(G(r,\varphi)) \cdot J_G(r,\varphi)\,.$$

$\cdot$ bezeichnet hierbei das Matrizenprodukt der beiden Funktionalmatrizen. In dieser Form lässt sich die Kettenregel auf allgemeine Vektorfelder verallgemeinern.

Satz (Kettenregel für Vektorfelder) Es sei $D \subset R^n$ offen, $U \subset R^m$ offen,

$$G : D \to U \qquad \text{und} \qquad F : U \to \mathbb{R}^k$$

stetig partiell differenzierbare Vektorfelder. Dann ist die Verkettung

$$\begin{aligned} H = F \circ G\ :& D \to \mathbb{R}^k \\ & X \mapsto F(G(X)) \end{aligned}$$

stetig (partiell) differenzierbar und es gilt für $X_0 \in D$

$$J_H(X_0) = J_{F\circ G}(X_0) = \underbrace{\underbrace{J_F(G(X_0))}_{\text{äußere}} \cdot \underbrace{J_G(X_0)}_{\text{innere}}}_{\text{Ableitung}}$$

Die Funktionalmatrix der Verkettung berechnet sich also als Matrizenprodukt der beiden Funktionalmatrizen $J_F(G(X_0))$ mit $J_G(X_0)$. Da $J_F(G(X_0))$ eine $k \times m$-Matrix und $J_G(X_0)$ eine $m \times n$-Matrix, ist das Matrizenprodukt eine $k \times n$-Matrix, wie es $J_H(X_0)$ auch sein muss.

5.2 Implizite Funktionen

Um eine Funktion $g : D \subset \mathbb{R} \to \mathbb{R}$ mit den Methoden der Differentialrechnung zu diskutieren, bedarf es eines expliziten Ausdrucks für die Funktion $g(x) = y$. Mitunter ist der Zusammenhang zwischen x und y nur durch eine Gleichung

$$f(x, y) = 0 \tag{5.1}$$

gegeben, und man hat zu gegebenem x nach $y = g(x)$ aufzulösen. Gesucht ist also eine Funktion g mit

$$f(x, g(x)) = 0\,. \tag{5.2}$$

Man sagt dann, dass g durch die Gleichung (5.1) **implizit** definiert ist.

Beispiele

(i) $f(x, y) = 3x + 2y - 4$

Die Gleichung $f(x, y) = 0$ führt auf die Gleichung $3x+2y = 4$, die wir nach y auflösen können:

$$y = -\frac{3}{2}x + 2\,, \qquad \text{also } g(x) = -\frac{3}{2}x + 2\,.$$

(ii) $f(x, y) = x^2 + y^2 - 1$

Die Gleichung $f(x, y) = 0$ führt auf die Gleichung $y^2 = 1 - x^2$ mit den Lösungen $y = \pm\sqrt{1 - x^2}$, $|x| \leq 1$. Dies führt auf die beiden implizit definierten Funktionen

$$g_1(x) = \sqrt{1 - x^2}\,, \qquad g_2(x) = -\sqrt{1 - x^2}$$

Wenn wir nun einen Punkt $[x_0, y_0]^T$ mit $f(x_0, y_0) = 0$ festhalten, so können wir, falls $y_0 \neq 0$, die Gleichung $f(x, y) = 0$ in einer Umgebung von $[x_0, y_0]^T$ eindeutig auflösen, denn in einer Umgebung wird das Vorzeichen von y nicht wechseln.

Satz Es sei $f : D =]a, b[\times]c, d[\to \mathbb{R}$ stetig partiell differenzierbar, $[x_0, y_0]^T \in D$ ein Punkt mit $f(x_0, y_0) = 0$ und $f_y(x_0, y_0) \neq 0$. Dann gibt es offene Intervalle $I \subset]a, b[$ mit $x_0 \in I$ und $J \subset]c, d[$ mit $y_0 \in J$, so dass $f_y(x, y) \neq 0$ für alle $[x, y]^T \in I \times J$, und es gibt eine stetig differenzierbare Funktion

$$g : I \to \mathbb{R} \qquad \text{mit } g(I) \subset]c, d[\,, g(x_0) = y_0\,,$$

so dass $f(x, g(x)) = 0$ für alle $x \in I$. Für die Ableitung von g gilt

$$g'(x) = -\frac{f_x(x, g(x))}{f_y(x, g(x))}\,, x \in I\,. \tag{5.3}$$

Beispiel Es sei $f(x, y) = y^2 + \sin x - 4$

Dann ist $[0, 2]^T$ Lösung von $f(x, y) = 0$ und $f_y(0, 2) = 4$, also ist die Gleichung

$$y^2 + \sin x - 4 = 0$$

in 0 lokal eindeutig nach y auflösbar. Für die durch die eindeutig bestimmte Lösung implizit definierte Funktion g gilt: $g(0) = 2$ und

$$g^2(x) + \sin x - 4 = 0$$

also $g(x) = \sqrt{4 - \sin x}$. Für die Ableitung gilt

$$g'(x) = -\frac{\cos x}{2g(x)} = -\frac{\cos x}{2\sqrt{4 - \sin x}}$$

was man durch direktes Ableiten von g bestätigen kann.

