

Mathematik IT 3 (Analysis)

für die Studiengänge Informatik/Diplom, Informatik/Bachelor, Informations- und Medientechnik/Bachelor
 und eBusiness/ Bachelor, 3. Semester im Wintersemester 2011/2012

Differentialrechnung im $\mathbb{R}^n$

Die nachfolgende Übersicht stellt im wesentlichen den Inhalt der Vorlesungen vom 01. und 02. Februar 2012 dar, in der die Grundidee der Berechnung von lokalen Extremwerten für Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (mit und ohne Nebenbedingungen) sowie einige Beispiele dazu behandelt wurden. Diese Zusammenfassung soll aber einige Ergänzungen dazu geben. So konnten aus zeitlichen Gründen (Ende der Vorlesungszeit) einige wichtige Begriffe, wie z.B. offene Mengen im $\mathbb{R}^n$, Richtungsableitungen oder Niveaumengen nicht behandelt werden. Dies soll hier kurz nachgeholt werden, andererseits wird ergänzend auch noch angegeben, wie Tangentialebenen berechnet werden können. Trotz allem bleiben aber einige Lücken bestehen, zum Beispiel der wichtige Begriff der Kompaktheit einer Menge im $\mathbb{R}^n$ oder hinreichende Kriterien für lokale Extrema unter Nebenbedingungen. Dazu sei auf die am Ende angegebene Literatur hingewiesen.

Bearbeitungsstand: 2. Februar 2016

Grundlegende Definitionen und Hinweise:

Für $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ bezeichne nachfolgend $\|x\| := \sqrt{x_1^2 + \dots + x_n^2}$ die euklidische Norm von x .

- Für $x \in \mathbb{R}^n$ und eine reelle Zahl $r > 0$ heißt die Menge $B(x, r) := \{x \in \mathbb{R}^n \mid \|x\| < r\}$ offene Kugel mit Mittelpunkt x und Radius r .
- Eine Teilmenge $D \subseteq \mathbb{R}^n$ heißt Umgebung eines Punktes $x \in D$, falls ein $\varepsilon > 0$ existiert, so dass $B(x, \varepsilon) \subset D$ gilt. Man nennt die offene Kugel $B(x, \varepsilon)$ auch ε -Umgebung von x .
- Eine Teilmenge $D \subseteq \mathbb{R}^n$ heißt offen, wenn zu jedem $x \in D$ ein $\varepsilon > 0$ mit $B(x, \varepsilon) \subset D$ existiert.
 - Beispiel: In $\mathbb{R}$ sind alle offenen Intervalle (a, b) offene Mengen, insbesondere ist $(x - \varepsilon, x + \varepsilon)$ für $\varepsilon > 0$ die ε -Umgebung eines Punktes $x \in \mathbb{R}$.
 - Beispiel: Im $\mathbb{R}^2$ sind zum Beispiel die Kreisscheiben $K := \{x = (x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 < r^2\}$ (also ohne den Rand $x_1^2 + x_2^2 = r^2$) Beispiele für offene Kugeln mit Mittelpunkt $(0, 0)$ und Radius r , denn für alle $x = (x_1, x_2) \in K$ hat man $\|x\| = \sqrt{x_1^2 + x_2^2} < r$. Die genannte Menge K ist zugleich auch ein Beispiel für eine offene Menge.

Nachfolgend sei $D \subseteq \mathbb{R}^n$ eine offene Menge ($n \in \mathbb{N} \setminus \{0\}$), $f : D \rightarrow \mathbb{R}$ eine Funktion und $x = (x_1, \dots, x_n) \in D$. Außerdem sei für $i = 1, 2, \dots, n$ mit e_i der i -te Einheitsvektor in $\mathbb{R}^n$ bezeichnet.

- Die Funktion $f(x)$ heißt in x nach der i -ten Koordinate x_i , $1 \leq i \leq n$ partiell differenzierbar, falls der Grenzwert

$$\frac{\partial f}{\partial x_i}(x) := \lim_{h \rightarrow 0} \frac{f(x + he_i) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{h}$$

existiert. $\frac{\partial f}{\partial x_i}(x)$ heißt dann die partielle Ableitung von $f(x)$ nach x_i im Punkt x . Die partielle Ableitung kann man auch als gewöhnliche Ableitung von Funktionen einer Veränderlichen interpretieren, d.h. zu festem $x = (x_1, \dots, x_n)$ betrachtet man die Funktionen $f_i(\alpha) := f(x_1, \dots, x_{i-1}, \alpha, x_{i+1}, \dots, x_n)$ für $i = 1, \dots, n$. Dann folgt aus der Definition der partiellen Ableitung

$$\frac{\partial f}{\partial x_i}(x) = \lim_{h \rightarrow 0} \frac{f_i(x_i + h) - f_i(x_i)}{h} = f'_i(x_i)$$

Die partielle Ableitung von f ist also nichts anderes, als die gewöhnliche Ableitung bzgl. der i -ten Variablen bei Festhaltung der übrigen $n - 1$ Variablen (siehe dazu Abb. 1 für $n = 2$).

