CSE203B Convex Optimization:
Chapter 9: Unconstrained Minimization

CK Cheng
Dept. of Computer Science and Engineering
University of California, San Diego

Chapter 9 Unconstrained Minimization

- Introduction
- Taylor's Expansion & Bounds
- Descent Methods
- Newton Method
- Summary

Introduction

Problem: $\min f(x)$ where $f: R^n \to R$ is convex and twice continuously differentiable

Theorem: Necessary and sufficient condition for a point x^* to be optimal is $\nabla f(x^*) = 0$.

Remark: keywords Taylor's expansion

Taylor's Expansion & Bounds: Scalar case

$f(x) = f(x_0) + \nabla f(x_0)^T(x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(z)(x - x_0)$

for some z on the segment $[x, x_0]$

(1)Scalar case: $f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(z)(x - x_0)^2$

We simplify the notations $f(x) = \frac{m}{2}(x - x_0)^2 + a(x - x_0) + b$

For fixed $m, a, and\ b$, the optimal solution can be derived as:

$$\nabla f(x) = 0 \Rightarrow m(x - x_0) + a = 0 \Rightarrow x - x_0 = -\frac{a}{m}$$

Thus, we have

$$f(x) = \frac{m}{2}\frac{a^2}{m^2} + a\frac{(-a)}{m} + b = \frac{a^2}{2m} - \frac{a^2}{m} + b = \frac{-a^2}{2m} + b$$

Or $f(x) - f(x_0) = -\frac{a^2}{2m}$

a. How far from opt. x^*? $x^* - x_0 = -\frac{a}{m}$

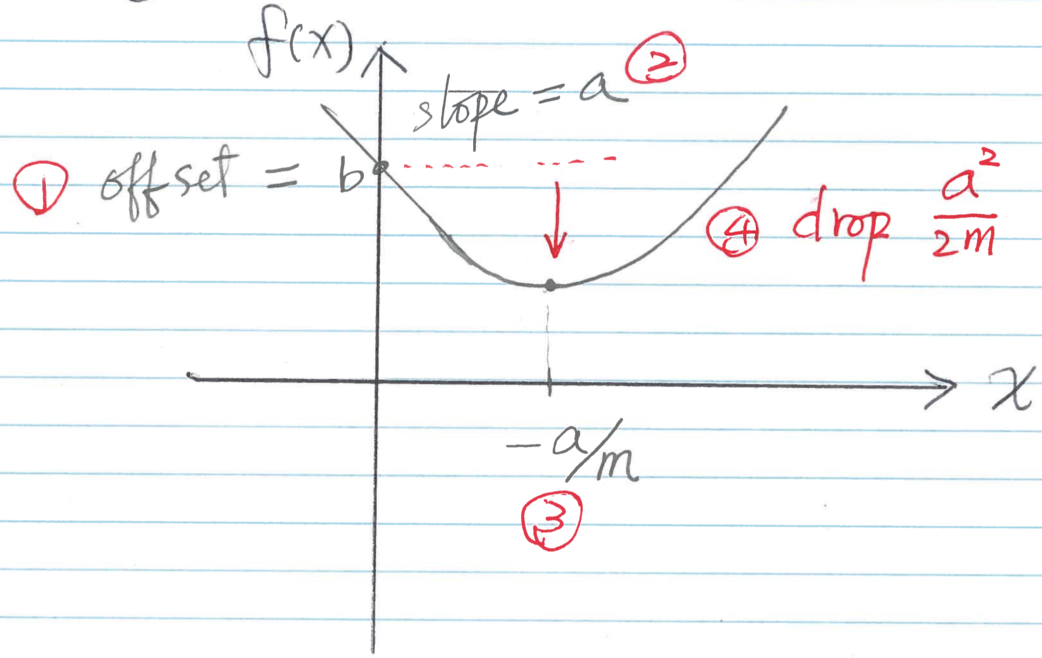
b. How much difference from opt. $f(x^*)$? $f(x_0) - f(x^*) = \frac{a^2}{2m}$

m is positive

$$f(x) = \frac{m}{2}x^2 + ax + b$$

$$= \frac{m}{2}\left(x + \frac{a}{m}\right)^2 - \frac{a^2}{2m} + b$$

(0) m, (1) b, (2) a, (3) a/m, (4) $\frac{a^2}{2m}$



A4.

