

# Notes for Lecture 6

Scribe: Tingkai Jia

## 1 Projected Subgradient Descent with Polyak's Stepsize

**Lemma 1.** *Projected subgradient update obeys*

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \leq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \eta_t^2 \|\mathbf{g}_t\|_2^2 - 2\eta_t(f(\mathbf{x}_t) - f^*).$$

*Proof.* It follows that

$$\begin{aligned} \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 &= \|\mathcal{P}_C(\mathbf{x}_t - \eta_t \mathbf{g}_t) - \mathbf{x}^*\|_2^2 \\ &\leq \|\mathbf{x}_t - \eta_t \mathbf{g}_t - \mathbf{x}^*\|_2^2 \\ &= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \eta_t^2 \|\mathbf{g}_t\|_2^2 - 2\eta_t \langle \mathbf{g}_t, \mathbf{x}_t - \mathbf{x}^* \rangle \\ &\leq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \eta_t^2 \|\mathbf{g}_t\|_2^2 - 2\eta_t(f(\mathbf{x}_t) - f^*), \end{aligned}$$

where the last inequality uses the convexity

$$f^* \geq f(\mathbf{x}_t) + \langle \mathbf{g}_t, \mathbf{x}^* - \mathbf{x}_t \rangle.$$

□

**Definition 1** (polyak's stepsize). *By treating the RHS of Inequality in Lemma 1 as a quadratic function with respect to  $\eta_t$ , we obtain a step size by minimizing this function*

$$\eta_t = \frac{f(\mathbf{x}_t) - f^*}{\|\mathbf{g}_t\|_2^2}.$$

### 1.1 Example: Projection onto Intersection of Convex Sets

**Example 1.** *Let  $\mathcal{C}_1, \mathcal{C}_2$  be closed convex sets and suppose  $\mathcal{C}_1 \cap \mathcal{C}_2 \neq \emptyset$ ,*

$$\text{minimize}_x \quad \max\{\text{dist}_{\mathcal{C}_1}(\mathbf{x}), \text{dist}_{\mathcal{C}_2}(\mathbf{x})\}$$

*where  $\text{dist}_C(\mathbf{x}) := \min_{\mathbf{z} \in C} \|\mathbf{x} - \mathbf{z}\|_2$ .*

For this problem, the subgradient method of polyak's stepsize will act as

$$\mathbf{x}_{t+1} = \mathcal{P}_{\mathcal{C}_1}(\mathbf{x}_t), \quad \mathbf{x}_{t+2} = \mathcal{P}_{\mathcal{C}_2}(\mathbf{x}_{t+1}).$$

*Proof.* First we consider the subgradient, it follows that

$$\mathbf{g}_t \in \partial \text{dist}_{\mathcal{C}_i}(\mathbf{x}_t)$$

where  $i = \arg \max_{i=1,2} \text{dist}_{\mathcal{C}_i}(\mathbf{x}_t)$ . If  $\text{dist}_{\mathcal{C}_i}(\mathbf{x}_t) \neq 0$ , we have

$$\mathbf{g}_t = \nabla \text{dist}_{\mathcal{C}_i}(\mathbf{x}_t) = \frac{\mathbf{x}_t - \mathcal{P}_{\mathcal{C}_i}(\mathbf{x}_t)}{\|\mathbf{x}_t - \mathcal{P}_{\mathcal{C}_i}(\mathbf{x}_t)\|_2}.$$

Then the polyak's stepsize is shown as

$$\eta_t = \frac{\text{dist}_{C_i}(\mathbf{x}_t) - 0}{\|\mathbf{g}_t\|_2^2} = \|\mathbf{x}_t - \mathcal{P}_{C_i}(\mathbf{x}_t)\|_2.$$

Adopting polyak's stepsize, we arrive at

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \eta_t \mathbf{g}_t = \mathbf{x}_t - \|\mathbf{x}_t - \mathcal{P}_{C_i}(\mathbf{x}_t)\|_2 \frac{\mathbf{x}_t - \mathcal{P}_{C_i}(\mathbf{x}_t)}{\|\mathbf{x}_t - \mathcal{P}_{C_i}(\mathbf{x}_t)\|_2} = \mathcal{P}_{C_i}(\mathbf{x}_t).$$

□

## 1.2 Convergence Rate with Polyak's Stepsize

**Theorem 1.** Suppose  $f$  is convex and  $L$ -Lipschitz continuous. Then the projected subgradient method with Polyak's stepsize obeys

$$f_{\text{best},t} - f^* \leq \frac{L\|\mathbf{x}_0 - \mathbf{x}^*\|_2}{\sqrt{t+1}}$$