- Ist f in allen Punkten $x \in D$ partiell differenzierbar nach x_i , dann heißt f partiell differenzierbar nach der Koordinate x_i . Entsprechend nennt man $\frac{\partial f}{\partial x_i}$ die partielle Ableitung von f nach x_i .
- Ist f für alle Koordinaten $x_1, x_2, \dots, x_n$ partiell differenzierbar, dann heißt die Funktion f partiell differenzierbar. Falls außerdem alle partiellen Ableitungen $\frac{\partial f}{\partial x_i}(x)$ ($i=1,2,\dots,n$) für alle $x \in D$ stetig sind, dann heißt $f(x)$ stetig partiell differenzierbar.

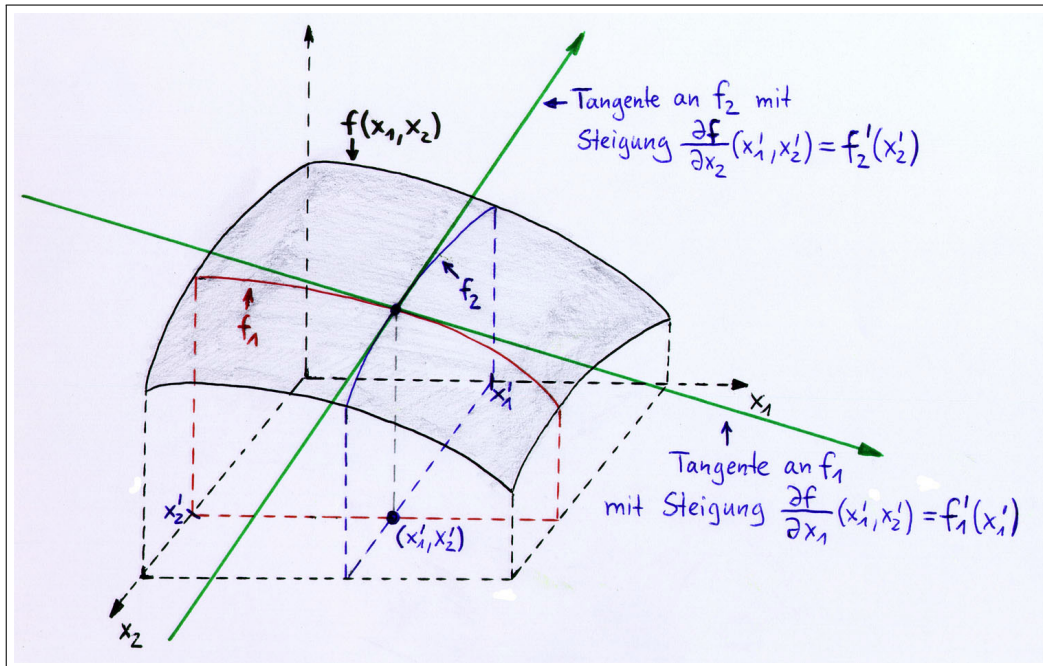


Abb. 1: Partielle Ableitungen einer Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ im Punkt (x_1', x_2') . Die partielle Ableitung nach der Variablen x_1 ergibt sich durch Festhalten der Variablen x_2 (hier also $x_2 := x_2'$) und ableiten der (nur von x_1 abhängigen) Funktion $f_1 : \mathbb{R} \rightarrow \mathbb{R}$ mit $f_1(x_1) := f(x_1, x_2')$, also $\frac{\partial f}{\partial x_1}(x_1, x_2') = \frac{df_1}{dx_1}(x_1) = f_1'(x_1)$. Die Tangente an die Funktion f_1 im Punkt x_1' hat damit die Steigung $\frac{\partial f}{\partial x_1}(x_1', x_2') = f_1'(x_1')$. Analog wird die partielle Ableitung nach x_2 gebildet, wobei sich die Rollen von x_1 und x_2 vertauschen.

- Sei f im Punkt x partiell differenzierbar. Der Zeilenvektor

$$\text{grad} f(x) := \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right)$$

heißt Gradient von $f(x)$ in x . Statt $\text{grad} f$ schreibt man auch ∇f , wobei ∇ (sprich: nabla) als vektorwertiger Differentialoperator zu verstehen ist, d.h. $\nabla = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right)$.