Bemerkung Die Formel (5.3) für die Ableitung der implizit definierten Funktion g folgt aus der Kettenregel, denn aus

$$f(x, g(x)) = 0$$

folgt durch Ableiten nach x:

$$f_x(x, g(x)) + f_y(x, g(x))\, g'(x) = 0$$

also

$$g'(x) = -\frac{f_x(x, g(x))}{f_y(x, g(x))}\,.$$

Der Satz über implizite Funktionen gilt auch allgemeiner:

Satz (Hauptsatz über implizite Funktionen) Es seien $U \subset \mathbb{R}^n$ offen, $V \subset \mathbb{R}^m$ offen und

$$F\,:\, U \times V \to \mathbb{R}^m\,, \qquad [X, Y]^T \mapsto F(X, Y)$$

stetig differenzierbar. Weiterhin sei $[X_0, Y_0]^T \in U \times V$ eine Lösung von $F(X, Y) = 0$ und die $m \times m$-Matrix

$$\frac{\partial F}{\partial Y}(X_0, Y_0) := \left[\frac{\partial F_i}{\partial y_j}(X_0)\right]$$

sei invertierbar (also $\det(\frac{\partial F}{\partial Y}(X_0, Y_0)) \neq 0$). Dann gibt es offene Mengen $U_0 \subset U$ mit $X_0 \in U_0$ und $V_0 \subset V$ mit $Y_0 \in V$, so dass $\frac{\partial F}{\partial Y}(X, Y)$ invertierbar für alle $[X, Y]^T \in U_0 \times V_0$, und es gibt ein stetig differenzierbares Vektorfeld

$$G\,:\, U_0 \to V_0 \qquad \text{mit} \qquad G(X_0) = Y_0$$

und

$$F(X, G(X)) = 0 \qquad \text{für alle} \qquad X \in U_0$$

sowie

$$J_G(X) = -\frac{\partial F}{\partial Y}(X, G(X))^{-1} \cdot \frac{\partial F}{\partial X}(X, G(X))\,.$$

5.3 Extrema unter Nebenbedingungen

Gegeben sei eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$. Mitunter ist es notwendig, ein Extremum von f nicht auf ganz $\mathbb{R}^n$ zu bestimmen, sondern nur auf einer Teilmenge, die mit Hilfe einer weiteren Funktion $g : \mathbb{R}^n \to \mathbb{R}$ durch eine Gleichung $g(X) = 0$ implizit definiert ist.

Definition Ein Punkt X_0 mit $g(X_0) = 0$ heißt **relatives Maximum** (bzw. **Minimum**) unter der **Nebenbedingung** $g(X) = 0$, falls

$$f(X_0) \geq f(X) \qquad (\text{ bzw. } f(X_0) \leq f(X))$$

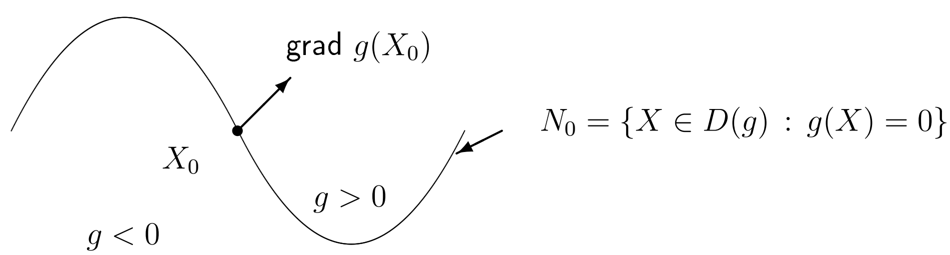
für alle $X \in U_\varepsilon(X_0) \cap \{X \in \mathbb{R}^n \,:\, g(X) = 0\}$.

Der fogende Satz liefert eine notwendige Bedingung die zum Auffinden relativer Extrema unter Nebenbedingungen nützlich ist.

