LECTURE - 3

Note Title 26-Jun-19

f is Convex

$\|\nabla f(x)\| \leq G$

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

GD: $\frac{1}{T} \sum_{t=1}^{T} \left(f(x_t) - f(x^*)\right) \leq \frac{GR}{\sqrt{T}}$

Running example : $f(x) = \underline{\frac{1}{2} \|Ax - b\|^2}$.

$x_0 = 0 \overset{\text{Sub.Gr.De.}}{\Longrightarrow}$

$R \triangleq \|x_0 - x^*\| = \|x^*\|$

$G \triangleq \underline{\max_{x : \|x\| \leq R} A^T(Ax - b)} \longrightarrow G = \|A\|^2 \cdot R + \|A\| \cdot \|b\|$

$$\frac{1}{T}\sum_{t=1}^{T}\left(f(x_t)-f(x^*)\right) \leq \frac{GR}{\sqrt{T}} = \frac{\|A\|^2 \cdot R^2 + \|A\| \cdot \|b\| \cdot R}{\sqrt{T}}.$$

Qn.: Can we do better?

↳ Say for Smooth functions.

Smoothness: $\|\nabla f(x) - \nabla f(y)\| \leq L \cdot \|x-y\|.$

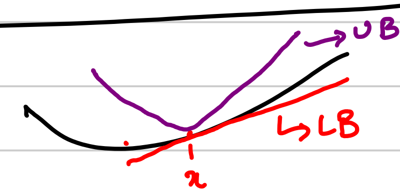
$$\downarrow$$

$$\|A^T(Ax-b) - A^T(Ay-b)\| = \|A^T A(x-y)\|$$

So, $L \triangleq \|A^T A\|.$

<u>This lecture</u>: ① Convergence rate of GD for smooth convex fns.

② Optimal method: Nesterov's accelerated gradient.



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

Convexity: $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$

Smoothness: $f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|x - y\|^2.$

Thm: If f is an L-smooth convex fn. then GD with $\eta = \boxed{\frac{1}{2L}}$, we have $\underbrace{f(x_T)}_{\text{Final iterate}} - f(x^*) \leq \frac{C \cdot L \cdot \|x_0 - x^*\|^2}{T}$.

Lemma: If $f(\cdot)$ is L-smooth then $\underbrace{\|\nabla f(x)\|^2 \leq 2L \cdot [f(x) - f(x^*)]}$.

Proof:
$$f(x^*) \leq f(\underbrace{x - \tfrac{1}{L}\nabla f(x)})$$
$$\leq f(x) + \langle \nabla f(x), -\tfrac{1}{L}\nabla f(x)\rangle + \tfrac{L}{2}\|\tfrac{1}{L}\nabla f(x)\|^2$$
$$= f(x) - \tfrac{1}{2L}\|\nabla f(x)\|^2.$$
∎

Proof of thm:

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

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

$$= \|x_t - x^*\|^2 - 2\eta \langle \nabla f(x_t), x_t - x^* \rangle + \eta^2 \underbrace{\|\nabla f(x_t)\|^2}_{\text{Use Lemma above}}$$

$$\leq \|x_t - x^*\|^2 - 2\eta \underbrace{\langle \nabla f(x_t), x_t - x^* \rangle}_{\text{Convexity}} + \eta^2 \cdot 2L\left[f(x_t) - f(x^*)\right]$$

$$\leq \|x_t - x^*\|^2 - 2\eta\left[f(x_t) - f(x^*)\right] + 2\eta^2 L\left[f(x_t) - f(x^*)\right]$$

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

Take telescopic Sum,

$$\|x_{T+1} - x^*\|^2 \leq \|x_0 - x^*\|^2 - 2\eta(1-\eta L)\sum_{t=0}^{T}\left[f(x_t) - f(x^*)\right]$$

$$\frac{1}{T+1}\sum_{t=0}^{T}\left[f(x_t) - f(x^*)\right] \leq \frac{\|x_0 - x^*\|^2 - \|x_{T+1} - x^*\|^2}{2\eta(1-\eta L)\cdot(T+1)} \quad +\text{ve}$$

Choose $\eta = \frac{1}{2L}$ $\therefore \leq \frac{2L\|x_0 - x^*\|^2}{T+1}.$

__Exercise__: For L-Smooth, Convex $f(\cdot)$, Show that GD with $\eta \le \frac{1}{2L}$ Satisfies $f(x_{t+1}) \le f(x_t)$. $\forall t$.