- Sei f auf D partiell differenzierbar. Induktiv seien die folgenden Funktionen für $i_1, \dots, i_k \in \{1, \dots, n\}$ und $k \geq 2$ definiert:

$$\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \dots \partial x_{i_1}} := \frac{\partial}{\partial x_{i_k}} \left(\frac{\partial^{k-1} f}{\partial x_{i_{k-1}} \dots \partial x_{i_1}} \right) \quad (1)$$

Die Funktion f heißt in x k -fach partiell differenzierbar nach $x_{i_1}, \dots, x_{i_k}$, falls der gemäß (1) definierte Grenzwert $\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \dots \partial x_{i_1}}(x)$ existiert.

- Die Funktion f heißt k -fach partiell differenzierbar auf D , falls alle partiellen Ableitungen der Ordnung k gemäß (1) existieren, d.h. für alle möglichen Kombinationen der Indizes $i_1, \dots, i_k \in \{1, \dots, n\}$ existieren die Grenzwerte $\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \dots \partial x_{i_1}}(x)$ für alle $x \in D$. Sind diese partiellen Ableitungen außerdem alle stetig auf D , dann heißt f k -fach stetig partiell differenzierbar auf D .

- Vertauschbarkeitssatz von Schwarz: Sei $f : D \rightarrow \mathbb{R}$ eine zweimal **stetig** (!) partiell differenzierbare Funktion. Dann gilt für alle $x \in D$ und alle $i, j = 1, 2, \dots, n$

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

- Sei $f : D \rightarrow \mathbb{R}$ eine zweimal stetig differenzierbare Funktion. Die symmetrische $n \times n$ -Matrix

$$H_f(x) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix}$$

heißt Hesse-Matrix von f im Punkt $x \in D$.

- Für $i = 1, \dots, m$ seien die Funktionen $f_i : D \rightarrow \mathbb{R}$ partiell differenzierbar. Setze $f(x) := (f_1(x), \dots, f_m(x))^T$. Die $m \times n$ -Matrix

$$J_f(x) := \begin{pmatrix} \nabla f_1(x) \\ \vdots \\ \nabla f_m(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}$$

heißt Jacobimatrix der vektorwertigen Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ im Punkt $x \in D$. Statt Jacobimatrix ist auch der Begriff Funktionalmatrix üblich.

- Die Menge $N_f(x^0) := \{x \in D \mid f(x) = f(x^0)\}$ heißt Niveaumenge. Dies sind in Worten alle $x \in D$, für die die Funktion f konstant ist, d.h. den Wert $f(x^0)$ hat. Im Fall $n = 2$ spricht man auch von Höhenlinien.