Satz Sind f, g auf einer Umgebung $U_\varepsilon(X_0)$ von X_0 stetig differenzierbar, hat f in X_0 ein relatives Extremum unter der Nebenbedingung $g(X) = 0$ und ist $\text{grad } g(X_0) \neq 0$, so gibt es ein $\lambda \in \mathbb{R}$ mit

$$\begin{aligned} & \text{grad } f(X_0) + \lambda \text{ grad } g(X_0) = 0\,, \\ \text{d.h.} \qquad & f_{x_i}(X_0) + \lambda \quad g_{x_i}(X_0) \quad = 0 \qquad \text{für } i = 1, \ldots, n\,. \end{aligned} \tag{5.4}$$

Geometrische Intepretation von (5.4)



Der Gradient $\text{grad } g(X_0)$ zeigt stets in eine Richtung, die aus $N_0 := \{X \in D(g) : g(X) = 0\}$ **hinausführt**, und zwar genauer in die Menge $g > 0$ **hinein**. Kurven, die ganz in N_0 verlaufen, haben Tangentialvektoren, die senkrecht zu $\text{grad } g(X_0)$ sind, denn: Ist

$$\gamma :]-\varepsilon, \varepsilon[\to N_0$$

stetig differenzierbare Kurve mit $\gamma(0) = X_0$, so folgt aus $g(\gamma(t)) = 0$ für alle $t \in]-\varepsilon, \varepsilon[$ mit Hilfe der Kettenregel

$$0 = \frac{d}{dt} g(\gamma(t)) = \text{ grad } g(X_0)^T \cdot \gamma'(t)$$

und somit insbesondere für $t = 0$

$$\text{grad } g(X_0)^T \cdot \gamma'(0) = 0\,.$$

Beispiel 5.1 Gegeben sei die Funktion

$$g(x, y, z) = x^2 + y^2 + z^2 - 1\,.$$

Dann ist $N_0 = \{[x, y, z]^T \in \mathbb{R}^3 \,:\, x^2 + y^2 + z^2 = 1\}$ also die Oberfläche der Einheitskugel S^1 im $\mathbb{R}^3$. Im folgenden bezeichne X_0 den Nordpol, also $X_0 = [0, 0, 1]^T$. Dann steht grad $g(X_0) = 2[0, 0, 1]^T$ senkrecht zur Tangentialebene

$$E \,:\, 0\,x + 0\,y + 1\,z = 1$$

an S^1 im Punkt X_0. Gegeben sei etwa die Kurve

$$\gamma(t) = [\sin t, 0, \cos t]^T\,, \qquad t \in]-\pi, \pi[\,.$$

Dann gilt $\gamma(0) = X_0$ und $\gamma'(0) = [\cos 0, 0, \sin 0]^T = [1, 0, 0]^T$.

Es sei nun f eine Funktion, die auf einer Umgebung von X_0 definiert ist. Der Gradient grad $f(X_0)$ zeigt immer in die Richtung des steilsten Anstieges von f. Ist also X_0 relatives Extremum, so wird der Gradient keinen Richtungsanteil haben, der tangential zu N_0 verläuft. Damit aber muss grad $f(X_0)$ in dieselbe Richtung wie grad $g(X_0)$ oder $-$ grad $g(X_0)$ zeigen. Folglich müssen grad $g(X_0)$ und grad $f(X_0)$ **linear abhängig** sein, d.h. es gibt ein $\lambda \in \mathbb{R}$ mit

$$\text{grad } f(X_0) + \lambda \text{ grad } g(X_0) = 0\,.$$

Beispiel 5.2 In der Situation des Beispiels 5.1 sei

$$f(x, y, z) = \frac{1}{2}x^4 + \frac{1}{2}y^4 + \frac{1}{2}z^4$$

gegeben. f ist monoton wachsend in x, y und z. Daher steht zu vermuten, dass f in X_0 sein Maximum unter der Nebenbedingung $g(X) = 0$ annimmt. In der Tat gilt

$$\text{grad } f(x, y, z) = 2[x^3, y^3, z^3]^T\,,$$

also

$$\text{grad } f(X_0) = 2[0, 0, 1]^T = \text{ grad } g(X_0)\,.$$

Damit ist die **notwendige Bedingung** für das Vorliegen eines relativen Extremums in X_0 erfüllt. Um zu prüfen, ob X_0 ein Maximum ist, müssen nun alle relativen Extrema untersucht werden. In diesem Falle finden wir alle relativen Extrema unter den Lösungen x, y, z und λ des nichtlinearen Gleichungssystems

$$\text{grad } f(x, y, z) + \lambda \text{ grad } g(x, y, z) = 0 \text{ mit } g(x, y, z) = 0\,,$$

also

$$2\begin{bmatrix} x^3 \\ y^3 \\ z^3 \end{bmatrix} + 2\lambda \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \qquad \text{mit} \qquad x^2+y^2+z^2=1\,.$$

Dies führt auf

$$\begin{aligned} x^3+\lambda x &= 0 \\ y^3+\lambda y &= 0 \\ z^3+\lambda z &= 0 \end{aligned} \qquad \Longrightarrow \qquad \begin{aligned} x(x^2+\lambda) &= 0 \\ y(y^2+\lambda) &= 0 \\ z(z^2+\lambda) &= 0\,. \end{aligned}$$

