

Praktische Optimierung

Wintersemester 2008/09

Prof. Dr. Günter Rudolph

Lehrstuhl für Algorithm Engineering

Fakultät für Informatik

TU Dortmund

Grundlagen von Optimierproblemen

Am Montag gelernt:

- Formalitäten
- Überblick über dieses Semester
- Def.: Optimierproblem
- Optimum, Optimalstelle
- lokales, globales Optimum
- Minimierung wie Maximierung
- konvexe Funktionen

Grundlagen klassischer Optimiertechniken

Themen der heutigen Vorlesung:

- Exkurs:
Differentialrechnung
- Optimalitätsbedingungen für unrestringierte Optimierprobleme
- Analyse von quadratischen, konvexen Funktionen
- Gradientenverfahren

Exkurs: Differentialrechnung

Grundlagen der Differentialrechnung

$f : \mathbb{R} \rightarrow \mathbb{R}$ differenzierbar in $x \iff$
 $\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$ existiert und ist endlich

Der Limes wird mit $f'(x)$ bezeichnet und heißt **Ableitung** von f an der Stelle x

Einige grundlegende Ableitungsregeln

Sei $f : \mathbb{R} \rightarrow \mathbb{R}$, $g : \mathbb{R} \rightarrow \mathbb{R}$, $x, \alpha, n \in \mathbb{R}$

$$(\alpha)' = 0$$

Konstante

$$(\alpha f)' = \alpha f'$$

Faktorregel

$$(f + g)' = f' + g'$$

Summenregel

$$(fg)' = gf' + fg'$$

Produktregel

$$\left(\frac{f}{g}\right)' = \frac{gf' - fg'}{g^2}$$

Quotientenregel

$$(f \circ g)'(x) = (f(g(x)))' = g'(x)f'(g(x))$$

Kettenregel

$$(x^n)' = nx^{n-1}$$

Potenzregel

$$(\ln x)' = \frac{1}{x}$$

logarithmische Ableitung

Partielle Funktion und Ableitung

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

Argument $x = (x_1, \dots, x_n)$ ist ein n -dimensionaler **Vektor**

$\hat{}$ zeigt an, dass eine Variable (Veränderliche) als konstant angenommen wird

Sei $\hat{x} \in \mathbb{R}^n$ und $f(\hat{x}_1, \dots, \hat{x}_{k-1}, x_k, \hat{x}_{k+1}, \dots, \hat{x}_n)$ eine partielle Funktion mit existierender Ableitung in x_k .

Partielle Ableitung nach x_i :

$$\frac{\partial f(x_1, \dots, x_n)}{\partial x_i} = \frac{\partial f}{\partial x_i}(x_1, \dots, x_n) = \frac{\partial f}{\partial x_i} = f_{x_i}$$

Handwerkliches:

Alle Veränderliche außer x_i als **Konstanten** auffassen und die nunmehr nur noch von x_i abhängende Funktion wie „gewohnt“ ableiten.

Beispiel: Einfache partielle Ableitungen

$$f(x, y, z) = x^2 + xy^2 + 2z^3$$

Partielle Ableitungen:

$$f_x(x, y, z) = 2x + y^2$$

$$f_y(x, y, z) = 2xy$$

$$f_z(x, y, z) = 6z^2$$

$$f(x, y) = y \log(2xy), \quad x, y > 0$$

Partielle Ableitungen:

$$f_x(x, y) = y \cdot 2y \cdot \frac{1}{2xy} = \frac{y}{x}$$

$$f_y(x, y) = \log(2xy) + y \cdot 2x \cdot \frac{1}{2xy} = \log(2xy) + 1$$

$$\text{Regel: } \log(g(x))' = \frac{g'(x)}{g(x)}$$

Partielle Ableitung zweiter Ordnung

Ordnung: Anzahl abzuleitender Variablen

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \dots, x_n) = f_{x_i x_j}$$

(höhere Ordnungen analog)

Prinzip: $\frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_j} \right)$

Handwerkliches:

erst nach x_j ableiten, dann Resultat nach x_i ableiten

Reihenfolge egal, falls $f \in C^m(G)$ (f m -fach stetig differenzierbar)

Satz von Schwarz:

Für jedes $f \in C^m(G)$ mit $\emptyset \neq G \subseteq \mathbb{R}^n$ offen ist die Reihenfolge der partiellen Ableitungen bis zur m -ten Ordnung irrelevant.

Beispiel: Partielle Ableitungen zweiter Ordnung

$$f(x, y) = x^3 - 2x^2y^2 + 4xy^3 + y^4 + 10$$

Partielle Ableitungen

$$f_x = 3x^2 - 4xy^2 + 4y^3 \qquad f_y = -4x^2y + 12xy^2 + 4y^3 \qquad \text{1. Ordnung}$$

$$f_{xx} = 6x - 4y^2 \qquad f_{yx} = -8xy + 12y^2 \qquad \text{2. Ordnung}$$

$$f_{xy} = -8xy + 12y^2 \qquad f_{yy} = -4x^2 + 24xy + 12y^2$$

$$f_{xy} = f_{yx} \quad \checkmark$$

Gradient

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

Gradient $\nabla f(x)$: Vektor der partiellen Ableitungen von $f(x)$.