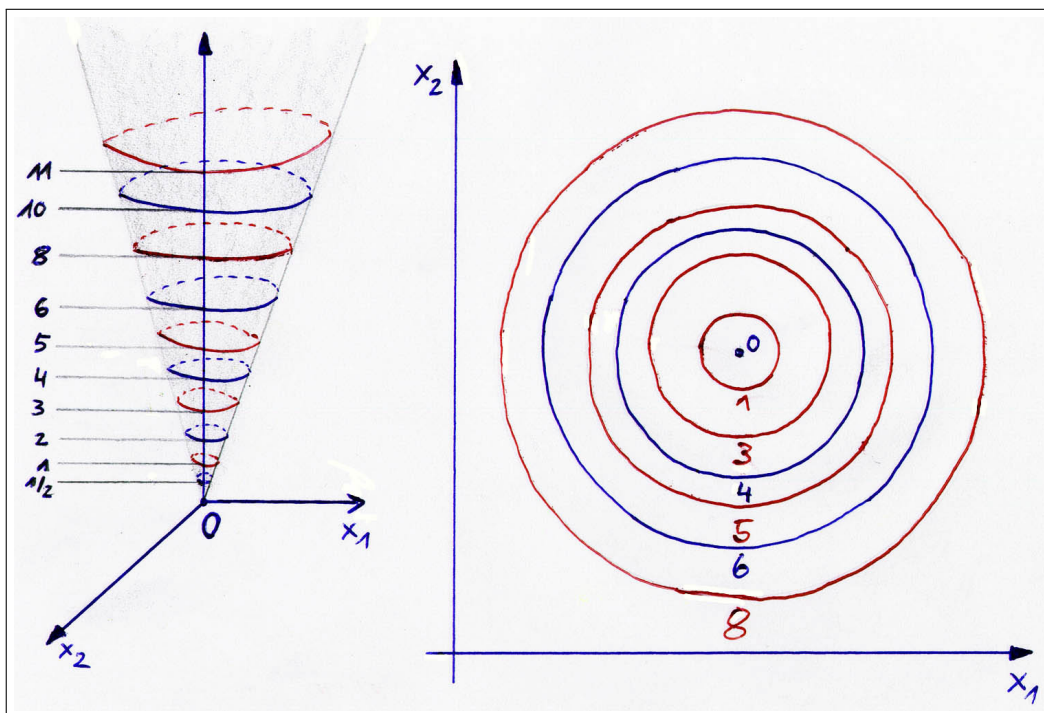


Abb. 2: Schematische Darstellung von Höhenlinien einer Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, hier ein nach oben geöffneter Kreiskegel mit Spitze im Nullpunkt. Links die Funktion selbst, auf der einige konstante Funktionswerte (anschaulich also die „Höhe“ bzw. die Entfernung von Null) durch farbliche Kreise hervorgehoben wurden. Diese Kreise sind anschaulich nichts anderes als die Schnittmenge des Graphen von f mit zur $\mathbb{R}^2$ -Ebene parallelen Ebenen. Die übliche Darstellung von Höhenlinien (rechts) erhält man, indem man die gesamte Funktion bzw. die links dargestellten Schnittmengen auf den $\mathbb{R}^2$ projiziert. Diese Kreise in der $\mathbb{R}^2$ -Ebene sind dann alle Punkte $(x_1, x_2) \in \mathbb{R}^2$, für die f die entsprechenden Werte annimmt.

- Ein Punkt $x \in D$ heißt lokales Maximum [bzw. lokales Minimum] von f , falls eine Umgebung $U \subset D$ von x existiert, so dass $f(x) \geq f(y)$ [bzw. $f(x) \leq f(y)$] für alle $y \in U$ gilt. Tritt der Fall $f(x) = f(y)$ nur für $y = x$ ein, dann spricht man von einem isolierten lokalen Maximum/Minimum.
- Sei $M \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix.
 - M heißt positiv definit, falls $\alpha^T \cdot M \cdot \alpha > 0$ für alle $\alpha \in \mathbb{R}^n \setminus \{0\}$ gilt.
 - M heißt negativ definit, falls $\alpha^T \cdot M \cdot \alpha < 0$ für alle $\alpha \in \mathbb{R}^n \setminus \{0\}$ gilt.
 - M heißt positiv semidefinit, falls $\alpha^T \cdot M \cdot \alpha \geq 0$ für alle $\alpha \in \mathbb{R}^n \setminus \{0\}$ gilt.
 - M heißt negativ semidefinit, falls $\alpha^T \cdot M \cdot \alpha \leq 0$ für alle $\alpha \in \mathbb{R}^n \setminus \{0\}$ gilt.
 - M heißt indefinit, falls es Vektoren $\alpha, \beta \in \mathbb{R}^n \setminus \{0\}$ mit $\alpha^T \cdot M \cdot \alpha > 0$ und $\beta^T \cdot M \cdot \beta < 0$
 - Alternative: Eigenwertberechnung $\rightarrow M$ ist genau dann positiv definit [bzw. negativ definit], wenn alle Eigenwerte von M positiv [bzw. negativ] sind. Entsprechend ist M genau dann positiv semidefinit [bzw. negativ semidefinit], wenn alle Eigenwerte von M größer [bzw. kleiner] gleich Null sind. M ist genau dann indefinit, wenn M einen positiven und einen negativen Eigenwert hat.
 - Alternative: Determinanten-Kriterium für Definitheit $\rightarrow$ Eine reelle symmetrische $n \times n$ -Matrix

$$M := \begin{pmatrix} m_{11} & \cdots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{n1} & \cdots & m_{nn} \end{pmatrix}$$
 ist genau dann positiv definit, wenn

$$\det \begin{pmatrix} m_{11} & \cdots & m_{1k} \\ \vdots & \ddots & \vdots \\ m_{k1} & \cdots & m_{kk} \end{pmatrix} > 0$$
 für alle $k = 1, 2, \dots, n$ gilt. Analoge Kriterien lassen sich für negative Definitheit, positive/negative Semidefinitheit und Indefinitheit einer symmetrischen $n \times n$ -Matrix angeben.

Berechnung lokaler Extremwerte (ohne Nebenbedingungen):

Sei $D \subseteq \mathbb{R}^n$ eine offene Menge ($n \in \mathbb{N} \setminus \{0\}$) und $f : D \rightarrow \mathbb{R}$ eine Funktion:

- notwendige Bedingung: Sei $f : D \rightarrow \mathbb{R}$ einmal stetig partiell differenzierbar. Besitzt f in $x \in D$ ein lokales Extremum, dann gilt: $\nabla f(x) = 0$. (Hinweis: Ein Punkt $x \in D$ mit $\nabla f(x) = 0$ wird auch als kritischer oder stationärer Punkt bezeichnet.)
- hinreichende Bedingung: Sei $f : D \rightarrow \mathbb{R}$ zweimal stetig partiell differenzierbar und $x \in D$ sei ein Punkt mit $\nabla f(x) = 0$.
 - Ist $H_f(x)$ positiv definit, so hat f in x ein isoliertes lokales Minimum.
 - Ist $H_f(x)$ negativ definit, so hat f in x ein isoliertes lokales Maximum.
 - Ist $H_f(x)$ indefinit, so besitzt f in x kein lokales Extremum.

