Comparing with GD: $\frac{1}{T^2}$ instead of $\frac{1}{T}$.

LECTURE - 4

$$f(x) = \frac{1}{2}\|Ax-b\|^2 \; ; \quad L = \|A\|^2$$

$$\overset{AGD}{\Longrightarrow} \quad f(x_T) - f(x^*) \leq \frac{6\,\|A\|^2 \cdot \|x_0 - x^*\|^2}{T^2}$$

Can we do better?

$\hookrightarrow$ Strong convexity.

Strong Convexity : f is μ- strongly convex

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|x - y\|^2 .$$

Linear regression : $\mu = \sigma_{min}(AA^T)$.

$\frac{1}{2} \|Ax - b\|^2$

GD: $x_{t+1} = x_t - \eta \nabla f(x_t)$.

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

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

$$+ \eta^2 \|\nabla f(x_t)\|^2$$

Strong Convexity

Smoothness

$$f(x^*) \geq f(x_t) + \langle \nabla f(x_t), x^* - x_t \rangle + \frac{\mu}{2}\|x_t - x^*\|^2$$

$$\|x_{t+1}-x^*\|^2 \leq \|x_t - x^*\|^2 - 2\eta\left[f(x_t)-f(x^*) + \frac{\mu}{2}\|x_t - x^*\|^2\right] + \eta^2 \cdot 2L\left[f(x_t)-f(x^*)\right]$$

$$\leq (1-\eta\mu)\|x_t - x^*\|^2 \underbrace{- 2\eta(1-\eta L)\left[f(x_t)-f(x^*)\right]}_{\leq 0}$$

If we choose $\boxed{\eta \leq \frac{1}{L}}$ then

$$\|x_{t+1}-x^*\|^2 \leq (1-\eta\mu)\|x_t - x^*\|^2.$$

$$\leq (1-\eta\mu)^{t+1}\|x_0 - x^*\|^2.$$

$$f(x_t) - f(x^*) \leq \frac{L}{2} \| x_t - x^* \|^2$$

$$\leq \frac{L}{2} \cdot (1 - \eta\mu)^t \| x_0 - x^* \|^2.$$

$$\leq \frac{L \cdot \| x_0 - x^* \|^2}{2} \cdot e^{-\eta \mu t}.$$

$\forall \omega \in \mathbb{R}$

FACT: $1 - \omega \leq e^{-\omega}$

Choose η as large as possible

So, $\eta = \frac{1}{L}$.

$$f(x_t) - f(x^*) \leq \frac{L \| x_0 - x^* \|^2}{2} \cdot e^{-\boxed{\frac{\mu}{L}} t}.$$

condition number

$\kappa = \frac{L}{\mu}$

≥ 1

$$f(x_t) - f(x^*) \leq \frac{L\|x_0 - x^*\|^2}{2} \cdot \underbrace{e^{-\frac{t}{\kappa}}}$$

Is this the best possible rate?

Acceleration for Smooth and Strongly Convex functions

Nesterov's AGD : $f(x_t) - f(x^*) \leq \frac{L\|x_0 - x^*\|^2}{2t^2}$.

$$\leq \frac{L}{\mu} \cdot \left[f(x_0) - f(x^*)\right] \cdot \frac{1}{t^2}$$

$$f(x_0) \geq f(x^*) + \langle \cancel{\nabla f(x^*)}, x_0 - x^* \rangle + \frac{\mu}{2}\|x_0 - x^*\|^2$$

$=0$

$$= \frac{K}{t^2}\left[f(x_0)-f(x^*)\right].$$

$$\boxed{\|x_0 - x^*\|^2 \leq \frac{2}{\mu}\left[f(x_0) - f(x^*)\right]}$$

Choosing $t = \sqrt{2K}$ iterations,

$$f(x_{\sqrt{2K}}) - f(x^*) \leq \frac{1}{2}\left[f(x_0) - f(x^*)\right]$$

$$f(x_{2\cdot\sqrt{2K}}) - f(x^*) \leq \frac{1}{2}\left[f(x_{\sqrt{2K}}) - f(x^*)\right]$$

$$f(x_{i\cdot\sqrt{2K}}) - f(x^*) \leq \frac{1}{2}\left[f(x_{(i-1)\cdot\sqrt{2K}}) - f(x^*)\right]$$

$$f(x_t) - f(x^*) \leq \left(\frac{1}{2}\right)^{\left(t/\sqrt{2\kappa}\right)} \left[f(x_0) - f(x^*)\right].$$