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

(∇ heißt Nabla-Operator, ' steht für transponiert)

Hessematrix $\nabla^2 f(x)$: Matrix der partiellen zweiten Ableitungen von $f(x)$.

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{pmatrix}$$

Falls $f \in C^2$

Satz von Schwarz
 \Rightarrow

Hessematrix symmetrisch

Taylor

Gemäß **Taylor-Formel** nähert man eine Funktion durch eine **Taylor-Reihe** bestehend aus **Taylor-Polynomen** an

$$f(x + h) = f(x) + (\nabla f(x))'h + \frac{1}{2}h'\nabla^2 f(x)h + \|h\|^2\rho(h)$$

mit $\lim_{h \rightarrow 0} \rho(h) = 0$

Abweichung kann beliebig klein werden

Andere Darstellung mit $h = x - x_0$

$$f(x) \approx f(x_0) + (\nabla f(x_0))'(x - x_0) + \frac{1}{2}(x - x_0)'\nabla^2 f(x_0)(x - x_0)$$

Approx. durch Gerade Approx. durch Paraboloid

Ende Exkurs Differentialrechnung

Optimalitätsbedingungen für unrestringierte Probleme

$f \in C^2$, $\exists x^* \in \mathbb{R}^n$ mit $\nabla f(x^*) = 0$ und $\nabla^2 f(x^*)$ positiv definit

$\Rightarrow f$ hat in x^* ein lokales Minimum

negativ definit \Rightarrow lokales Maximum, indefinit \Rightarrow kein lokales Optimum

Definitheitskriterium für $(n \times n)$ -Matrix A :

Sei $\Delta_k := \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{vmatrix}$ k -te Abschnittsdeterminante

(von Hauptminor: quadr. Untermatrix beginnend links oben)

A pos. def. $\iff \forall k : \Delta_k > 0$

A neg. def. $\iff \forall k : (-1)^k \Delta_k > 0$

Beispiel: $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$

$a > 0$ und $ac - b^2 > 0 \Rightarrow$ pos. def.

$a < 0$ und $ac - b^2 > 0 \Rightarrow$ neg. def.

$a < 0$ und $ac - b^2 < 0 \Rightarrow$ indef.

Beispiel: „Methodik bestimmt lokales Optium“

$$f(x, y) = x^3 + y^3 - 3xy$$

$$f_x = 3x^2 - 3y$$

$$f_y = 3y^2 - 3x$$

partielle Ableitungen bestimmen

Nullstellen der partiellen Ableitungen suchen

⇒ Gleichungssystem mit zwei Lösungen: (0, 0) und (1, 1)

$$\nabla^2 f(x, y) = \begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix}$$

Hessematrix berechnen

Eigenschaften der Hessematrix an Lösungsstellen

$$\nabla^2 f(0, 0) = \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix} \quad \text{indefinit} \Rightarrow (0, 0) \text{ keine Optimalstelle}$$

$$\nabla^2 f(1, 1) = \begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix} \quad \begin{cases} \Delta_1 = 6 > 0 \text{ und} \\ \Delta_2 = 6 \cdot 6 - (-3)(-3) = 27 > 0 \end{cases} \Rightarrow \text{pos.def.}$$

⇒ lokales Optimum: $f(1, 1) = -1$

globales Optimum?

$$f(x, x) \rightarrow +\infty \text{ f\"ur } x \rightarrow +\infty \quad \text{und} \quad f(x, x) \rightarrow -\infty \text{ f\"ur } x \rightarrow -\infty$$

⇒ f unbeschränkt ⇒ lokales Opt. ist kein globales Opt.

Beispiel: „Methodik hat erste Schwierigkeiten“

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

$$f_x = 2x - 2y$$

$$f_y = 2y - 2x$$

partielle Ableitungen bestimmen

Nullstellen der partiellen Ableitungen suchen

⇒ alle (x, y) mit $x = y$

Hessematrix berechnen

$$\nabla^2 f(x, y) = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} \quad \begin{cases} \Delta_1 > 0 \\ \Delta_2 = 0 \end{cases}$$

so keine Entscheidung möglich

Analyse durch Nachdenken und Abschätzung der Funktion:

$$f(x, y) = x^2 + y^2 - 2xy + 1 = \underbrace{(x - y)^2}_{\geq 0} + 1 \geq 1$$

Lösung raten:

$$\forall x : f(x, x) = 1$$

⇒ alle Elemente in $\{(x, y) \in \mathbb{R}^2 : x = y\}$ sind lokale und globale Minimalstellen