Nesterov's accelerated gradient

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

$$x_{t+1} \triangleq \arg\min_{x} \underbrace{\left[\underbrace{f(x_t) + \langle \nabla f(x_t), x - x_t \rangle}_{\text{First order Taylor exp.}} + \underbrace{\frac{1}{2\eta} \|x - x_t\|^2}_{\text{Quadratic term}} \right]}_{\text{Local upper bound when } \eta \le \frac{1}{L}.}$$

$$f(x) \leq f(x_t) + \langle \nabla f(x_t), x - x_t \rangle + \underbrace{\frac{L}{2} \|x - x_t\|^2}_{\leq \frac{1}{2\eta} \|x - x_t\|^2}$$

If $L \leq \frac{1}{\eta}$ then

Estimate Sequences

$$\phi_0(x) = \underline{f(v_0)} + \frac{L}{2} \| x - v_0 \|^2.$$

$$\phi_{t+1}(x) = (1 - \alpha_t) \phi_t(x) + \alpha_t \Big[\underbrace{f(y_t) + \langle \nabla f(y_t), x - y_t \rangle}_{\leq f(x)} \Big]$$

$$\triangleq \phi_t^*$$

Lemma: If $\exists\, x_t$ s.t. $f(x_t) \leq \min_x \phi_t(x)$ then

$$f(x_t) - f(x^*) \leq \underbrace{\left[\prod_{s=0}^{t-1}(1-\alpha_s)\right]}\left[\phi_0(x^*) - f(x^*)\right].$$

Proof: $f(x_t) - f(x^*) \overset{\text{Hypothesis}}{\leq} \phi_t^* - f(x^*)$

$$\leq \underline{\phi_t(x^*) - f(x^*)}.$$

$-ve$

$$= (1-\alpha_{t-1})\phi_{t-1}(x^*) + \alpha_{t-1}\left[f(y_{t-1}) + \langle \nabla f(y_{t-1}), x^* - y_{t-1}\rangle\right]$$
$$- (1-\alpha_{t-1}) f(x^*) - \alpha_{t-1}\left[f(x^*)\right]$$

$$\phi_t(x^*) - f(x^*) \leq (1-\alpha_{t-1})\left[\phi_{t-1}(x^*) - f(x^*)\right]$$

$$\leq \prod_{s=0}^{t-1}(1-\alpha_s)\left[\phi_0(x^*) - f(x^*)\right]. \quad (11)$$

Given : x_t s.t. $f(x_t) \leq \min_x \phi_t(x)$.

Also given ϕ_t

Task : ① Choose y_t and query $\nabla f(y_t)$.

② choose $\alpha_t \longrightarrow \phi_{t+1}$

③ Find x_{t+1} s.t. $f(x_{t+1}) \leq \min_x \phi_{t+1}(x)$.

Lemma : If $\phi_{t+1}(x) = (1-\alpha_t)\,\phi_t(x) + \alpha_t\left[f(y_t) + \langle \nabla f(y_t), x - y_t\rangle\right]$

then $\phi^*_{t+1} \triangleq \min_x \phi_{t+1}(x) =$ ______

and $V_{t+1} \triangleq \arg\min_x \phi_{t+1}(x) = \underline{V_t - \frac{\alpha_t}{\lambda_t(1-\alpha_t)L}\nabla f(y_t)}$

Proof :

$\phi^*_t \triangleq \min_x \phi_t(x)$

$V_t \triangleq \arg\min_x \phi_t(x)$

$\lambda_t \triangleq \prod_{s=0}^{t-1}(1-\alpha_s)$

$\phi_0(x) = \phi_0^* + \frac{L}{2}\|x - v_0\|^2$

$\phi_1(x) = (1-\alpha_0)\,\phi_0(x) + \alpha_0[a + \langle b, x\rangle]$

$$\phi_t(x) = \phi_t^* + \frac{\lambda_t L}{2} \| x - v_t \|^2.$$

$$\phi_{t+1}(x) = (1-\alpha_t)\,\phi_t(x) + \alpha_t \left[f(y_t) + \langle \nabla f(y_t), x - y_t \rangle \right]$$

$$= (1-\alpha_t)\,\phi_t^* + \frac{(1-\alpha_t)\lambda_t L}{2} \| x - v_t \|^2$$

$$+ \alpha_t f(y_t) + \alpha_t \langle \nabla f(y_t), x - y_t \rangle.$$

$$= \frac{(1-\alpha_t)\lambda_t L}{2} \left[\|x\|^2 - 2 \langle v_t, x \rangle + \underline{\|v_t\|^2} \right]$$

