EECS 127 Lecture 12

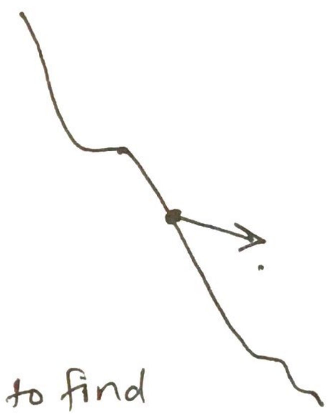
$$p^* = \min_{x \in \mathbb{R}^n} f_0(x)$$

Unconstained optimization.

$$f_0(\vec{x}) = \|A\vec{x} - \vec{b}\|_2^2$$

Optimum solution: $f_0(\vec{x}^*)$

Gradient Descent



$$f(x + \Delta x) = f(x) + \nabla f(x)^T \cdot \Delta x + \cdots$$

$$f(\vec{x} + s\vec{v}) = f(\vec{x}) + \nabla f(\vec{x})^T \cdot s \cdot \vec{v}$$

s: stepsize. scalar

$$= f(\vec{x}) + s \cdot \langle \nabla f(\vec{x}), \vec{v} \rangle$$

$\langle \nabla f(\vec{x}), \vec{v} \rangle > 0$ We don't want this, we want to find a min.

$\langle \nabla f(\vec{x}), \vec{v} \rangle < 0$

$$\langle \nabla f(\vec{x}), \vec{v} \rangle^2 \leq \|\nabla f(\vec{x})\|_2^2 \, \|\vec{v}\|_2^2$$

To minimize this inner product, Cauchy Schwartz : $\vec{v} = -\nabla f(\vec{x})$

GD: $\vec{x}_{k+1} = \vec{x}_k - \eta \cdot \nabla f(\vec{x}_k)$

x_0 : initial point.

η : stepsize.

How + when does gradient descent converge to the minimum point?

$\vec{x}_1, \vec{x}_2 \dots \vec{x}_n \dots \vec{x}_*$?

Example: $f(\vec{x}) = \|A\vec{x} - \vec{b}\|_2^2$

$$= (A\vec{x} - \vec{b})^T (A\vec{x} - \vec{b})$$

$$\nabla f(x) = 2A^TA\vec{x} - 2A^T\vec{b}$$

GD: $\vec{x}_{k+1} = \vec{x}_k - \eta \cdot \nabla f(\vec{x}_k)$.

$$= \vec{x}_k - \eta \cdot (2 \cdot A^TA\vec{x}_k - 2A^T\vec{b}).$$

$$\vec{x}_{k+1} = (I - 2\eta \cdot A^TA)\vec{x}_k - 2\eta \cdot A^T\vec{b}.$$

Care about: $\lambda_{abs}(I - 2\eta \cdot A^TA) < 1.$

$y_{k+1} = a \cdot y_k + b.$

$y_0 = 1.$

$a = 2$ $\quad \lim_{k\to\infty} y_k = \infty$

$a = \frac{1}{2}$ $\quad \lim_{k\to\infty} y_k \neq \infty$

$\sum a^i b$

A full col. rank.

$$\left(\vec{x}_{k+1} - (A^TA)^{-1}A^T\vec{b}\right) = (I - 2\eta A^TA)\vec{x}_k + 2\eta \cdot A^T\vec{b} - (A^TA)^{-1}A^T\vec{b}$$

$$= \text{——} \;''\; \text{——} + 2\eta \cdot (A^TA) \cdot (A^TA)^{-1} \cdot A^T\vec{b} - (A^TA)^{-1}A^T\vec{b}$$

$$= \text{——} \;''\; \text{——} + (2\eta \cdot A^TA - I) \cdot (A^TA)^{-1}A^T \cdot \vec{b}$$

$$= (I - 2\eta \cdot A^TA) \cdot \left(\vec{x}_k - (A^TA^{-1})A^T \cdot \vec{b}\right)$$

$$\vdots$$

$$= \underbrace{(I - 2\eta \cdot A^TA)^{k+1}}\left(\vec{x}_0 - (A^TA)^{-1}A^T\vec{b}\right)$$

$$\left| \text{eigenvalues of } (I - 2\eta \cdot A^TA) \right| < 1.$$

If $\vec{x}_k = \vec{x}_*$

$$\vec{x}_{k+1} = \vec{x}_* - \eta \cdot \underbrace{\nabla \cdot f(\vec{x}_*)}_{0} = \vec{x}_*$$

$\underline{(A^TA)^{-1}} A^T \vec{b}$: $O(n^3)$

Gradient computation: $O(n^2)$

$n = 10^6$

How do we generalize this idea?