*Proof.* With Lemma 1 and substituting  $\eta_t$ , we obtain

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

Applying it resursively, we get

$$\begin{aligned} \sum_{k=0}^t (f(\mathbf{x}_k) - f^*)^2 &\leq [\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2] L^2 \\ (t+1)(f_{\text{best},t} - f^*)^2 &\leq [\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2] L^2 \\ (f_{\text{best},t} - f^*)^2 &\leq \frac{L^2\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2}{t+1} \end{aligned}$$

which completes the proof. □

## 2 Projected Subgradient Descent with Other Stepsizes

**Lemma 2.** Suppose  $f$  is convex and  $L$ -Lipschitz continuous. Then the projected subgradient update obeys

$$f_{\text{best},t} - f^* \leq \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + L^2 \sum_{i=0}^t \eta_i^2}{2 \sum_{i=0}^t \eta_i}.$$

*Proof.* Using Lemma 1 and summing it recursively, we obtain

$$\begin{aligned} 2 \sum_{i=0}^t \eta_i (f(\mathbf{x}_i) - f^*) &\leq \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 + \sum_{i=0}^t \eta_i^2 \|\mathbf{g}_i\|_2^2 \\ 2(f_{\text{best},t} - f^*) \sum_{i=0}^t \eta_i &\leq \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 + \sum_{i=0}^t \eta_i^2 \|\mathbf{g}_i\|_2^2 \\ 2(f_{\text{best},t} - f^*) \sum_{i=0}^t \eta_i &\leq \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 + \sum_{i=0}^t \eta_i^2 L^2 \\ f_{\text{best},t} - f^* &\leq \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + L^2 \sum_{i=0}^t \eta_i^2}{2 \sum_{i=0}^t \eta_i}, \end{aligned}$$

thus we complete the proof. □

## 2.1 Convergence with $1/\sqrt{t+1}$ Stepsize

Considering the inequality in Lemma 2, we aim to make its RHS approach zero as the subgradient method update, which means  $\sum_{i=0}^t \eta_i^2 < \infty$  and  $\sum_{i=0}^t \eta_i \rightarrow \infty$ . Now we can consider  $\eta_t = \frac{1}{\sqrt{t+1}}$ .

**Theorem 2.** *Suppose  $f$  is convex and  $L$ -Lipschitz continuous. Then the projected subgradient method with  $\eta_t = \frac{1}{\sqrt{t+1}}$  obeys*

$$f_{\text{best},t} - f^* \lesssim \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + L^2}{\sqrt{t}}.$$

*Proof.* With the fact that  $\frac{2}{\sqrt{t+1} + \sqrt{t+2}} \leq \frac{1}{\sqrt{t+1}} \leq \frac{2}{\sqrt{t+1} + \sqrt{t+1}}$ , we can get a lower and upper bound in  $\sum_{i=0}^t \eta_i$ ,

$$\begin{aligned} \sum_{k=0}^t \frac{2}{\sqrt{k+1} + \sqrt{k+2}} &\leq \sum_{i=0}^t \eta_i \leq \sum_{k=0}^t \frac{2}{\sqrt{k} + \sqrt{k+1}} \\ 2 \sum_{k=0}^t (\sqrt{k+2} - \sqrt{k+1}) &\leq \sum_{i=0}^t \eta_i \leq 2 \sum_{k=0}^t (\sqrt{k+1} - \sqrt{k}) \\ 2(\sqrt{t+2} - 1) &\leq \sum_{i=0}^t \eta_i \leq 2(\sqrt{t+1}), \end{aligned}$$

and consider sequence  $\frac{1}{k}$ , the upper bound of its sum is  $\log t + 1$ , now we get

$$\begin{aligned} f_{\text{best},t} - f^* &\leq \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + L^2 \sum_{i=0}^t \eta_i^2}{2 \sum_{i=0}^t \eta_i} \\ &\leq \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + L^2 (\log t + 1)}{4(\sqrt{t+2} - 1)} \\ &\lesssim \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + L^2 \log t}{\sqrt{t}}. \end{aligned}$$

Through this approach, we find that the convergence contains a  $\log t$  in the numerator. now, we attempt to eliminate this term.