Constant

Quadratic

Linear

$$+\ \alpha_t \underline{\langle \nabla f(y_t), x\rangle} \quad \boxed{\begin{aligned}&+ (1-\alpha_t)\phi_t^* + \alpha_t f(y_t)\\ &- \alpha_t \langle \nabla f(y_t), y_t\rangle\end{aligned}}\ \hat{\theta}$$

$$= \frac{(1-\alpha_t)d_t^2}{2}\Big[\|x\|^2 - 2\left\langle v_t - \frac{\alpha_t}{(1-\alpha_t)d_t^2}\nabla f(y_t), x\right\rangle$$

$$+ \underline{\left\| v_t - \frac{\alpha_t}{(1-\alpha_t)d_t^2}\nabla f(y_t)\right\|^2}$$

Extra terms
$$\left\{\begin{aligned}&+ \frac{2\alpha_t}{(1-\alpha_t)d_t^2}\langle \nabla f(y_t), v_t\rangle\\ &- \left(\frac{\alpha_t}{(1-\alpha_t)d_t^2}\right)^2 \|\nabla f(y_t)\|^2\Big] + \hat{\theta}\end{aligned}\right.$$

$$= \frac{(1-\alpha_t)\, d_t L}{2} \Big\| x - \underbrace{v_t + \frac{\alpha_t}{(1-\alpha_t) d_t L} \nabla f(y_t)}_{v_{t+1}} \Big\|^2 + \text{Extra terms} + \hat{\theta}$$

$$\phi^*_{t+1} = (1-\alpha_t)\phi^*_t + \alpha_t f(y_t) - \frac{\alpha_t^2}{2 d_t (1-\alpha_t) L} \|\nabla f(y_t)\|^2 + \alpha_t \langle \nabla f(y_t), v_t - y_t \rangle .$$

$\blacksquare$

How to find x_{t+1} s.t. $f(x_{t+1}) \leqslant \phi^*_{t+1}$?

By smoothness: $f(\underbrace{y_t - \overset{\frac{1}{L}}{\overset{\shortparallel}{\eta}} \nabla f(y_t)}_{x_{t+1}}) \leq f(y_t) - \eta \, \|\nabla f(y_t)\|^2 + \frac{\eta^2 L}{2} \|\nabla f(y_t)\|^2$

$$\leq f(y_t) - \frac{1}{2L} \|\nabla f(y_t)\|^2$$

$$\underbrace{f(x_{t+1})} - \underbrace{\phi^*_{t+1}} \leq \underbrace{f(y_t) - \frac{1}{2L} \|\nabla f(y_t)\|^2} - (1-\alpha_t)\,\underline{\phi_t^*} - \alpha_t f(y_t) + \frac{\alpha_t^2}{2\alpha_t(1-\alpha_t)L} \|\nabla f(y_t)\|^2 - \alpha_t \langle \nabla f(y_t), v_t - y_t \rangle$$

$$\leq (1-\alpha_t) f(y_t) - (1-\alpha_t)\, \underline{f(x_t)} - \alpha_t \langle \nabla f(y_t), v_t - y_t \rangle$$

(Convexity: $x = y_t$, $y = x_t$)

$$\leq \langle \nabla f(y_t), (1-\alpha_t)(y_t - x_t) \rangle - \alpha_t \langle \nabla f(y_t), v_t - y_t \rangle$$

$$= \langle \nabla f(y_t), y_t - (1-\alpha_t) x_t - \alpha_t v_t \rangle$$

$$\leq 0$$

Want: $\dfrac{\alpha_t^2}{d_t(1-\alpha_t)} \leq 1$

Requirements

$$\frac{\alpha_t^2}{2 d_t (1-\alpha_t) L} \leq \frac{1}{2L}$$

Algorithm

$y_t = (1-\alpha_t) x_t + \alpha_t v_t$

$x_{t+1} = y_t - \frac{1}{L} \nabla f(y_t)$

$v_{t+1} =$ Update from Lemma.

and minimize $\lambda_t = \prod_{s=0}^{t-1} (1-\alpha_s)$.

Can choose α_t: $\boxed{\lambda_{t+1} \leq \frac{4}{t^2}}$.

$$
\begin{aligned}
f(x_T) - f(x^*) &\leq \lambda_T \left[\Phi_0(x^*) - f(x^*) \right] \\
&\leq \frac{4}{(T-1)^2} \left[f(v_0) + \frac{L}{2} \|x^* - v_0\|^2 - f(x^*) \right] \\
&\leq \frac{4}{(T-1)^2} \left[L \|x^* - v_0\|^2 \right].
\end{aligned}
$$