Stochastic Gradient Descent

Recall GD: • f convex • f ℓ-lipschitz • $\|x^*\| \le B$

Bound: 1) $f(\bar{x}) - f(x^*) \le \frac{1}{T}\sum (v^t)^T(x^t - x^*)$ By Jenson's and convexity

$\star$ 2) $\frac{1}{T}\sum (v^t)^T(x^t - x^*) \le \frac{B\ell}{\sqrt{T}}$ By • $\|v^t\| \le \ell$

• $x^{t+1} = x^t - \eta v^t$

• $\eta = \frac{B}{\ell\sqrt{T}}$

SGD

v^t random w/ $E[v^t|x^t] \in \partial f(x^t)$

Assume $P[\|v^t\| \le \ell] = 1$

Denote $v_{1:t} = v_1, \dots, v_t$

$$\underset{v_{1:t}}{E}\left[f(\bar{x}) - f(x^*)\right] \le \underset{v_{1:t}}{E}\left[\frac{1}{T}\sum_{t=1}^{T} f(x^t) - f(x^*)\right] \quad (1)$$

Aside: Fix $v_{1:t}$, By $\star$

$$\frac{1}{T}\sum (v^t)^T(x^t - x^*) \le \frac{B\ell}{\sqrt{T}}$$

$$\underset{v_{1:T}}{E}\left[\frac{1}{T}\sum (v^t)^T(x^t - x^*)\right] \le \frac{B\ell}{\sqrt{T}} \quad (2)$$

Lemma: $$\underset{v_{1:T}}{E}\left[\frac{1}{T}\sum f(x^t) - f(x^*)\right] \le \underset{v_{1:T}}{E}\left[\frac{1}{T}\sum (v^t)^T(x^t - x^*)\right] \quad (3)$$

By (1),(2),(3)

$$\underset{v_{1:T}}{E}[f(\bar{x})] - f(x^*) \le \frac{B\ell}{\sqrt{T}}$$

Pf (3):

$$\underset{v_{1:T}}{E}\left[\frac{1}{T}\sum (v^t)^T(x^t - x^*)\right] = \frac{1}{T}\sum \underbrace{\underset{v_{1:T}}{E}\left[(v^t)^T(x^t - x^*)\right]}$$

$$= \underset{v_{1:t}}{E}\left[(v^t)^T(x^t - x^*)\right]$$

$$= \underset{v_{1:t-1}}{E} \; \underset{v_{1:t}}{E}\left[(v^t)^T(x^t - x^*) \,\middle|\, v_{1:t-1}\right]$$

defines x^t

$$= \underset{v_{1:t-1}}{E} \; \underbrace{\underset{v_t}{E}\left[v_t^T \,|\, v_{t-1}\right]}_{\in \partial f(x^t)} (x^t - x^*)$$

$$\geq \underset{v_{1:t-1}}{E}\left[f(x^t) - f(x^*)\right]$$

$$\geq \frac{1}{T} \sum \underset{v_{1:t-1}}{E}\left[f(x^t) - f(x^*)\right]$$

$$= \underset{v_{1:T}}{E}\left[\frac{1}{T}\sum f(x^t) - f(x^*)\right] \qquad []$$

trong Convexity

n. f is λ-strongly convex if $\forall x, y \in \mathbb{R}^n$ $\forall \alpha \in (0,1)$

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) - \frac{\lambda}{2}\alpha(1-\alpha)\|x-y\|^2$$

hm $\forall x, y \in \mathbb{R}^n$, $\forall v \in \partial f(x)$

$$v^T(x-y) \geq f(x) - f(y) + \frac{\lambda}{2}\|x-y\|^2$$

Thm f is λ-strongly convex. $E[\|v^t\|^2] \le \rho^2$ $\qquad \eta_t = \frac{1}{\lambda t}$

$$E[f(\bar{x})] - f(x^*) \le \frac{\rho^2}{2\lambda T}(1 + \log T)$$

Pf: let $g^t = E[v^t \mid x^t] \rightarrow g^t \in \partial f(x^t)$

$$\underbrace{(g^t)^T(x^t - x^*)} \ge f(x^t) - f(x^*) + \frac{\lambda}{2}\|x^t - x^*\|^2 \qquad (1)$$

Aside.

$$\|x^t - x^*\|^2 - \|x^{t+1} - x^*\|^2 = \cancel{(x^t - x^*)^T(x^t - x^*)} \underbrace{- (x^{t+1} - x^*)^T(x^{t+1} - x^*)}$$

$$-(x^t - \eta v^t - x^*)^T(x^t - \eta v^t - x^*)$$

$$= -((x^t - x^*)^T - \eta v^t)^T((x^t - x^*) - \eta v^t)$$

$$= \cancel{-(x^t - x^*)^T(x^t - x^*)} + 2\eta(v^t)^T(x^t - x^*) - \eta^2\|v^t\|^2$$

$$= 2\eta(v^t)^T(x^t - x^*) - \eta^2\|v^t\|^2$$

$$\cancel{E[(v^t)^T(x^t - x^*) \mid x^t] = E\left[\frac{\|x^t - x^*\|^2 - \|x^{t+1} - x^*\|}{2\eta_t}\right] + \frac{\eta_t}{2}\|v^t\|^2}$$

$$\cancel{(g^t)^T(x^t - x^*) \le \frac{\|x^t - x^*\|^2 - E[\|x^{t+1} - x^*\| \mid x^t]}{2\eta_t} + \frac{\eta_t}{2}\rho^2}$$

Thm: f is λ-strongly convex, $E[\|v_t\|^2] \le \rho^2$

$$E[f(\bar{x})] - f(x^*) \le \frac{\rho^2}{2\lambda T}$$

Take $\bar{x} = \sum_{t=\frac{T}{2}+1}^{T} x^t$

$$E\left[(v^t)^T(x^t - x^*)\right] = E\left[\frac{\|x^t - x^*\|^2 - \|x^{t+1} - x^*\|^2}{2\eta_t} + \frac{\eta_t}{2}\|v^t\|^2\right]$$

$$E\left\{E\left[(v^t)^T(x^t - x^*) \mid x^t\right]\right.$$

$$E\left[(g^t)^T(x^t - x^*)\right] \le \frac{E\left[\|x^t - x^*\|^2 - \|x^{t+1} - x^*\|^2\right]}{2\eta_t} + \frac{\eta_t}{2}\rho^2$$

$$E\left[f(x^t) - f(x^*)\right] \le E\left[\underbrace{\frac{\|x^t - x^*\|^2 - \|x^{t+1} - x^*\|^2}{2\eta_t}}_{} - \frac{\lambda}{2}\|x^t - x^*\|^2\right] + \frac{\eta_t}{2}\rho^2$$

$$\frac{1}{T}\sum$$

$$-\lambda T\|x^{T+1} - x^*\|^2 \le 0$$

$$\frac{1}{T}\sum E\left[f(x^t) - f(x^*)\right] \le \frac{\rho^2}{2\lambda T}\sum \frac{1}{t} \le \frac{\rho^2}{2\lambda T}(1 + \log T)$$

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