1. Fall $x \neq 0$, also $\lambda = -x^2$ und damit

$$\begin{aligned} y &= 0 \text{ oder } y = \pm x \\ z &= 0 \text{ oder } z = \pm x \end{aligned}$$

(I) $y = z = 0$ führt wegen der Nebenbedingung $g(x,y,z,) = 0$ auf $x = \pm 1$, also

$$f(0,0,\pm 1) = \frac{1}{2} = f(X_0)\,.$$

(II) $y = \pm x$, $z = 0$ führt entsprechend auf $x = \pm\frac{1}{\sqrt{2}}$, $y = \pm\frac{1}{\sqrt{2}}$, also

$$f\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}, 0\right) = \frac{1}{4}\,.$$

(III) $y = \pm x$, $z = \pm x$ führt auf $x = \pm\frac{1}{\sqrt{3}}$, $y = \pm\frac{1}{\sqrt{3}}$, $z = \pm\frac{1}{\sqrt{3}}$, also

$$f\left(\pm\frac{1}{\sqrt{3}}, \pm\frac{1}{\sqrt{3}}, \pm\frac{1}{\sqrt{3}}\right) = \frac{1}{6}\,.$$

Die übrigen Fälle $y \neq 0$ und $z \neq 0$ sind analog zu behandeln.

Insgesamt gibt es

- 6 Punkte vom Typ (I), die allesamt relative Maxima sind. Insbesondere ist also der Nordpol X_0 ein relatives Maximum.
- 12 Punkte vom Typ (II), die allesamt Sattelpunkte sind.
- 8 Punkte vom Typ (III), die allesamt relative Minima sind.

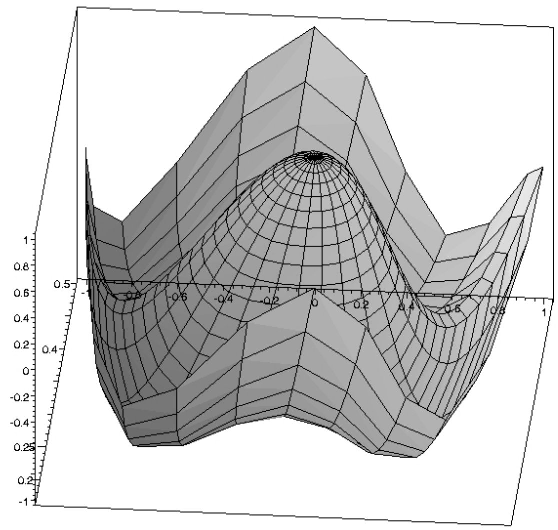
Für einen Punkt $[x,y,z]^T$ mit $z > 0$ können wir schreiben

$$z = z(x,y) = \sqrt{1-x^2-y^2}$$

und damit erhält man

$$f(x,y,z) = f(x,y,z(x,y)) = \frac{1}{2}x^4 + \frac{1}{2}y^4 + \frac{1}{2}\left(1-x^2-y^2\right)^2$$

als Funktion der zwei Variablen x und y. Die folgende Grafik enthält den zugehörigen Funktionsgraphen. Man erkennt deutlich die relativen Extrema vom Typ (I) und (III):



Beispiel Gesucht ist der minimale Abstand des Punktes $P = [1,1,1]^T \in \mathbb{R}^3$ zur Kugeloberfläche

$$K := \{X \in \mathbb{R}^3 \,:\, \|X\| = 1\}\,.$$

Gesucht ist also das Minimum des Abstandsquadrats

$$f(X) := \|X - P\|^2 \qquad\qquad (\text{ grad } f(X) = 2(X-P))$$

unter der Nebenbedingung

$$g(X) := \|X\|^2 - 1 = 0 \qquad\qquad (\text{ grad } g(X) = 2X)\,.$$

Mögliche Kandidaten für Minima sind die Punkte $X \in K$ mit

$$\text{grad } f(X) + \lambda \text{ grad } g(X) = 0$$

für ein $\lambda \in \mathbb{R}$, also diejenigen $X \in K$ mit

$$2(X-P) + 2\lambda X = 0 \qquad \Longrightarrow \qquad (1+\lambda)X = P\,.$$

Da $X \in K$, muss $|1+\lambda| = \|P\| = \sqrt{3}$ sein, also $\lambda = -1 \pm \sqrt{3}$, und damit $X = \pm\frac{1}{\sqrt{3}}P$. Es ist klar, dass $-\frac{1}{\sqrt{3}}P$ (relatives) Maximum und $\frac{1}{\sqrt{3}}P$ (relatives) Minimum der Abstandes ist. Für den minimalen Abstand erhält man also

$$\left\| \frac{1}{\sqrt{3}}P - P \right\| = \sqrt{3} - 1\,.$$