Notes for Lecture 12

Scribe: Tingkai Jia

1 Convergence Analysis of SVRG

Lemma 1.

$$\mathbb{E}[\|\mathbf{g}_s^t\|_2^2] \le 4L[F(\mathbf{x}_s^t) - F(\mathbf{x}^*) + F(\tilde{\mathbf{x}}_s) - F(\mathbf{x}^*)]$$

Proof.

$$\begin{aligned}
&\mathbb{E}[\|\nabla f_{i_t}(\mathbf{x}_s^t) - \nabla f_{i_t}(\tilde{\mathbf{x}}_s) + \nabla F(\tilde{\mathbf{x}}_s)\|_2^2] \\
&= \mathbb{E}[\|\nabla f_{i_t}(\mathbf{x}^{t_s}) - \nabla f_{i_t}(\mathbf{x}^*) - (\nabla f_{i_t}(\tilde{\mathbf{x}}_s) - \nabla f_{i_t}(\mathbf{x}^*) - \nabla F(\tilde{\mathbf{x}}_s))\|_2^2] \\
&\le 2\mathbb{E}[\|\nabla f_{i_t}(\mathbf{x}_s^t) - \nabla f_{i_t}(\mathbf{x}^*)\|_2^2] + 2\mathbb{E}[\|\nabla f_{i_t}(\tilde{\mathbf{x}}_s) - \nabla f_{i_t}(\mathbf{x}^*) - \nabla F(\tilde{\mathbf{x}}_s)\|_2^2] \\
&= 2\mathbb{E}[\|\nabla f_{i_t}(\mathbf{x}_s^t) - \nabla f_{i_t}(\mathbf{x}^*)\|_2^2] + 2\mathbb{E}[\|\nabla f_{i_t}(\tilde{\mathbf{x}}_s) - \nabla f_{i_t}(\mathbf{x}^*) - \mathbb{E}[\nabla f_{i_t}(\tilde{\mathbf{x}}_s) - \nabla f_{i_t}(x^*)]\|_2^2] \\
&\le 2\mathbb{E}[\|\nabla f_{i_t}(\mathbf{x}_s^t) - \nabla f_{i_t}(\mathbf{x}^*)\|_2^2] + 2\mathbb{E}[\|\nabla f_{i_t}(\tilde{\mathbf{x}}_s) - \nabla f_{i_t}(\mathbf{x}^*)\|_2^2] \\
&\le 4L[F(\mathbf{x}_s^t) - F(\mathbf{x}^*) + F(\tilde{\mathbf{x}}_s) - F(\mathbf{x}^*)]
\end{aligned}$$

The last inequality would hold if we could justify:

$$\frac{1}{n}\sum_{i=1}^n \|\nabla f_i(\mathbf{x}) - \nabla f_i(\mathbf{x}^*)\|^2 \le 2L[F(\mathbf{x}) - F(\mathbf{x}^*)],$$

which comes from smoothness and convexity of f_i that

$$\frac{1}{2L}\|\nabla f_i(\mathbf{x}) - \nabla f_i(\mathbf{x}^*)\|_2^2 \le f_i(\mathbf{x}) - f_i(\mathbf{x}^*) - \nabla f_i(\mathbf{x}^*)^\top(\mathbf{x} - \mathbf{x}^*),$$

and summing over all i yield

$$\begin{aligned}
\frac{1}{2L}\|\nabla f_i(\mathbf{x}) - \nabla f_i(\mathbf{x}^*)\|_2^2 &\le nF(\mathbf{x}) - nF(\mathbf{x}^*) - n(\nabla F(\mathbf{x}^*))^\top(\mathbf{x} - \mathbf{x}^*) \\
&= nF(\mathbf{x}) - nF(\mathbf{x}^*).
\end{aligned}$$

□

Theorem 1. *Assume each f_i is convex and L-smooth, and F is μ-strongly convex. Choose m large enough s.t. $\rho = \frac{1}{\mu\eta(1-2L\eta)m} + \frac{2L\eta}{1-2L\eta} < 1$, then*

$$\mathbb{E}[F(\tilde{\mathbf{x}}_s) - F(\mathbf{x}^*)] \le \rho^s[F(\tilde{\mathbf{x}}_0) - F(\mathbf{x}^*)].$$

Proof. Let $\mathbf{g}_s^t = \nabla f_{i_t}(\mathbf{x}_s^t) - \nabla f_{i_t}(\tilde{\mathbf{x}}_s) + \nabla F(\tilde{\mathbf{x}}_s)$ for simplcity. As usual, conditional on everything prior to $\mathbf{x}_s^{t+1}$, one has

$$\begin{aligned}
\mathbb{E}[\|\mathbf{x}_s^{t+1} - \mathbf{x}^*\|_2^2] &= \mathbb{E}[\|\mathbf{x}_s^t - \eta\mathbf{g}_s^t - \mathbf{x}^*\|_2^2] \\
&= \|\mathbf{x}_s^t - \mathbf{x}^*\|_2^2 - 2\eta(\mathbf{x}_s^t - \mathbf{x}^*)^\top\mathbb{E}[\mathbf{g}_s^t] + \eta^2\mathbb{E}[\|\mathbf{g}_s^t\|_2^2] \\
&\le \|\mathbf{x}_s^t - \mathbf{x}^*\|_2^2 - 2\eta(\mathbf{x}_s^t - \mathbf{x}^*)^\top\nabla F(\mathbf{x}_s^t) + \eta^2\mathbb{E}[\|\mathbf{g}_s^t\|_2^2] \\
&\le \|\mathbf{x}_s^t - \mathbf{x}^*\|_2^2 - 2\eta(F(\mathbf{x}_s^t) - F(\mathbf{x}^*)) + \eta^2\mathbb{E}[\|\mathbf{g}_s^t\|_2^2],
\end{aligned}$$

