

# Notes for Lecture 9

Scribe: Tingkai Jia

## 1 Convergence of Proximal Gradient Descent

**Lemma 1.** Let  $\mathbf{y}^+ = \text{prox}_{\frac{1}{L}h}(y - \frac{1}{L}\nabla f(\mathbf{y}))$ , then

$$F(\mathbf{y}^+) - F(\mathbf{x}) \leq \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 - \frac{L}{2}\|\mathbf{x} - \mathbf{y}^+\|_2^2 - g(\mathbf{x}, \mathbf{y})$$

where  $g(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - f(\mathbf{y}) - \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$ .

*Proof.* Define  $\phi(\mathbf{z}) = f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{z} - \mathbf{y} \rangle + \frac{L}{2}\|\mathbf{z} - \mathbf{y}\|_2^2 + h(\mathbf{z})$ . It is true that  $\mathbf{y}^+ = \arg \min_{\mathbf{z}} \phi(\mathbf{z})$  and  $\phi(\mathbf{z})$  is  $L$ -strongly convex, which means

$$\phi(\mathbf{x}) \geq \phi(\mathbf{y}) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|_2^2.$$

From smoothness,

$$\begin{aligned} \phi(\mathbf{y}^+) &= f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{y}^+ - \mathbf{y} \rangle + \frac{L}{2}\|\mathbf{y}^+ - \mathbf{y}\|_2^2 + h(\mathbf{y}^+) \\ &\geq f(\mathbf{y}^+) + h(\mathbf{y}^+) = F(\mathbf{y}^+). \end{aligned}$$

Now we get

$$\phi(\mathbf{x}) \geq \phi(\mathbf{y}) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|_2^2$$

which together with the definition of  $\phi(\mathbf{x})$  gives

$$\begin{aligned} f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + h(\mathbf{x}) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 &\geq F(\mathbf{y}^+) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}^+\|_2^2 \\ f(\mathbf{x}) + h(\mathbf{x}) - g(\mathbf{x}, \mathbf{y}) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 &\geq F(\mathbf{y}^+) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}^+\|_2^2 \\ F(\mathbf{x}) - g(\mathbf{x}, \mathbf{y}) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 &\geq F(\mathbf{y}^+) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}^+\|_2^2. \end{aligned}$$

□

**Theorem 1** (PGD for convex problems). Suppose  $f$  is convex and  $L$ -smooth. If  $\eta_t \equiv 1/L$ , then

$$F(\mathbf{x}_t) - F(\mathbf{x}_*) \leq \frac{L\|\mathbf{x}_0 - \mathbf{x}_*\|_2^2}{2t}.$$

*Proof.* With Lemma 1 in mind, set  $\mathbf{x} = \mathbf{x}_*$ ,  $\mathbf{y} = \mathbf{x}_t$  to obtain

$$\begin{aligned} F(\mathbf{x}_{t+1}) - F(\mathbf{x}_*) &\leq \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}_*\|_2^2 - \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2^2 - g(\mathbf{x}_*, \mathbf{x}_t) \\ &\leq \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}_*\|_2^2 - \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2^2. \end{aligned}$$

Apply it recursively and add up all inequalities to get

$$\sum_{k=0}^{t-1} (F(\mathbf{x}_{k+1}) - F(\mathbf{x}_*)) \leq \frac{L}{2}\|\mathbf{x}_0 - \mathbf{x}_*\|_2^2 - \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}_*\|_2^2,$$

with  $t(F(\mathbf{x}_t) - F(\mathbf{x}_*)) \leq \sum_{k=0}^{t-1} (F(\mathbf{x}_{k+1}) - F(\mathbf{x}_*))$ , we finally get

$$F(\mathbf{x}_t) - F(\mathbf{x}_*) \leq \frac{L\|\mathbf{x}_0 - \mathbf{x}_*\|_2^2}{2t}.$$

□

**Theorem 2** (PGD for strongly convex problems). *Suppose  $f$  is  $\mu$ -strongly convex and  $L$ -smooth. if  $\eta_t \equiv 1/L$ , then*

$$\|\mathbf{x}_t - \mathbf{x}_*\|_2^2 \leq \left(1 - \frac{\mu}{L}\right)^t \|\mathbf{x}_0 - \mathbf{x}_*\|_2^2.$$

*Proof.* With Lemma 1 in mind, set  $\mathbf{x} = \mathbf{x}_*$ ,  $\mathbf{y} = \mathbf{x}_t$  to obtain

$$\begin{aligned} F(\mathbf{x}_{t+1}) - F(\mathbf{x}_*) &\leq \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}_*\|_2^2 - \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2^2 - g(\mathbf{x}_*, \mathbf{x}_t) \\ &\leq \frac{L-\mu}{2}\|\mathbf{x}_t - \mathbf{x}_*\|_2^2 - \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2^2. \end{aligned}$$

This taken collectively with  $F(\mathbf{x}_{t+1}) - F(\mathbf{x}_*) \geq 0$  yields

$$\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2^2 \leq \left(1 - \frac{\mu}{L}\right) \|\mathbf{x}_t - \mathbf{x}_*\|_2^2.$$

Applying it recursively concludes the proof.

□