Taylor's Expansion & Bounds: Example

$$f(x) = x^2 + 4x + 1$$

For the format

$$f(x) = \frac{m}{2}x^2 + ax + b, \qquad m = 2, a = 4, b = 1.$$

Let $x_0 = 0$, we have the answer.

a. *How far?* $x^* - x_0 = -\frac{a}{m} = -2$

b. *How much?* $f(x_0) - f(x^*) = \frac{a^2}{2m} = 4$

Taylor's Expansion & Bounds: Bounds

(2) Vector case:

Assumption A: $\nabla^2 f(x)$ is bounded, *i.e.* $mI \preccurlyeq \nabla^2 f(x) \preccurlyeq MI$

$MI - \nabla^2 f(x) \geq 0$

Theorem A*:* We have the following bounds

$$\frac{1}{M}\|\nabla f(x_0)\|_2 \overset{④}{\leq} \|x_0 - x^*\|_2 \overset{①}{\leq} \frac{2}{m}\|\nabla f(x_0)\|_2$$

$$\frac{1}{2M}\|\nabla f(x_0)\|_2^2 \overset{③}{\leq} f(x_0) - p^* \overset{②}{\leq} \frac{1}{2m}\|\nabla f(x_0)\|_2^2$$

Proof: ①

$$f(x) + \nabla f(x)^T(y - x) + \frac{m}{2}\|y - x\|_2^2 \leq f(y)$$

$$\leq f(x) + \nabla f(x)^T(y - x) + \frac{M}{2}\|y - x\|_2^2$$

(Taylor's Expansion + Assumption A)

Taylor's Expansion & Bounds: Bounds

***Proof* ①:** $\|x - x^*\|_2 \leq \frac{2}{m}\|\nabla f(x)\|_2$

$p^* = f(x^*) \geq f(x) + \nabla f(x)^T(x^* - x) + \frac{m}{2}\|x^* - x\|_2^2$ *(Taylor's exp + Assumption A.)*

$$\geq f(x) - \|\nabla f(x)\|_2\|x^* - x\|_2 + \frac{m}{2}\|x^* - x\|_2^2$$

We shift f(x) to the left hand side.

$0 \geq p^* - f(x) \geq -\|\nabla f(x)\|_2\|x^* - x\|_2 + \frac{m}{2}\|x^* - x\|_2^2$

Shift $-\|\nabla f(x)\|_2\|x^* - x\|_2$ *to the left,*

$\|\nabla f(x)\|_2\|x^* - x\|_2 \geq \frac{m}{2}\|x^* - x\|_2^2$

Therefore we have

1. $\|\nabla f(x)\|_2 \geq \frac{m}{2}\|x^* - x\|_2$

2. $\|x^* - x\|_2 \leq \frac{2}{m}\|\nabla f(x)\|_2$

Taylor's Expansion & Bounds: Bounds

***Proof* ②:** $f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{m}{2}\|y - x\|_2^2$

(Taylor's exp + assumption A.)

$\geq f(x) - \frac{1}{2m}\|\nabla f(x)\|_2^2$ *(Minimization with y)*

Thus, we have

$$f(x) - f(y) \leq \frac{1}{2m}\|\nabla f(x)\|_2^2, \qquad \forall y$$

Therefore

$$f(x) - f(x^*) \leq \frac{1}{2m}\|\nabla f(x)\|_2^2$$

Taylor's Expansion & Bounds

Remark:

(1) If $\|\nabla f(x)\|_2 \leq (2m\epsilon)^{\frac{1}{2}}$

We have $\|x - x^*\|_2 \leq \frac{2}{m}(2m\epsilon)^{\frac{1}{2}}$

$$f(x) - f(x^*) \leq \frac{\|\nabla f(x)\|_2^2}{2m} = \epsilon$$

(2) The bounds can be used to design algorithms.

prove the convergence.

(3) If $M \gg m$ $(e.g. 10^{10})$

Impact on the bounds become very loose

→ Efficiency of gradient descent approaches.