GD: $e^{-t/\kappa}$; AGD: $e^{-t/\sqrt{\kappa}}$

$\|x_0 - x^*\| \le R$

CLASS	Algorithm	Guarantee
G-Lipschitz $\|\nabla f\| \le G$	SUBGRADIENT DESCENT	$\frac{GR}{\sqrt{T}}$
L-SMOOTH $\|\nabla f(x) - \nabla f(y)\| \le L\|x-y\|$	G.D.	$\frac{LR^2}{T}$
	A.G.D.	$\frac{LR^2}{T^2}$
L-SMOOTH μ- Str. Cvx.	G.D.	$LR^2 \cdot \exp\left(\frac{-\mu T}{L}\right)$
	A.G.D.	$LR^2 \cdot \exp\left(-\sqrt{\frac{\mu}{L}} \cdot T\right)$

G- Lipschitz μ- Str. cvx.	Sub. G.D	$\frac{G^2}{\mu T}$

CONSTRAINED OPTIM.

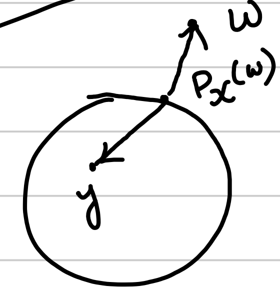
$$\min_{x \in \mathcal{X}} f(x)$$

$\mathcal{X}$ is a Simple closed Convex set.

We have access to projection operator

Given x, $\quad P_{\mathcal{X}}(x) = \underset{y \in \mathcal{X}}{\operatorname{argmin}} \|x - y\|^2.$

$\mathcal{X} = B_1(w)$ is a simple set



Pythagoras theorem: $\forall w$, we have

EXERCISE

i) $\langle w - P_{\mathcal{X}}(w), y - P_{\mathcal{X}}(w) \rangle \leq 0$

ii) $\| P_{\mathcal{X}}(w) - y \|^2 \leq \| w - y \|^2 \quad \boxed{\forall y \in \mathcal{X}}$

Projected GD: $x_{t+1} = P_{\mathcal{X}}\left(x_t - \eta \nabla f(x_t)\right)$.

Equivalent

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \underbrace{\left[\overbrace{f(x_t) + \langle \nabla f(x_t), x - x_t \rangle}^{\text{Linear}} + \overbrace{\frac{1}{2\eta} \|x - x_t\|^2}^{\frac{1}{\eta}\text{-str. conv. quadratic}} \right]}_{\hat{f}_\eta(x_t;\, x)}.$$

Doing now

Conv. rate of PGD for L-Smooth fn.

Exercise: $\hat{f}_\eta(x_t; x)$ is $\frac{1}{\eta}$ - Strongly Convex

(1) $\longrightarrow$ $\hat{f}_\eta(x_t; x) \geq \hat{f}_\eta(x_t; x_{t+1}) + \frac{1}{2\eta}\|x - x_{t+1}\|^2$

$\boxed{+ < \quad >}$ ↳ Use Pythagoras thm. $\left(\because x_{t+1} \text{ is } \arg\min \hat{f}_\eta(x_t; x)\right)$

For $\eta \leq \frac{1}{L}$; $f(x) \leq \hat{f}_\eta(x_t; x)$ [by L-smoothness of $f(\cdot)$]

$$f(x) \overset{\text{Conv.}}{\geq} f(x_t) + \langle \nabla f(x_t), x - x_t \rangle$$

$$= \hat{f}_\eta(x_t; x) - \frac{1}{2\eta}\|x - x_t\|^2$$

$$\overset{1}{\geq} \hat{f}_\eta(x_t; x_{t+1}) + \frac{1}{2\eta}\|x - x_{t+1}\|^2 - \frac{1}{2\eta}\|x - x_t\|^2$$

$$\overset{\text{Smoothness}}{\geq} f(x_{t+1}) + \frac{1}{2\eta}\|x - x_{t+1}\|^2 - \frac{1}{2\eta}\|x - x_t\|^2$$

$$\|x - x_{t+1}\|^2 \le \|x - x_t\|^2 - 2\eta\left[f(x_{t+1}) - f(x)\right]$$

$x = x^*$ and iterate

$$\|x^* - x_{t+1}\|^2 \le \|x^* - x_0\|^2 - 2\eta \underbrace{\sum_{s=1}^{t+1}\left[f(x_s) - f(x^*)\right]}.$$

$$\frac{1}{t+1}\sum_{s=1}^{t+1}\left[f(x_s) - f(x^*)\right] \le \frac{\|x^* - x_0\|^2 - \|x^* - x_{t+1}\|^2}{2\eta(t+1)}.$$

$$\left[\eta = \frac{1}{L}\right] \le \frac{L\|x_0 - x^*\|^2}{2(t+1)}. \quad \square$$

Ref: INTRODUCTORY LECTURES ON CONVEX OPT.

— YURII NESTEROV