# HW3-sol-Optimization in machine learning

Tian QIU

August 28, 2024

## 1 SGD for Smooth and Strongly Convex Functions

### 1.1 a

By $l$-smoothness:

$$\begin{aligned}\mathbb{E}\left[f(x_{s+1}) - f(x_s)\right] &\leq \mathbb{E}\left[\langle \nabla f(x_s), x_{s+1} - x_s \rangle + \frac{l}{2}\|x_{s+1} - x_s\|^2\right] \\ &= \mathbb{E}\left[\langle \nabla f(x_s), -\eta g(x_s) \rangle\right] + \frac{\eta^2 l}{2}\mathbb{E}\left[\|g(x_s)\|^2\right] \\ &= -\eta \nabla f(x_s)^T \mathbb{E}[g(x_s)] + \frac{\eta^2 l}{2}\mathbb{E}\left[\|g(x_s)\|^2\right].\end{aligned}$$

By the property and assumption of variance:

$$\mathbb{E}\left[\|g(x_s)\|^2\right] - \|\mathbb{E}[\nabla f(x_s)]\|^2 \leq \sigma^2.$$

We have:

$$\begin{aligned}\mathbb{E}\left[f(x_{s+1}) - f(x_s)\right] &\leq -\eta \|\nabla f(x_s)\|^2 + \frac{\eta^2 l}{2}\left(\sigma^2 + \|\nabla f(x_s)\|^2\right) \\ &= \left(\frac{\eta^2 l}{2} - \eta\right)\|\nabla f(x_s)\|^2 + \frac{\eta^2 l \sigma^2}{2}.\end{aligned}$$

So:

$$\mathbb{E}\left[f(x_{s+1}) - f(x_s)\right] \leq \left(\frac{\eta^2 l}{2} - \eta\right)\|\nabla f(x_s)\|^2 + \frac{\eta^2 l \sigma^2}{2}. \tag{1}$$

By Strongly Convexity:

$$\begin{aligned}
\mathbb{E}(f(x_s) - f(x^\star)) &\leq \mathbb{E}\left[\langle \nabla f(x_s), x_s - x^\star\rangle - \frac{\alpha}{2}\|x_s - x^\star\|^2\right] \\
&= \mathbb{E}\left[\langle \mathbb{E}[g(x_s)], x_s - x^\star\rangle\right] - \frac{\alpha}{2}\mathbb{E}\left[\|x_s - x^\star\|^2\right] \\
&= \mathbb{E}\left[\left\langle \frac{1}{\eta}(x_s - x_{s+1}), x_s - x^\star \right\rangle\right] - \frac{\alpha}{2}\mathbb{E}\left[\|x_s - x^\star\|^2\right] \\
&= \frac{1}{2\eta}\mathbb{E}\left[\|x_s - x_{s+1}\|^2 + \|x_s - x^\star\|^2 - \|x_{s+1} - x^\star\|^2\right] - \frac{\alpha}{2}\mathbb{E}\left[\|x_s - x^\star\|^2\right] \\
&= \left(-\frac{\alpha}{2} + \frac{1}{2\eta}\right)\mathbb{E}\left[\|x_s - x^\star\|^2\right] - \frac{1}{2\eta}\mathbb{E}\left[\|x_{s+1} - x^\star\|^2\right] + \frac{\eta}{2}\mathbb{E}\left[\|g(x_s)\|^2\right] \\
&\leq \left(-\frac{\alpha}{2} + \frac{1}{2\eta}\right)\mathbb{E}\left[\|x_s - x^\star\|^2\right] - \frac{1}{2\eta}\mathbb{E}\left[\|x_{s+1} - x^\star\|^2\right] + \frac{\eta}{2}\left(\sigma^2 + \|\nabla f(x_s)\|^2\right).
\end{aligned}$$