Beispiel: „Methodik versagt“

$$f(x, y) = \sqrt{x^2 + y^2}$$

offensichtlich: globales Minimum in $(0, 0)$

Analyse durch partielle Ableitung:

$$f_x = 2x \cdot \frac{1}{2\sqrt{x^2+y^2}}$$

$$f_y = 2y \cdot \frac{1}{2\sqrt{x^2+y^2}}$$

suche Nullstelle der partiellen Ableitung:

$$f_x(0, y) = 0 \text{ f\"ur } y \neq 0 \text{ und}$$

$$f_y(x, 0) = 0 \text{ f\"ur } x \neq 0,$$

aber $f_x(0, 0)$ und $f_y(0, 0)$ undefiniert

⇒ es existieren keine partiellen Ableitungen in $(0, 0)$

Optimum einer quadratischen, konvexen Funktion (1)

$$f(x) = \frac{1}{2}x'Ax + b'x + c$$

(Erinnerung: ' für transponiert)

$$\nabla f(x) = Ax + b \stackrel{!}{=} 0$$

suche Optimalstellen

$$\nabla^2 f(x) = A$$

A pos. def., da f konvex \Rightarrow lokales Opt. = globales Opt.

Umformung des Gradienten

$$Ax = -b \quad |A^{-1} \text{ linksseitig}$$

$$A^{-1}Ax = -A^{-1}b$$

$$x = -A^{-1}b \quad =: x^* \text{ Optimalstelle eines globalen Optimums}$$

Inverse Matrix A^{-1} existiert immer:

$$A \text{ pos. def.} \Rightarrow \det A > 0$$

$$\det A \neq 0 \iff A^{-1} \text{ existiert}$$

Optimum einer quadratischen, konvexen Funktion (2)

Umwandlung von f durch ausmultiplizieren

$$\begin{aligned}f(x) &= \frac{1}{2}x'Ax + b'x + c \\&= \frac{1}{2}(x_1, x_2) \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (b_1, b_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + c && \text{A symmetrisch} \\&= \frac{1}{2}(x_1 a_{11} + x_2 a_{12}, x_1 a_{12} + x_2 a_{22}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + b_1 x_1 + b_2 x_2 + c \\&= \frac{1}{2}(a_{11}x_1^2 + 2a_{12}x_1x_2 + x_2 a_{22}) + b_1 x_1 + b_2 x_2 + c\end{aligned}$$

Gradient

$$\begin{aligned}\nabla f(x) &= \begin{pmatrix} f_{x_1} \\ f_{x_2} \end{pmatrix} \\&= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + b_1 \\ a_{22}x_1 + a_{12}x_1 + b_2 \end{pmatrix} \\&= Ax + b\end{aligned}$$

Optimum einer quadratischen, konvexen Funktion (3)

Berechne globales Optimum an Optimalstelle x^*

$$\begin{aligned} f(x^*) &= f(-A^{-1}b) \\ &= \frac{1}{2}(-A^{-1}b)'AA^{-1}b - b'A^{-1}b + c \\ &= \frac{1}{2}b'A^{-1}b - b'A^{-1}b + c \\ &= -\frac{1}{2}b'A^{-1}b + c \end{aligned}$$

Gradientenverfahren

$\nabla f(x)$ zeigt in Richtung des stärksten Anstiegs an der Stelle x
 $\Rightarrow -\nabla f(x)$ stärkster Abstieg

Idee:

iterative Annäherung an lokales Optimum durch „Verfolgen“ des neg. Gradienten

Schrittweite $s \in \mathbb{R}^+$, Hochzahl bezeichnet Iterationsschritt

$$\begin{aligned}x^{(k+1)} &= x^{(k)} - s^{(k)} \frac{\nabla f(x^{(k)})}{\|\nabla f(x^{(k)})\|} && \text{Vektorlänge normiert} \\ &= x^{(k)} - s^{(k)} \cdot d^{(k)}\end{aligned}$$

optimale Schrittweite:

wähle $s^{(k)}$ so, dass $f(x^{(k)} - s^{(k)} d^{(k)}) = \min\{f(x^{(k)} - s^{(k)} d^{(k)}) : s \in \mathbb{R}^+\}$

Wie macht man das?

Fortsetzung am Montag...

Zusammenfassung

Heute (mindestens) gelernt:

- Grundlagen der Differenzialrechnung ...
sind immer noch wie in der Schule
- Partielle Ableitungen ...
betrachten nicht abzuleitende Variablen als Konstanten
existieren nicht immer
sind oft **nicht** zum Auffinden einer Optimalstelle geeignet
- Konvexe, quadratische Funktion ...
sind harmlos
haben als lokale Optima nur das globale Optimum
lassen sich mit Hilfe partieller Ableitungen lösen
- Gradientenverfahren ...
sind Optimierverfahren
arbeiten iterativ
benötigen eine passend gewählte Schrittweite

Literatur zu Analysis

- H. Heuser: Lehrbuch der Analysis (Teil 1)
3. Auflage, Teubner, Stuttgart, 1984 (Kapitel 6)
- H. Heuser: Lehrbuch der Analysis (Teil 2)
9. Auflage, Teubner, Stuttgart, 1995 (Kapitel 20)
- (ganz viele andere)