Ist die Hessematrix $H_f(x)$ (positiv oder negativ) semidefinit, so lassen sich keine Aussagen hinsichtlich lokaler Extrema machen. In diesen Fällen sind ggf. weitere Untersuchungen notwendig. Ist $\lambda = 0$ ein Eigenwert der Hessematrix $H_f(x)$, dann heißt der Punkt x ausgeartet. Ist $H_f(x)$ im Fall $n = 2$ indefinit, dann sagt man auch, dass f in x einen Sattelpunkt hat.

- Beispiel: $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ mit $f(x, y) := x^2 + y^2$. Es gilt $\nabla f(x, y) = (2x, 2y) = (0, 0)$ genau dann, wenn $(x, y) = (0, 0)$. Die Hesse-Matrix $H_f(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ ist positiv definit, denn für alle $(x, y) \neq (0, 0)$ gilt $(x, y) \cdot H_f(0, 0) \cdot (x, y)^T = 2x^2 + 2y^2 > 0$. Also hat f im Punkt $(0, 0)$ ein isoliertes lokales Minimum.
- Beispiel: $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ mit $f(x, y) := 2xy + 2x - 2y$. Es gilt $\nabla f(x, y) = (2y + 2, 2x - 2) = (0, 0)$ genau dann, wenn $(x, y) = (1, -1)$. Die Hesse-Matrix $H_f(1, -1) = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$ hat die Eigenwerte $\lambda_1 = 2$ und $\lambda_2 = -2$, d.h. sie ist indefinit. Also hat f im Punkt $(1, -1)$ kein Extremum, sondern lediglich einen Sattelpunkt.

Berechnung lokaler Extremwerte (mit Gleichungsnebenbedingungen):

Sei $D \subseteq \mathbb{R}^n$ eine offene Menge ($n \in \mathbb{N} \setminus \{0\}$) und $f : D \rightarrow \mathbb{R}$ eine Funktion. Gesucht sind die lokalen Extrema von f , wobei allerdings nur solche Punkte $x \in D$ zugelassen sind, die ein Gleichungssystem der Form

$$g_1(x) = 0 \quad , \quad g_2(x) = 0 \quad , \quad \dots \quad , \quad g_m(x) = 0 \quad (2)$$

erfüllen, wobei $g_i : D \rightarrow \mathbb{R}$ für $i = 1, 2, \dots, m$ und $m \geq 1$. Die Extremwerte werden also nicht auf ganz D , sondern auf der kleineren Menge $G := \{x \in D \mid g_1(x) = 0, \dots, g_m(x) = 0\}$ bestimmt. Man spricht in diesem Zusammenhang auch von einem Optimierungsproblem mit Gleichungsrestriktionen, die Menge G wird dabei auch als zulässige Menge bezeichnet.

- Lösungsansatz 1: (Lokale) Auflösung¹ des Gleichungssystems (2) nach m Variablen, zum Beispiel nach $x_{n-m+1}, \dots, x_n$. Setzt man diese in die Funktion f ein, so sind (z.B.) die lokalen Extremwerte der Funktion

$$h(x_1, \dots, x_{n-m}) := f(x_1, \dots, x_{n-m}, x_{n-m+1}, \dots, x_n)$$

zu bestimmen (unrestringiertes Problem). *Aber:* Die Auflösung des Gleichungssystems (2) nach m Variablen ist häufig gar nicht möglich und/oder nicht eindeutig!

