

Lecture 13. Strongly convex. Condition number and exact line search.

Gradient descent with fixed step size t : $\sum_{k=1}^T f(x_k) - f(x^*) \leq \frac{\|x_0 - x^*\|^2}{2t}$

If $f(x_{k+1}) < f(x_k)$. $\Rightarrow f(x_T) - f(x^*) \leq \frac{1}{2tT} \|x_0 - x^*\|^2$.

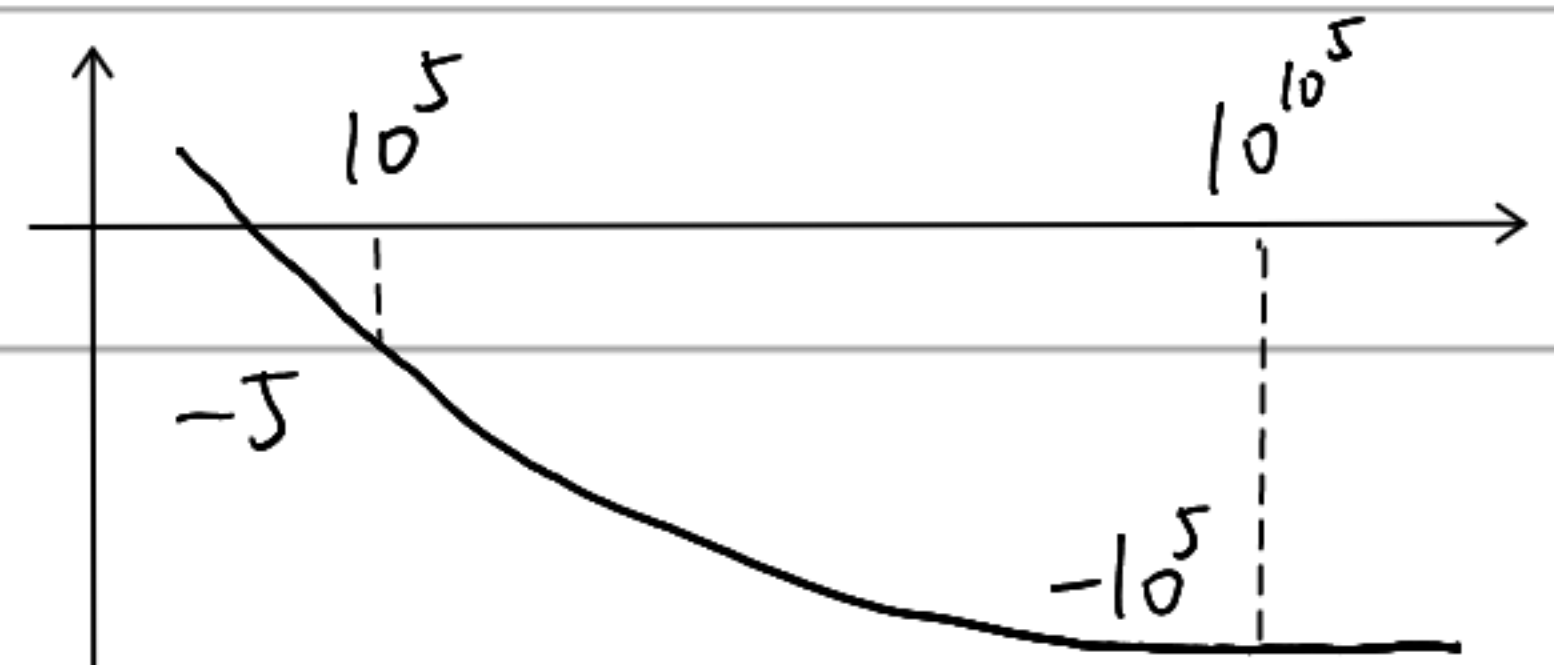
o.w. $f(\frac{1}{T} \sum x_k) \leq \frac{1}{T} \sum f(x_k)$ by convexity. f is L -smooth.

$\Rightarrow f(\frac{1}{T} \sum x_k) - f(x^*) \leq \frac{1}{2tT} \|x_0 - x^*\|^2$. $t < 1/L$.

Remark: rate of convergence is $O(1/T)$; $T = O(1/\epsilon)$ to get ϵ -approximation.

