AA Lecture 22 4/1

TE: $f(x+\delta) = f(x) + \nabla f(x)^T \cdot \delta + \frac{1}{2}\delta^T \nabla^2 f(y)\,\delta$

$\delta = -\eta \cdot \nabla f(x)$.

A#1: $\lambda_{max}(\nabla^2 f(y)) \le \beta$.

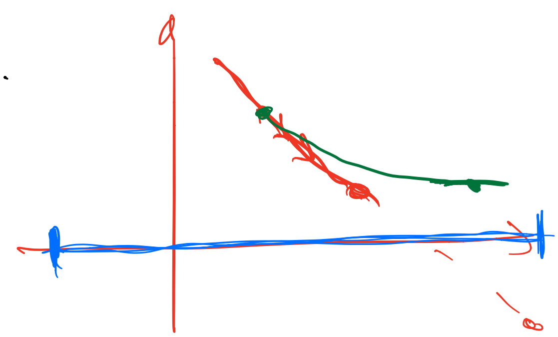
$\eta = \frac{1}{2\beta}$.

A#2: $\lambda_{min}(\nabla^2 f(y)) \geq 0$.

$\boxed{f(x) - f(x^*) \le \varepsilon}$

I: $T = O\left(\frac{\beta \cdot D^2}{\varepsilon}\right)$.

$D = \max_{x:} \|x - x^*\|$.



A#3: $\lambda_{min}(\nabla^2 f(y)) \geq \alpha > 0$.

f: α – strongly convex

Dependency of $\|x - x^*\|$ vs $f(x) - f(x^*)$.

TE + α-s-c:

$$f(x) = f(x^* + \underbrace{(x-x^*)}_{\delta}) \geq f(x^*) + \cancelto{0}{\nabla f(x^*)}\cdot\delta + \frac{1}{2}\delta^T \cdot \alpha \cdot \delta$$

$$f(x) - f(x^*) \geq \frac{\alpha}{2}\cdot\|x-x^*\|^2 \qquad (*)$$

Progress in each GD step:

$x^t \to x^{t+1}$:

$$f(x^{t+1}) - f(x^*) \leq f(x^t) - f(x^*) - \frac{1}{2\beta}\|\nabla f(x^t)\|^2$$

$$\overset{\text{last time}}{\leq} f(x^t) - f(x^*) - \frac{1}{2\beta}\,\frac{[f(x^t)-f(x^*)]^2}{\|x^t-x^*\|^2}$$

$$\overset{(*)}{\leq} f(x^t) - f(x^*) - \frac{1}{2\beta}\cdot\frac{[f(x^t)-f(x^*)]^2}{2/\alpha\cdot(f(x^t)-f(x^*))}$$

$$= [f(x^t)-f(x^*)]\cdot\left[1-\frac{\alpha}{4\beta}\right].$$

$\Rightarrow$ ~~E~~ in T steps:

$$f(x^T) - f(x^*) \leq [f(x^0)-f(x^*)]\cdot\underbrace{\left(1-\frac{\alpha}{4\beta}\right)^T}.$$

for $T = \frac{4\beta}{\alpha}\cdot\ln\frac{f(x^0)-f(x^*)}{\varepsilon}$

$$\hookrightarrow \le e^{-T\cdot\frac{\alpha}{4\beta}} = \frac{\varepsilon}{f(x^0)-f(x^*)}.$$

$$\Rightarrow f(x^T) - f(x^*) \le \varepsilon.$$

$\underline{T}$: α-s-c, β-sm function f,

$$T = O\left(\frac{\beta}{\alpha}\cdot \lg \frac{f(x^0)-f(x^*)}{\varepsilon}\right).$$

$$\frac{\beta}{\alpha} = \frac{\lambda_{max}(\nabla^2 f(y))}{\lambda_{min}(\nabla^2 f(y))} = K(\nabla^2 f(y)).$$

$\underline{TE}$:

$$f(x+\delta) = f(x) + \nabla f(x)^T\cdot\delta + \frac{1}{2}\delta^T \nabla^2 f(y)\,\delta$$

<u>Newton's method</u>: (2^{nd} order method)

<u>Idea</u>: change of variables.

$\Delta = A\cdot\delta$, A = full-rank matrix $n\times n$.