- Beispiel: Gesucht ist das Minimum der Funktion $f(x, y) := x^2 + y^2$ unter der Nebenbedingung $g_1(x, y) := y - x^2 - a = 0$, wobei $a < -\frac{1}{2}$ gelte. Die Nebenbedingung $y - x^2 - a = 0$ können wir nach x auflösen, also $x = \pm\sqrt{y-a}$. Dies eingesetzt in die Funktion $f(x, y)$ liefert als „neue“ zu minimierende Funktion $h(y) := f(x, y) = y - a + y^2$. Dies ist ein Problem, das nur noch von einer Variablen abhängt. Die Funktion h hat in y ein Extremum, wenn $h'(y) = 2y + 1 = 0$ gilt, also wenn $y = -\frac{1}{2}$ gilt. Weiter ist $h''(y) = 2 > 0$ für alle $y \in \mathbb{R}$, also hat h in $y = -\frac{1}{2}$ ein lokales Minimum. Es ergibt sich $x = \pm\sqrt{-\frac{1}{2}-a}$. Daraus folgt, dass die Punkte $(\sqrt{-\frac{1}{2}-a}, -\frac{1}{2})$ und $(-\sqrt{-\frac{1}{2}-a}, -\frac{1}{2})$ die Nebenbedingung erfüllen und somit die Funktion $f(x, y)$ minimieren (unter der Nebenbedingung $g_1(x, y) = 0$).
- Lösungsansatz 2: Ankoppeln der Gleichungen $g_i(x) = 0$ für $i = 1, \dots, m$ mit unbekannten *Lagrange-Multiplikatoren* $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ an die Funktion f . Dazu wird die sog. *Lagrange-Funktion*

$$F(x) := f(x) + \sum_{i=1}^m \lambda_i g_i(x) \quad (3)$$

betrachtet und deren lokale Extrema auf D für feste $\lambda_1, \dots, \lambda_m$ bestimmt (unrestringiertes Problem).

- Minimiert [bzw. maximiert] $x \in D$ die Lagrange-Funktion F für feste $\lambda_1, \dots, \lambda_m$ und löst x das Gleichungssystem (2), dann ist x zugleich Minimum [bzw. Maximum] von f über G .
- Seien $f, g_1, \dots, g_m$ stetig partiell differenzierbar, $g(x) := (g_1(x), \dots, g_m(x))^T$ und $x^0 \in D$ sei ein lokales Extremum von f unter der Nebenbedingung $g(x) = 0$. Ferner gelte die Regularitätsbedingung $\text{rang}(J_g(x^0)) = m$. Dann existieren Lagrange-Multiplikatoren $\lambda_1, \dots, \lambda_m \in \mathbb{R}$, so dass für die Lagrange-Funktion F gemäß (3) die folgende notwendige Bedingung gilt: $\nabla F(x^0) = 0$.
- Sind $f, g_1, \dots, g_m$ stetig partiell differenzierbar, so erhält man als notwendige Bedingung für eine Extremstelle $\nabla F(x) = \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) = 0$. Dies ist ein (allgemein nichtlineares) Gleichungssystem mit $n+m$ Gleichungen und den $n+m$ Variablen $x_1, \dots, x_n, \lambda_1, \dots, \lambda_m$, dessen Lösungen eben Kandidaten für die gesuchten Extremalstellen sind.
- Sind $f, g_1, \dots, g_m$ zweimal partiell differenzierbar, $x \in G$ ein Punkt mit $\nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) = 0$ und ist die Hesse-Matrix $H_F(x)$ positiv positiv [bzw. negativ] definit, so hat F und damit f in x ein lokales Minimum [bzw. Maximum]. Diese hinreichende Bedingung ist aber in vielen Anwendungen eine zu strenge Forderung, d.h. die Bedingung ist nicht erfüllt, aber F bzw. f hat in x trotzdem eine

¹Für mehr Details sei hier auf die Theorie der impliziten Funktionen verwiesen, die aber nicht Gegenstand dieser Vorlesung ist.

lokale Extremstelle. Insbesondere kann aus der Indefinitheit von $H_F(x)$ nicht geschlossen werden, dass x keine Extremstelle ist.²

- Beispiel: Gesucht ist das Minimum der Funktion $f(x, y) := x^2 + y^2$ unter der Nebenbedingung $g_1(x, y) := y - x^2 - a = 0$, wobei $a < -\frac{1}{2}$ gelte (vgl. oben). Die zugehörige Lagrange-Funktion ist $F(x, y) = f(x, y) + \lambda g_1(x, y) = x^2 + y^2 + \lambda(y - x^2 - a)$. Die notwendige Bedingung $\nabla F(x) = 0$ ist äquivalent zur Lösung des folgenden Gleichungssystems:

$$\frac{\partial F}{\partial x}(x, y, \lambda) = 2x - 2\lambda x = 0 \quad (\text{I})$$

$$\frac{\partial F}{\partial y}(x, y, \lambda) = 2y + \lambda = 0 \quad (\text{II})$$

$$\frac{\partial F}{\partial \lambda}(x, y, \lambda) = y - x^2 - a = 0 \quad (\text{III})$$

Aus Gleichung (I) folgt sofort $\lambda = 1$. Damit ergibt sich aus (II) dann $y = -\frac{1}{2}$. Eingesetzt in (III) ergibt sich $x = \pm\sqrt{-\frac{1}{2} - a}$. Daraus folgt (wie nicht anders zu erwarten, siehe oben), dass die Punkte $\left(\sqrt{-\frac{1}{2} - a}, -\frac{1}{2}\right)$ und $\left(-\sqrt{-\frac{1}{2} - a}, -\frac{1}{2}\right)$ die Nebenbedingung erfüllen und somit die Funktion $f(x, y)$ minimieren (unter der Nebenbedingung $g_1(x, y) = 0$).