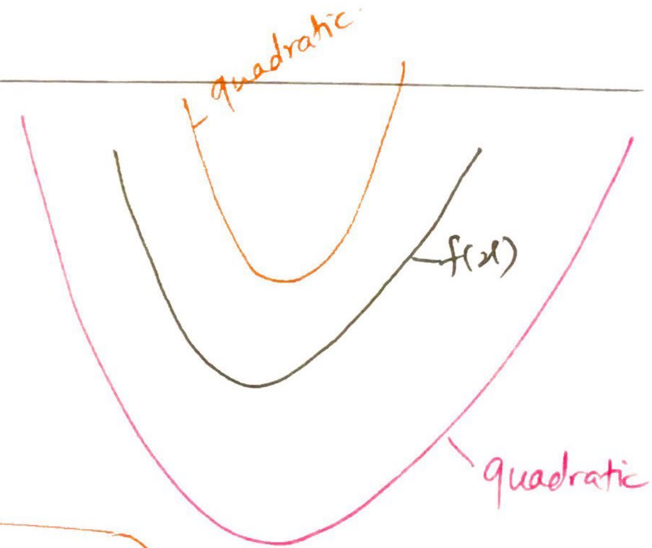
GD for smooth + strongly convex functions.

μ- strongly convex : $\forall x, x'$

$$f(x') \geq f(x) + \nabla f(x)^T (x' - x) + \frac{\mu}{2} \|x' - x\|_2^2$$

L-smooth $\forall x, x'$

$$f(x') \leq f(x) + \nabla f(x)^T (x' - x) + \frac{L}{2} \| x' - x\|_2^2 .$$



Main Thm: Convergence of GD

$$\| x_{t+1} - x^* \|_2^2 \leq (C)^{t+1} \cdot \| x_0 - x^* \|_2^2$$

Lemma: L-smooth.

$$\| \nabla f(x) \|_2^2 \leq 2L \cdot (f(x) - f(x^*)).$$

$(\eta = \frac{1}{L}).$

Proof: We are at x. $\vec{x} \longrightarrow \vec{x} - \frac{\nabla f(\vec{x})}{L}$

$$f(\vec{x}^*) \leq f(\vec{x})$$

$$f(\vec{x}^*) \leq f\left(\vec{x} - \frac{\nabla f(\vec{x})}{L}\right)$$

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

$$= f(x) - \frac{1}{L} \cdot \| \nabla f(\vec{x}) \|_2^2 + \frac{L}{2} \| \nabla f(\vec{x}) \|_2^2 \cdot \frac{1}{L^2}$$

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

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

$$\Rightarrow \|\nabla f(\vec{x})\|_2^2 \leq 2L \cdot (f(x) - f(x^*)).$$

Proof of Main Thm:

μ- strongly convex: $f(x^*) \geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2} \|x^* - x\|_2^2$

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

Proof:

Proof:

$$\|x_{t+1} - x^*\|_2^2 = \| x_t - \eta \cdot \nabla f(x_t) - x^*\|_2^2$$ Grad. step.

$$= \| \underbrace{(x_t - x^*)} - \underbrace{\eta \cdot \nabla f(x_t)} \|_2^2$$

$$= \|x_t - x^*\|_2^2 + \eta^2 \cdot \|\nabla f(x_t)\|_2^2 - 2\eta \cdot \langle \nabla f(x_t), x_t - x^* \rangle$$

$$\leq \|x_t - x^*\|_2^2 + \eta^2 \cdot 2L\left(f(x_t) - f(x^*)\right) - 2\eta \left(\underbrace{f(x_t) - f(x^*) + \frac{\mu}{2}\|x_t - x^*\|_2^2} \right)$$

true by Lemma
L-smooth.

μ-strongly convex bound

$$= (1 - \eta \cdot \mu)\|x_t - x^*\|_2^2 + (2\eta^2 L - 2\eta)\left(f(x_t) - f(x^*)\right)$$

Choose $\eta \neq 0$
$\eta = \frac{1}{L}$.

$$= \left(1 - \frac{\mu}{L}\right)\|x_t - x^*\|_2^2$$

Recursing: $\|x_{t+1} - x^*\|_2^2 \leq \left(1 - \frac{\mu}{L}\right)^{t+1} \|x_0 - x^*\|_2^2$

f: <u>differentiable</u>:

f: strongly convex, smooth: $O(\exp(-t))$.

f: strongly convex, Lipschitz. $O(\frac{1}{t})$

f: convex, smooth $O(\frac{1}{t})$

f: convex, Lipschitz $O(\frac{1}{\sqrt{t}})$