Consider the following function:

$f(x) = -\log x$. $x < 10^{10^5}$ and -10^5 o.w.



$f'(x) = 1/x$. $x_{k+1} \leftarrow x_k + t/x_k$. stop at $|f'(x)| < 10^{-5}$. Bad!

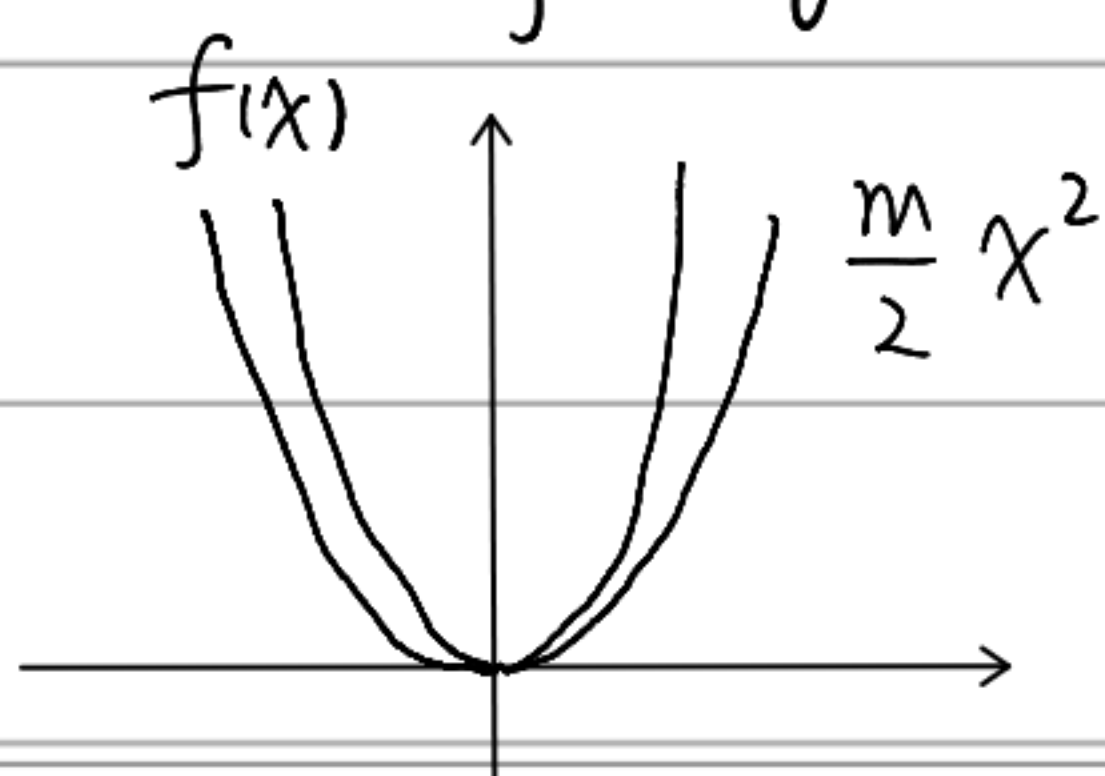
Another bad example: $f(x) = x^4$. if $|x| \leq 1$. $4|x| - 3$. o.w.

$x_{k+1} \leftarrow x_k - 4x_k^3 \cdot t = x_k(1 - 4tx_k^2)$. $x_k \sim (8+k)^{-1/2}$. $f(x_k) \sim (8+k)^{-2}$

Good example: $f(x) = 6x^2$. $x_{k+1} = x_k(1 - 12t)$. $f(x_k) = 6x_0^2(1 - 12t)^{2k}$.

Strongly convex: f is strongly convex with $m > 0$ or m -strongly convex:

if $g(x) = f(x) - \frac{m}{2} \|x\|^2$ is convex.



should be strongly convex everywhere.

$$f(x) - \frac{m}{2} \|x - y\|^2 = g(x) - \frac{m}{2} (\|y\|^2 - 2x^T y).$$

First-order condition: a differentiable f is m -strongly convex.

$$\text{iff } f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{m}{2} \|y-x\|^2 \quad \forall x, y.$$

Proof: $g(x) = f(x) - \frac{m}{2} \|x\|^2$ convex $\Leftrightarrow g(y) \geq g(x) + \nabla g(x)^T (y-x)$

$$\Leftrightarrow f(y) - \frac{m}{2} \|y\|^2 \geq f(x) - \frac{m}{2} \|x\|^2 + (\nabla f(x) - mx)^T (y-x).$$

$$\Leftrightarrow f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{m}{2} (\|x\|^2 + \|y\|^2 - 2x^T y). \quad \square$$

Remark: L -smoothness $\Rightarrow f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{L}{2} \|y-x\|^2$

Second-order condition: a twice continuous differentiable f is

m -strongly convex iff $\nabla^2 f(x) \succeq mI$. i.e. $\lambda_{\min}(\nabla^2 f(x)) \geq m$.

