CS292F Lecture 17

Convergence of SVRG for L-Lipschitz f_i m-Strongly convex $\frac{1}{n}\sum f_i =: F(x)$

Alg: inner loop, $\tilde{x}$

$$x_{k+1} = x_k - t\underbrace{\left(g_{i_k}^k - g_{i_k}(\tilde{x}) + \frac{1}{n}\sum_{i=1}^n g_i(\tilde{x})\right)}_{=: v_k}$$

$x^* = \arg\min_x F(x)$ $g_i = \nabla f_i$

$$\|x_{k+1} - x^*\|^2 = \|x_k - t v_k - x^*\|^2$$

$$= \|x_k - x^*\|^2 - 2t(x_k - x^*)^T v_k + t^2\|v_k\|^2$$

take expectation $\mathbb{E}[\cdot \mid x_k]$

$$\mathbb{E}[\|x_{k+1} - x^*\|^2 \mid x_k] = \|x_k - x^*\|^2 - 2t(x_k - x^*)^T \nabla F(x_k) + t^2\,\mathbb{E}[\|v_k\|^2 \mid x_k]$$

$$\leq \|x_k - x^*\|^2 - 2t\left(F(x_k) - F(x^*)\right) + 4Lt^2\left(F(x_k) - F(x^*) + F(\tilde{x}) - F(x^*)\right)$$

$$= \|x_k - x^*\|^2 - 2t(1 - 2Lt)\left(F(x_k) - F(x^*)\right) + 4Lt^2\left(F(\tilde{x}) - F(x^*)\right)$$

Lemma (Variance bound of v_k)

$$\boxed{\mathbb{E}[\|v_k\|^2 \mid x_k] \leq 4L\left(F(x_k) - F(x^*) + F(\tilde{x}) - F(x^*)\right)}$$

Let's take $\tilde{x} = \tilde{x}_s$, take full expectation

Take full expectation and sum up everything for $k=0,\dots,N$

$$\mathbb{E}\left[\|x_{k+1}-x^*\|^2 \mid x_k\right] \le \|x_k-x^*\|^2 - 2t(1-2Lt)\left(f(x_k)-f(x^*)\right) + 4Lt^2\left(F(\tilde{x})-f(x^*)\right)$$

$$E\left[\|x_N-x^*\|^2\right] \le E\left[\|x_0-x^*\|^2\right] - 2t(1-2Lt)\Big(\underbrace{E\left[\sum F(x_k)\right]}_{=\,N\,E\left[F(\tilde{x}_{s+1})\right]} - F(x^*)\Big) + 4Lt^2 N\, E\left[F(\tilde{x})-f(x^*)\right]$$

$\Leftarrow$ algorithm on how to update outerloop snapshot

$$\underbrace{\cancel{E\left[\|x_N-x^*\|^2\right]}}_{\ge 0} + 2t(1-2Lt)N\, E\left[F(\tilde{x}_{s+1})-F(x^*)\right]$$

$$\le E\left[\|\tilde{x}_s-x^*\|^2\right] + 4Lt^2 N\, E\left[F(\tilde{x}_s)-F(x^*)\right]$$

~~m-PL condition~~

m-StrongConvex

m-quadratic growth condition

$$\le \frac{2}{m} E\left[F(\tilde{x}_s)-f(x^*)\right] + 4Lt^2 N\, E\left[F(\tilde{x}_s)-F(x^*)\right]$$

$$= 2\left(\frac{1}{m} + 2LNt^2\right) E\left[F(\tilde{x}_s)-F(x^*)\right]$$

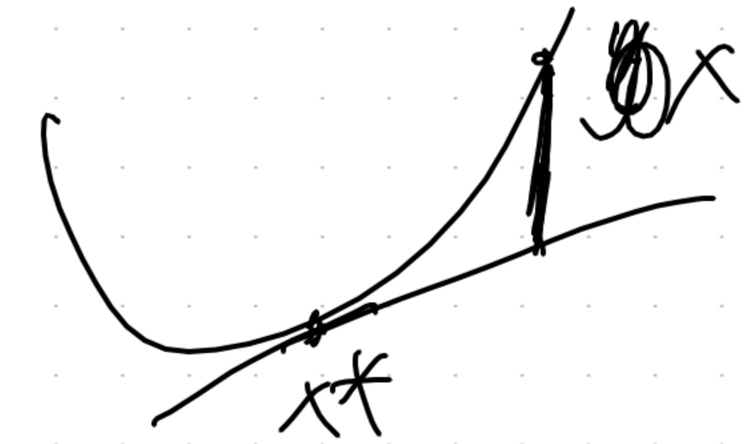
this implies:

$$E\left[F(\tilde{x}_{s+1})-F(x^*)\right] \le \left[\frac{1}{mt(1-2Lt)N} + \frac{2Lt}{1-2Lt}\right] E\left[F(\tilde{x}_s)-F(x^*)\right]$$

$< \frac{2}{3}$

Bregman divergence

$$g_i(x) = f_i(x) - f_i(x^*) - \nabla f_i(x^*)^T (x - x^*)$$



g_i is convex, also g_i is also L-smooth

$$\nabla g_i(x) = \nabla f_i(x) - \nabla f_i(x^*)$$

Descent Lemma if $t = \frac{1}{L}$ * this t is not the learning rate used by SVRG. it's just an arbitrary way to induce the descent lemma

$$0 \le g_i(x - t\nabla g_i(x)) \le g_i(x) - \frac{\|\nabla g_i(x)\|^2}{2L}$$

$$\|\nabla g_i(x)\|^2 = \|\nabla f_i(x) - \nabla f_i(x^*)\|^2$$

Apply descent lemma to g_i

$$0 \le f_i(x) - f_i(x^*) - \nabla f_i(x^*)^T(x - x^*) - \frac{\|\nabla f_i(x) - \nabla f_i(x^*)\|^2}{2L}$$

add up for all $i = 1, 2, \ldots, n$.

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\|\nabla f_i(x) - \nabla f_i(x^*)\|^2}{2L} \le \frac{1}{n}\sum_{i=1}^{n} f_i(x) - f_i(x^*) - \nabla f_i(x^*)^T(x - x^*)$$

$$\le F(x) - F(x^*) - \underbrace{\left(\frac{1}{n}\sum \nabla f_i(x^*)\right)^T(x - x^*)}_{0}$$

≥ 0 using x^* is optimal

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} \|\nabla f_i(x) - \nabla f_i(x^*)\|^2 \le 2L\left(F(x) - F(x^*)\right)} \quad (*)$$

$\boxed{\text{Proof of the Variance Reduction Lemma}}$

$$v_k = \nabla f_{i_k}(x_k) - \nabla f_{i_k}(\tilde{x}) + \nabla F(\tilde{x})$$

$(x+y)^2 \le 2x^2 + 2y^2$

$$\boxed{\|x+y\|^2 \le 2\|x\|^2 + 2\|y\|^2}$$

inequality (a)

$$E[\|v_k\|^2 \mid x_k]$$

(a)

$$\le 2E\|\nabla f_{i_k}(x_{k-1}) - \nabla f_{i_k}(x^*)\|^2 + 2E\|\nabla f_{i_k}(\tilde{x}) - \nabla f_{i_k}(x^*) - \nabla F(\tilde{x})\|^2$$

(b)

$$\le 2E\|\nabla f_{i_k}(x_{k-1}) - \nabla f_{i_k}(x^*)\|^2 + 2E\|\nabla f_{i_k}(\tilde{x}) - \nabla f_{i_k}(x^*)\|^2$$

$$\nabla F(\tilde{x}) = E_{i_k}(\nabla f_{i_k}(\tilde{x})) = E_{i_k}[\nabla f_{i_k}(\tilde{x}) - \nabla f_{i_k}(x^*)]$$

why?

$\nabla F(x^*) = 0$

apply (*) from previous page

$$\le 4L(F(x_k) - F(x^*)) + 4L(F(\tilde{x}) - F(x^*))$$

$\square$

$$\boxed{\begin{aligned} E\|X - EX\|^2 &\le E\|X\|^2 \\ \mathrm{Var}(X) &\le E[X^2] \end{aligned}}$$

inequality (b)