Goal: optimize $f(x+\delta)$ as function of δ.

$\Rightarrow \delta = A^{-1}\Delta$.

TE: $f(x+\delta) = f(x) + \nabla f(x)^T \cdot A^{-1} \cdot \Delta +$

$+ \Delta^T \cdot (A^{-1})^T \cdot \nabla^2 f(y) \cdot A^{-1} \cdot \Delta.$

Best change of vars A s.t.

$$(A^{-1})^T \cdot \nabla^2 f(y) \cdot A^{-1} = I = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}.$$

$A = [\nabla^2 f(y)]^{1/2}$. ← defined as long as $\lambda_{min}(\nabla^2 f(y)) \geq 0$.

$$\delta = \arg\min_{\delta:\ \Delta = A\delta} \nabla f(x)^T \cdot A^{-1} \cdot \Delta + \frac{1}{2} \underbrace{\Delta^T \Delta}_{= \|\Delta\|^2}$$

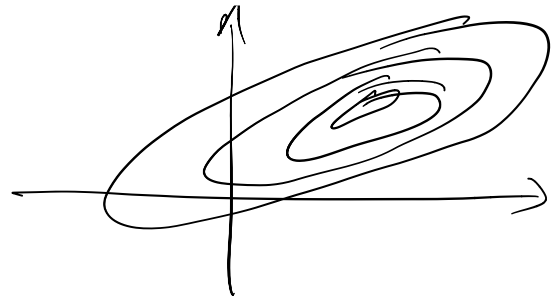
$\Rightarrow \Delta = -\eta \cdot (A^{-1})^T \cdot \nabla f(x).$

$\Rightarrow \Delta = -\eta \cdot [\nabla^2 f(y)]^{-1/2} \cdot \nabla f(x).$

opt η is $\eta = 1$.

$\Rightarrow \delta = A^{-1} \cdot \Delta = -\eta \cdot [\nabla^2 f(y)]^{-1} \cdot \nabla f(x).$

$x^t, \quad x^{t+1} = x^t + \delta$ ✓



Issue: we don't know $\nabla^2 f(y)$ or y.

Note: time to compute δ is now $\sim n^3$: time to invert $\nabla^2 f(y)$

Theorem [Newton's m.]:

$$n(x) \triangleq -\left[\nabla^2 f(x)\right]^{-1} \cdot \nabla f(x).$$

Suppose $\exists\, r > 0$ s.t. $\forall\, x, y$ at distance $\leq r$ from x^*:

1) $\lambda_{min}(\nabla^2 f(x)) \geq \alpha$.

2) $\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L \cdot \|x - y\|$

operator norm.

Then: $\forall x^0$ at dist $\leq r$ from x^*:

$$x^1 = x^0 + n(x^0)$$

$$\|x^1 - x^*\| \le \frac{L}{2\alpha} \cdot \|x^0 - x^*\|^2.$$

Let $\gamma = \frac{2\alpha}{L}$.

$$\Rightarrow \left\|\frac{x^1 - x^*}{\gamma}\right\| \le \frac{L^2}{(2\alpha)^2}\|x^0 - x^*\|^2 = \left\|\frac{x^0 - x^*}{\gamma}\right\|^2.$$

Suppose x^0 is such that $\left\|\frac{x^0 - x^*}{\gamma}\right\| \le 0.9$

$\Rightarrow$ ith T iterations:

$$\left\|\frac{x^T - x^*}{\gamma}\right\| \le \left\|\frac{x^0 - x^*}{\gamma}\right\|^{2^T} \le 0.9^{2^T} \le \varepsilon$$

when $T = \lg_2 \lg_{\frac{1}{0.9}} \frac{1}{\varepsilon} = O(\lg \lg 1/\varepsilon)$.

Remark: $\|x^0 - x^*\| \le 0.9 \cdot \frac{2\alpha}{L}$.

warm start

next Interior Point Methods:

$$\min_{x \in K} f(x)$$

eg: $K = \{x : Ax \le b\}$.
$f(x) = c^T x$.

How to put it into unc. opt. framework of GD?

<u>Idea</u>: define $F(x) = \begin{cases} f(x), & x \in K \\ +\infty, & x \notin K. \end{cases}$