Example. x^4 , $-\log x$ are not strongly convex. $f'' = 12x^2 - m$.

Example. $f(x) = w^T x$ is not strongly convex. $f'' = -m$.

$f(x) = x^T Q x$ is $2\lambda_{\min}(Q)$ -strongly convex if $Q \succ 0$.

Consider $f(x) = ax^2$, $a > 0$. $x_{k+1} = x_k(1-2t) = (1-2t)^{k+1} x_0$.

Theorem: If f is m -strongly convex and L -smooth. fix $t < 1/L$ and

assume $x^* = \text{argmin } f$. and $\{x_k\}$ given by the gradient descent method.

then. $\|x_k - x^*\|^2 \leq (1-2mt)^k \|x_0 - x^*\|^2$

Remark: $\Rightarrow f(x_k) - f(x^*) \leq \frac{L}{2} (1-2mt)^k \|x_0 - x^*\|^2$.

Proof. $\|x_{k+1} - x^*\|^2 = \|x_k - x^*\|^2 + t^2 \|\nabla f(x_k)\|^2 - 2t \nabla f(x_k)^T (x_k - x^*)$.

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

$$\Rightarrow \|x_{k+1} - x^*\|^2 \leq (1 - mt) \|x_k - x^*\|^2 + t^2 \|\nabla f(x_k)\|^2 + 2t(f(x^*) - f(x_k))$$

(recall $f(x_{k+1}) \leq f(x_k) - \frac{t}{2} \|\nabla f(x_k)\|^2$ by L -smoothness)

$$\Rightarrow \|x_{k+1} - x^*\|^2 \leq (1 - mt) \|x_k - x^*\|^2 + 2t(f(x^*) - f(x_{k+1})) \quad \square$$

Key difficulty: how to select step size t ?

Convergence rate for quadratic function $f(x) = x^T Q x$. $Q > 0$

$f(x)$ is $2\lambda_{\max}$ -smooth and $2\lambda_{\min}$ -strongly convex. select $t < 1/\lambda_{\max}$.

let $Q = U^T \Lambda U$ where $\Lambda = \text{diag}\{\lambda_{\min}, \dots, \lambda_{\max}\}$. $\nabla f(x) = 2Qx$.

$$\Rightarrow x_{k+1} = (I - 2tQ)x_k = U^T \Lambda' U x_k \text{ where } \Lambda' = I - 2t \text{diag}\{\lambda_{\min}, \dots, \lambda_{\max}\}$$

$$\Rightarrow x_k = (U^T \Lambda' U)^k x_0 = U^T (\Lambda')^k U x_0. \text{ let } y_k = U x_k. y^* = U x^* = 0.$$

$$y_k = (\Lambda')^k y_0 = \left(\text{diag}\{1 - 2t\lambda_{\min}, \dots, 1 - 2t\lambda_{\max}\} \right)^k y_0$$

$$= \text{diag}\{(1 - 2t\lambda_{\min})^k, \dots, (1 - 2t\lambda_{\max})^k\} y_0.$$

$$\Rightarrow \|x_k\|^2 = \|y_k\|^2 = \sum_{i=1}^n (1 - 2t\lambda_i)^{2k} y_{0i}^2 \quad \lambda_{\min} \leq \lambda_i \leq \lambda_{\max}$$

$$\text{convergence rate} = \max \{(1 - 2t\lambda_i)^{2k}\} = \max \{(1 - 2t\lambda_{\min})^{2k}, (1 - 2t\lambda_{\max})^{2k}\}$$

$$\text{select } t \text{ to } \min \max \{|1 - mt|, |1 - Lt|\} \Rightarrow t = 2/(m+L).$$