then we try to control $\mathbb{E}[\|\mathbf{g}_s^t\|_2^2]$ using Lemma 1. Thus we have

$$\begin{aligned}
&\mathbb{E}[\|\mathbf{x}_s^{t+1} - \mathbf{x}^*\|_2^2] \\
&\leq \|\mathbf{x}_s^t - \mathbf{x}^*\|_2^2 - 2\eta(F(\mathbf{x}_s^t) - F(\mathbf{x}^*)) + 4L\eta^2[F(\mathbf{x}_s^t) - F(\mathbf{x}^*) + F(\tilde{\mathbf{x}}_s) - F(\mathbf{x}^*)] \\
&= \|\mathbf{x}_s^t - \mathbf{x}^*\|_2^2 - 2\eta(1 - 2L\eta)[F(\mathbf{x}_s^t) - F(\mathbf{x}^*)] + 4L\eta^2[F(\tilde{\mathbf{x}}_s) - F(\mathbf{x}^*)]. \qquad (1)
\end{aligned}$$

Taking expectation w.r.t. all history, we have

$$\begin{aligned}
&2\eta(1 - 2L\eta)m\mathbb{E}\big[F(\tilde{\mathbf{x}}_{s+1}) - F(\mathbf{x}^*)\big] \\
&= 2\eta(1 - 2L\eta)\sum_{t=0}^{m-1}\mathbb{E}\big[F(\mathbf{x}_s^t) - F(\mathbf{x}^*)\big] \\
&\leq \mathbb{E}\big[\|\mathbf{x}_{s+1}^m - \mathbf{x}^*\|_2^2\big] + 2\eta(1 - 2L\eta)\sum_{t=0}^{m-1}\mathbb{E}\big[F(\mathbf{x}_s^t) - F(\mathbf{x}^*)\big],
\end{aligned}$$

then we apply (1) recursively to obtain

$$\begin{aligned}
&2\eta(1 - 2L\eta)m\mathbb{E}\big[F(\tilde{\mathbf{x}}_{s+1}) - F(\mathbf{x}^*)\big] \\
&\leq \mathbb{E}\big[\|\mathbf{x}_{s+1}^0 - \mathbf{x}^*\|_2^2\big] + 4Lm\eta^2\mathbb{E}\big[F(\tilde{\mathbf{x}}_s) - F(\mathbf{x}^*)\big] \\
&= \mathbb{E}\big[\|\tilde{\mathbf{x}}_s - \mathbf{x}^*\|_2^2\big] + 4Lm\eta^2\mathbb{E}\big[F(\tilde{\mathbf{x}}_s) - F(\mathbf{x}^*)\big] \\
&\leq \frac{2}{\mu}\mathbb{E}\big[F(\tilde{\mathbf{x}}_s) - F(\mathbf{x}^*)\big] + 4Lm\eta^2\mathbb{E}\big[F(\tilde{\mathbf{x}}_s) - F(\mathbf{x}^*)\big] \\
&= \left(\frac{2}{\mu} + 4Lm\eta^2\right)\mathbb{E}\big[F(\tilde{\mathbf{x}}_s) - F(\mathbf{x}^*)\big].
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathbb{E}[F(\tilde{\mathbf{x}}_{s+1}) - F(\mathbf{x}^*)] &\leq \frac{\frac{2}{\mu} + 4Lm\eta^2}{2\eta(1 - 2L\eta)m}\mathbb{E}[F(\tilde{\mathbf{x}}_s) - F(\mathbf{x}^*)] \\
&= \left(\frac{1}{\mu\eta(1 - 2L\eta)m} + \frac{2L\eta}{1 - 2L\eta}\right)\mathbb{E}[F(\tilde{\mathbf{x}}_s) - F(\mathbf{x}^*)].
\end{aligned}$$

Finally we apply this bound recursively to finish the proof. □

2 Nonconvex Problems

Theorem 2. *Let f be L-smooth and $\eta_k \equiv \eta = 1/L$. Assume t is even. GD obeys*

$$\min_{0 \leq k < t} \|\nabla f(\mathbf{x}_k)\|_2 \leq \sqrt{\frac{2L(f(\mathbf{x}_0) - f(\mathbf{x}^*))}{t}}.$$

Proof. From the smoothness assumption,

$$\begin{aligned}
f(\mathbf{x}_{t+1}) - f(\mathbf{x}_t) &\leq \nabla f(\mathbf{x}_t)^\top(\mathbf{x}_{t+1} - \mathbf{x}_t) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 \\
&= -\eta\|\nabla f(\mathbf{x}_t)\|_2^2 + \frac{\eta^2 L}{2}\|\nabla f(\mathbf{x}_t)\|_2^2 \\
&= -\frac{1}{2L}\|\nabla f(\mathbf{x}_t)\|_2^2.
\end{aligned}$$

Apply it recursively, we have

$$\frac{1}{2L}\sum_{k=0}^{t-1}\|\nabla f(\mathbf{x}_t)\|_2^2 \le \sum_{k=0}^{t-1}(f(\mathbf{x}_0)-f(\mathbf{x}_t)) = f(\mathbf{x}_0)-f(\mathbf{x}_t)$$
$$\le f(\mathbf{x}_0)-f(\mathbf{x}^*),$$

which means

$$\min_{0\le k<t}\|\nabla f(\mathbf{x}_k)\|_2 \le \sqrt{\frac{2L(f(\mathbf{x}_0)-f(\mathbf{x}^*))}{t}}.$$

□