So:

$$\mathbb{E}\left[f(x_s) - f(x^\star)\right] \leq \left(-\frac{\alpha}{2} + \frac{1}{2\eta}\right)\mathbb{E}\left[\|x_s - x^\star\|^2\right] - \frac{1}{2\eta}\mathbb{E}\left[\|x_{s+1} - x^\star\|^2\right] + \frac{\eta}{2}\left(\sigma^2 + \|\nabla f(x_s)\|^2\right). \tag{2}$$

Add up (1) and (2):

$$\begin{aligned}
\mathbb{E}\left[f(x_{s+1}) - f(x^\star)\right] &\leq \left(-\frac{\alpha}{2} + \frac{1}{2\eta}\right)\mathbb{E}\left[\|x_s - x^\star\|^2\right] - \frac{1}{2\eta}\mathbb{E}\left[\|x_{s+1} - x^\star\|^2\right] \\
&\quad + \left(\frac{\eta^2 l\sigma^2}{2} + \frac{\eta\sigma^2}{2}\right) + \left(\frac{\eta^2 l}{2} - \frac{\eta}{2}\right)\|\nabla f(x_s)\|^2 \\
&\leq \left(-\frac{\alpha}{2} + \frac{1}{2\eta}\right)\mathbb{E}\left[\|x_s - x^\star\|^2\right] - \frac{1}{2\eta}\mathbb{E}\left[\|x_{s+1} - x^\star\|^2\right] + \eta\sigma^2
\end{aligned}$$

Do some simplification and we gain the final ans:

$$\mathbb{E}\|x_{s+1} - x^\star\|^2 \leq (1 - \eta\alpha)\mathbb{E}\|x_s - x^\star\|^2 - 2\eta\mathbb{E}\left[f(x_{s+1}) - f(x^\star)\right] + 2\eta^2\sigma^2 \tag{3}$$

### 1.2 b

Do telescope of the inequality in (b):

$$\mathbb{E}\sum_{s=2}^{t+1}\lambda_s(f(x_s)-f(x^\star)) \le \sum_{s=2}^{t+1}\frac{\lambda_s}{2\eta}\mathbb{E}[(1-\alpha\eta)||x_{s-1}-x^\star||^2-||x_s-x^\star||^2+2\eta^2\sigma^2]$$

Plug in the value of $\lambda_s$

$$\le \eta\sigma^2+\frac{1}{2\eta\sum_{s=2}^{t+1}(1-\eta\alpha)^{t+1-s}}\sum_{s=2}^{t+1}[(1-\alpha\eta)^{t+2-s}||x_{s-1}-x^\star||^2-(1-\alpha\eta)^{t+1-s}||x_s-x^\star||^2]$$

$$=\eta\sigma^2+\frac{1}{2\eta\sum_{s=2}^{t+1}(1-\eta\alpha)^{t+1-s}}[(1-\alpha\eta)^t||x_1-x^\star||^2-(1-\eta\alpha)^0||x_{t+1}-x^\star||^2]$$

$$\le \eta\sigma^2+\frac{1}{2\eta\sum_{s=2}^{t+1}(1-\eta\alpha)^{t+1-s}}e^{-\eta\alpha t}||x_1-x^\star||^2$$

$$\le \eta\sigma^2+\frac{e^{-\eta\alpha t}||x_1-x^\star||^2}{2\eta}$$

We can easily know that $\eta\alpha\in(0,1)$.The last step relax the denominator to 1.
We prove that:

$$\mathbb{E}\sum_{s=2}^{t+1}\lambda_s(f(x_s)-f(x^\star)) \le \eta\sigma^2+\frac{e^{-\eta\alpha t}||x_1-x^\star||^2}{2\eta} \tag{4}$$

### 1.3 c

Use convexity?or use strong cvx? simply relaxation in parameter may not work Do the same operations just like in note 11,consider the dominator in small t and large t. Plug in different $\eta$ in the two separate terms.But there still exists a strange constant,which doesn't cause too much trouble.

## 2 Catalyst Acceleration for Finite-sum Problems