Ergänzung: Eigenschaften des Gradienten, Tangentialebenen

- Seien $D \subseteq \mathbb{R}^n$ eine offene Menge, $x \in D$, $f : D \rightarrow \mathbb{R}$ sowie $v \in \mathbb{R}^n$ ein Vektor mit $\|v\| = 1$. Unter der Richtungsableitung von f im Punkt x in Richtung v versteht man (im Fall der Existenz) den Grenzwert

$$D_v f(x) := \left. \frac{d}{dt} f(x + tv) \right|_{t=0} = \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t}.$$

Speziell für $v = e_i$ erhält man $D_{e_i} f(x) = \frac{\partial f}{\partial x_i}(x)$.

- Sei $D \subseteq \mathbb{R}^n$ eine offene Menge und $f : D \rightarrow \mathbb{R}$. Für alle $x \in D$ und für alle $v \in \mathbb{R}^n$ mit $\|v\| = 1$ gilt

$$D_v f(x) = \langle v, \nabla f(x) \rangle = \sum_{i=1}^n v_i \cdot \frac{\partial f}{\partial x_i}(x).$$

- Ist $\nabla f(x) \neq 0$, dann ist zwischen den Vektoren v und $\nabla f(x)$ ein Winkel ϕ definiert und es gilt $D_v f(x) = \langle v, \nabla f(x) \rangle = \|\nabla f(x)\| \cdot \cos(\phi)$. Daraus folgt, dass $D_v f(x)$ maximal wird, falls v und $\nabla f(x)$ die gleiche Richtung haben. Dies bedeutet: $\nabla f(x)$ gibt die Richtung des steilsten Anstiegs von f im Punkt x an, entsprechend gibt $-\nabla f(x)$ die Richtung des steilsten Abstiegs von f im Punkt x an.³
- Ist $x^0 \in D$, dann steht der Gradient $\nabla f(x^0)$ senkrecht auf der Niveaumenge $N_f(x^0)$. Verallgemeinerung: Liegt der Punkt $a \in D$ auf der Niveaumenge $N_f(x^0)$, dann steht $\nabla f(a)$ ebenfalls senkrecht auf $N_f(x^0)$, d.h. es gilt $\langle v, \nabla f(a) \rangle = 0$, wobei v in $t = 0$ Tangentialvektor einer Kurve $\gamma(t)$ aus $N_f(x^0)$ mit $\gamma(0) = a$ ist (zur Veranschaulichung siehe Abbildung 3).
- Sei $D \subseteq \mathbb{R}^n$ offen und $f : D \rightarrow \mathbb{R}$ eine partiell differenzierbare Funktion. Seien $x^0 \in D$ und $a \in N_f(x^0)$ mit $\nabla f(a) \neq 0$ gegeben. Durch die Gleichung

$$\nabla f(a) \cdot \begin{pmatrix} x_1 - a_1 \\ \vdots \\ x_n - a_n \end{pmatrix} = 0$$

wird die Tangentialebene im Punkt a an die Niveaumenge $N_f(x^0)$ definiert.

²Auf schwächere hinreichende Bedingungen bzw. Alternativen wollen und können wir hier nicht eingehen, da dies den zeitlichen sowie inhaltlichen Rahmen dieser Vorlesung sprengen würde. Dazu sei auf die angegebene Literatur verwiesen.

³Diese Überlegungen sind in der Optimierung bei Gradientenverfahren bzw. Verfahren des steilsten Abstiegs wichtig.