Note that  $\sum_{k=\lceil \frac{t}{2} \rceil}^t \frac{1}{k} \approx \log t - \log \lceil \frac{t}{2} \rceil \leq \log 3$  and  $\sum_{k=\lceil \frac{t}{2} \rceil}^t \frac{1}{\sqrt{k}} \approx 2\sqrt{t} - 2\sqrt{\lceil \frac{t}{2} \rceil} = (2 - \sqrt{2})\sqrt{t}$ . Thus, we modify the inequality in Lemma 2 and obtain

$$\begin{aligned} 2 \sum_{i=\lceil \frac{t}{2} \rceil}^t \eta_i (f(\mathbf{x}_i) - f^*) &\leq \|\mathbf{x}_{\lceil \frac{t}{2} \rceil} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 + L^2 \sum_{i=\lceil \frac{t}{2} \rceil}^t \eta_i^2 \\ 2(f_{\text{best},t} - f^*) \sum_{i=\lceil \frac{t}{2} \rceil}^t \eta_i &\leq \|\mathbf{x}_{\lceil \frac{t}{2} \rceil} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 + L^2 \sum_{i=\lceil \frac{t}{2} \rceil}^t \eta_i^2 \\ 2(f_{\text{best},t} - f^*) \sum_{i=\lceil \frac{t}{2} \rceil}^t \eta_i &\leq \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + L^2 \sum_{i=\lceil \frac{t}{2} \rceil}^t \eta_i^2 \\ f_{\text{best},t} - f^* &\leq \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + L^2 \sum_{i=\lceil \frac{t}{2} \rceil}^t \eta_i^2}{2 \sum_{i=\lceil \frac{t}{2} \rceil}^t \eta_i} \\ f_{\text{best},t} - f^* &\leq \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + L^2 \log 3}{2(2 - \sqrt{2})\sqrt{t}} \\ f_{\text{best},t} - f^* &\lesssim \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + L^2}{\sqrt{t}}, \end{aligned}$$

which we finish the proof. □

### 3 Strongly Convex and Lipschitz Problems

**Theorem 3.** Let  $f$  be  $\mu$ -strongly convex and  $L$ -Lipschitz continuous over  $\mathcal{C}$ . If  $\eta_t \equiv \eta = \frac{2}{\mu(t+1)}$ , then

$$f_{\text{best},t} - f^* \leq \frac{2L^2}{\mu(t+1)}$$

*Proof.* Consider strongly convex situation in Lemma 1, we have

$$\begin{aligned} \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 &= \|\mathcal{P}_{\mathcal{C}}(\mathbf{x}_t - \eta_t \mathbf{g}_t) - \mathbf{x}^*\|_2^2 \\ &\leq \|\mathbf{x}_t - \eta_t \mathbf{g}_t - \mathbf{x}^*\|_2^2 \\ &= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \eta_t^2 \|\mathbf{g}_t\|_2^2 - 2\eta_t \langle \mathbf{g}_t, \mathbf{x}_t - \mathbf{x}^* \rangle \\ &\leq (1 - \mu\eta_t) \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \eta_t^2 \|\mathbf{g}_t\|_2^2 - 2\eta_t (f(\mathbf{x}_t) - f^*). \end{aligned}$$

For the last inequality, please see problem 1 in the homework. Since  $\eta_t \equiv \eta = \frac{2}{\mu(t+1)}$ , we have

$$\begin{aligned} f(\mathbf{x}_t) - f^* &\leq \frac{\mu(t-1)}{4} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 - \frac{\mu(t+1)}{4} \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 + \frac{1}{\mu(t+1)} \|\mathbf{g}_t\|_2^2 \\ t(f(\mathbf{x}_t) - f^*) &\leq \frac{\mu t(t-1)}{4} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 - \frac{\mu t(t+1)}{4} \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 + \frac{t}{\mu(t+1)} \|\mathbf{g}_t\|_2^2 \\ &\leq \frac{\mu t(t-1)}{4} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 - \frac{\mu t(t+1)}{4} \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 + \frac{1}{\mu} \|\mathbf{g}_t\|_2^2. \end{aligned}$$

Summing over all iterations before  $t$ , we get

$$\sum_{k=0}^t k(f(\mathbf{x}_k) - f^*) \leq 0 - \frac{\mu t(t+1)}{4} \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 + \frac{1}{\mu} \sum_{i=0}^t \|\mathbf{g}_i\|_2^2 \leq \frac{tL^2}{\mu},$$

which means

$$f_{\text{best},t} - f^* \leq \frac{tL^2}{\mu \sum_{k=0}^t k} \leq \frac{2L^2}{\mu(t+1)}.$$

Thus we finish the proof. □