(4) Quadratic obj. with sparse matrix (A)

$\frac{1}{2}x^T Ax + b^T x + c$

is a preferred formulation in terms of algorithm efficiency.

II. Descent Methods

Given convex function, twice continuously differentiable $f(x)$ and an initial point $x_o \in dom\, f$.

Repeat

1. Determine a descent direction Δx $(\nabla f(x)^T \Delta x < 0)$
2. Line Search, choose a step size $t > 0$.
3. Update $x = x + t\Delta x$

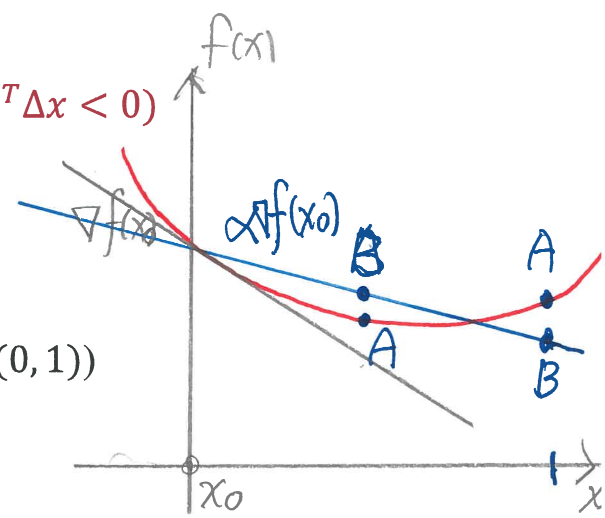
Until stopping criterion is met.

Line Search : $t = arg \min_{t>0} f(x + t\Delta x)$

Backtracking line search $(\alpha \in (0, 1/2), \beta \in (0, 1))$

Start at $t = 1$, repeat $t := \beta t$

until $f(x + t\Delta x) < f(x) + \alpha t \nabla f(x)^T \Delta x$



Stopping criterion $\|\nabla f(x)\|_2 \leq \eta$ $\eta = (2m\epsilon)^{\frac{1}{2}}$ (Theorem A (2))

Taylor's Expansion & Bounds: Bounds

***Proof* ③:** $f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{m}{2}\|y - x\|_2^2$

(Taylor's exp + assumption A.)

$\geq f(x) - \frac{1}{2m}\|\nabla f(x)\|_2^2$ *(Minimization with y)*

Let $y = x - \frac{1}{M}\nabla f(x)$, *we have*

$$f\left(x - \frac{1}{M}\nabla f(x)\right) \leq f(x) + \nabla f(x)^T \frac{-1}{M}\nabla f(x) + \frac{M}{2}\left\|\frac{1}{M}\nabla f(x)\right\|_2^2$$

$$= f(x) - \frac{1}{2M}\|\nabla f(x)\|_2^2$$

Shift the terms on the left and right, we have

$$\frac{1}{2M}\|\nabla f(x)\|_2^2 \leq f(x) - f\left(x - \frac{1}{M}\nabla f(x)\right)$$

$$\leq f(x) - f(x^*)$$

Taylor's Expansion & Bounds: Bounds

(4) Proof: $f(y) \leq f(x) + \nabla f(x)^T(y - x) + \frac{M}{2}\|y - x\|_2^2$

(Taylor's exp. + assumption A)

(i) Let $x = x^*$, we have $\nabla f(x^*) = 0$,

thus, we can write the above eq.

$f(y) \leq f(x^*) + \frac{M}{2}\|y - x^*\|_2^2$

or $f(y) - p^* \leq \frac{M}{2}\|y - x^*\|_2^2$

(ii) From (3), we have

$\frac{1}{2M}\|\nabla f(x_o)\|_2^2 \leq f(x_o) - p^*$

(iii) From (i)&(ii), we have

$\frac{1}{2M}\|\nabla f(x_o)\|_2^2 \leq \frac{M}{2}\|x_o - x^*\|_2^2$

Therefore, we have

$\frac{1}{M}\|\nabla f(x_o)\|_2 \leq \|x_o - x^*\|_2$