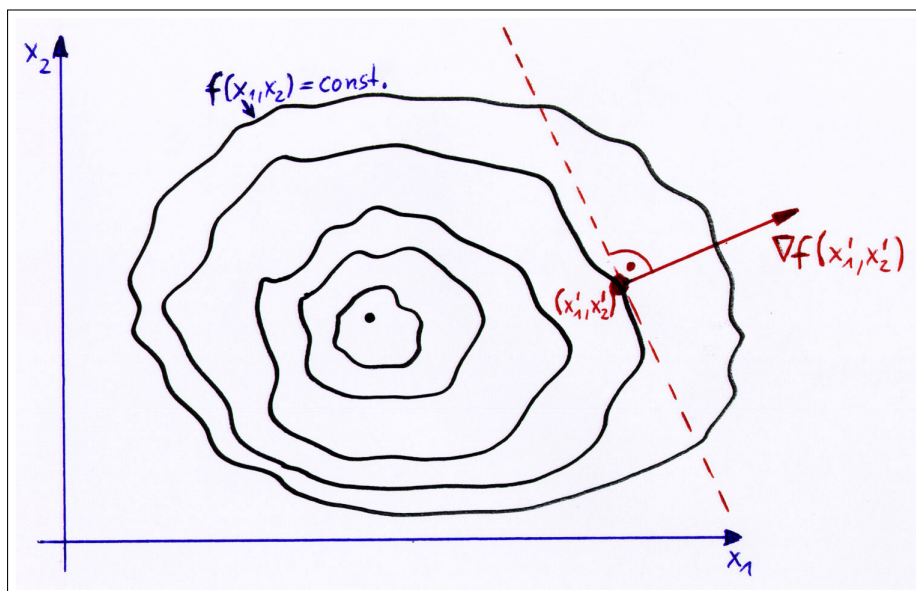


Abb. 3: Die Höhenlinien einer Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ und der Gradient in einem Punkt (x'_1, x'_2) .

- Beispiel: Die Richtungsableitung von $f(x, y) := x^3y + xy^3$ in Richtung des Einheitsvektors $v = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ im Punkt $(x, y) = (1, 1)$ ist $D_v f(1, 1) = \langle v, \nabla f(1, 1) \rangle = \left\langle \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), (4, 4) \right\rangle = \frac{8}{\sqrt{2}} = 4\sqrt{2}$, wobei $\nabla f(x, y) = (3x^2y + y^3, x^3 + 3xy^2)$.
- Beispiel: Gesucht ist die maximale Steigung der Funktion $f(x, y) = x^2 - y^2$ im Punkt $(x, y) = (1, 1)$. Es gilt $\nabla f(x, y) = (2x, -2y)$, also steigt f im Punkt $(1, 1)$ in der Richtung $\nabla f(1, 1) = (2, -2)$ am schnellsten. Setze $v := \frac{\nabla f(1, 1)}{\|\nabla f(1, 1)\|} = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$. Die maximale Steigung von f im Punkt $(1, 1)$ ist damit $D_v f(1, 1) = \langle v, \nabla f(1, 1) \rangle = \left\langle \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right), (2, -2) \right\rangle = 2\sqrt{2}$.
- Beispiel: Sei $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ eine partiell differenzierbare Funktion. Setzt man $F(x, y, z) := f(x, y) - z$, dann ist der Graph von f die Niveaumenge $N := \{(x, y, z) \mid F(x, y, z) = F(x, y, f(x, y)) = 0\}$. Es gilt $\nabla F(x, y, z) = \left(\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y), -1\right)$. Als Gleichung der Tangentialebene an den Graphen von f im Punkt (a, b) [bzw. an F im Punkt $(a, b, f(a, b))$] erhält man damit

$$\nabla F(a) \cdot \begin{pmatrix} x - a \\ y - b \\ z - f(a, b) \end{pmatrix} = \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) - (z - f(a, b)) = 0,$$

also

$$z(x, y) = f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b)$$

- Beispiel: Gesucht sei die Gleichung der Tangentialebene an die Funktion $f(x, y, z) = 3xy + z^2$ im Punkt $(x, y, z) = (1, 1, 1)$. Es gilt $\nabla f(x, y, z) = (3y, 3x, 2z)$, also $\nabla f(1, 1, 1) = (3, 3, 2)$. Daraus folgt

$$\nabla f(1, 1, 1) \cdot \begin{pmatrix} x - 1 \\ y - 1 \\ z - 1 \end{pmatrix} = 3x - 3 + 3y - 3 + 2z - 2 = 0,$$

also $3x + 3y + 2z = 8$.

Literatur:

Alle hier gegebenen Definitionen sowie insbesondere auch die Beweise der angegebenen Aussagen sind in jedem guten Analysis-Lehrbuch zu finden. Im Rahmen dieser Vorlesung ist dabei zum Beispiel

- Oberguggenberger, M.; Ostermann, A.: *Analysis für Informatiker*, Springer, Berlin/Heidelberg 2009
- Ansorge, R.; Oberle, H.-J.: *Mathematik für Ingenieure, Band 2*, Akademie Verlag, Berlin 1994

zu empfehlen. Darüber hinausgehende Informationen erhält man zum Beispiel in

- Forster, O.: *Analysis 2 - Differentialrechnung im $\mathbb{R}^n$. Gewöhnliche Differentialgleichungen*, Vieweg, Wiesbaden 1999
- Heuser, H.: *Lehrbuch der Analysis - Teil 2*, Teubner, Stuttgart 1995
- Walter, W.: *Analysis 2*, Springer, Berlin/Heidelberg 2002