Notes for Lecture 6

Scribe: Tingkai Jia

1 Some Properties of Projection

Property 1. *Suppose f is a convex function and $\mathcal{C}$ is a closed convex set. Let*

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}) \quad \textit{and} \quad \mathbf{x}_* = \arg\min_{\mathbf{x}\in\mathcal{C}} f(\mathbf{x})$$

It is possible that

$$\mathbf{x}_* \neq \mathcal{P}_{\mathcal{C}}(\hat{\mathbf{x}}).$$

Consider a counterpart example, $\mathbf{x} \in \mathbb{R}^2, f(\mathbf{x}) = x_1^2 + 5x_2^2, \mathcal{C} = \{\mathbf{x}|x_1 + x_2 = 1\}$. The minimizer of f is $\mathbf{x}_* = 0$, and $\mathcal{P}_{\mathcal{C}}(\mathbf{x}_*) = (\frac{1}{2}, \frac{1}{2})^\top$, then we see

$$f(\mathcal{P}_{\mathcal{C}}(\mathbf{x}_*)) = \frac{3}{2} < f(1,0) = 1,$$

which means $\mathcal{P}_{\mathcal{C}}(\mathbf{x}_*)$ is not minimizer in constrained set.

Property 2. *Let $\mathcal{C} \in \mathbb{R}^d$ be closed and convex, $\mathbf{z} \in \mathcal{C}$, $\mathbf{x} \in \mathbb{R}^d$. Then*

$$\langle \mathbf{x} - \mathcal{P}_{\mathcal{C}}(\mathbf{x}), \mathbf{z} - \mathcal{P}_{\mathcal{C}}(\mathbf{x})\rangle \leq 0.$$

Proof. Let $\forall \mathbf{y} \in \mathcal{C}, g(\mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2^2$, then $\mathcal{P}_{\mathcal{C}}(\mathbf{x}) = \min_{\mathbf{y}} g(\mathbf{y})$. Consider that $\mathcal{P}_{\mathcal{C}}(\mathbf{x})$ is the minimizer of g in $\mathcal{C}$, with $\forall \mathbf{z} \in \mathcal{C}$, we have

$$\langle \nabla g(\mathcal{P}_{\mathcal{C}}(\mathbf{x})), \mathbf{z} - \mathcal{P}_{\mathcal{C}}(\mathbf{x})\rangle \geq 0,$$

which means

$$\langle -2(\mathbf{x} - \mathcal{P}_{\mathcal{C}}(\mathbf{x})), \mathbf{z} - \mathcal{P}_{\mathcal{C}}(\mathbf{x})\rangle \geq 0.$$

Thus we finish the proof. □

Property 3. *Let $\mathcal{C} \in \mathbb{R}^d$ be closed and convex. For any $\mathbf{x}, \mathbf{z} \in \mathbb{R}^d$, we have*

$$\|\mathcal{P}_{\mathcal{C}}(\mathbf{x}) - \mathcal{P}_{\mathcal{C}}(\mathbf{z})\|_2 \leq \|\mathbf{x} - \mathbf{z}\|_2.$$

Proof. With Property 2 mentioned above, we first have

$$\langle \mathbf{x} - \mathcal{P}_{\mathcal{C}}(\mathbf{x}), \mathcal{P}_{\mathcal{C}}(\mathbf{z}) - \mathcal{P}_{\mathcal{C}}(\mathbf{x})\rangle \leq 0 \quad \text{and} \quad \langle \mathbf{z} - \mathcal{P}_{\mathcal{C}}(\mathbf{z}), \mathcal{P}_{\mathcal{C}}(\mathbf{x}) - \mathcal{P}_{\mathcal{C}}(\mathbf{z})\rangle \leq 0.$$

Combining the two inequalities

$$\langle \mathbf{x} - \mathcal{P}_{\mathcal{C}}(\mathbf{x}), \mathcal{P}_{\mathcal{C}}(\mathbf{z}) - \mathcal{P}_{\mathcal{C}}(\mathbf{x})\rangle \leq \langle \mathcal{P}_{\mathcal{C}}(\mathbf{z}) - \mathbf{z}, \mathcal{P}_{\mathcal{C}}(\mathbf{x}) - \mathcal{P}_{\mathcal{C}}(\mathbf{z})\rangle,$$

and rearranging the terms, we get

$$\begin{aligned}\|\mathcal{P}_{\mathcal{C}}(\mathbf{x}) - \mathcal{P}_{\mathcal{C}}(\mathbf{z})\|_2^2 &\leq \langle \mathbf{x} - \mathbf{z}, \mathcal{P}_{\mathcal{C}}(\mathbf{x}) - \mathcal{P}_{\mathcal{C}}(\mathbf{z})\rangle \\ &\leq \|\mathbf{x} - \mathbf{z}\|_2 \|\mathcal{P}_{\mathcal{C}}(\mathbf{x}) - \mathcal{P}_{\mathcal{C}}(\mathbf{z})\|_2,\end{aligned}$$

thus we finish the proof. □

2 Strongly Convex and Smooth Constrained Optimization

Lemma 1. *Suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is μ-strongly convex and L-smooth. Thus*

$$\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq \frac{\mu L}{\mu + L} \|\mathbf{x} - \mathbf{y}\|_2^2 + \frac{1}{\mu + L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2.$$

Theorem 1. *Let f be L-smooth and μ-strongly convex. If $\eta_t \equiv \eta = \frac{2}{\mu + L}$, then PGD obeys*

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

Proof. Let $S(\mathbf{x}) = \mathbf{x} - \eta \nabla f(\mathbf{x})$, we have

$$\|S(\mathbf{y}) - S(\mathbf{x})\|_2^2 = \|\mathbf{y} - \mathbf{x} + \eta(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))\|_2^2,$$

then expand the equation and substitute Lemma 1, we get

$$\begin{aligned}
\|S(\mathbf{y}) - S(\mathbf{x})\|_2^2 &= \|\mathbf{y} - \mathbf{x}\|_2^2 + \eta^2 \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2 - 2\eta \langle \mathbf{x} - \mathbf{y}, \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}) \rangle \\
&\leq \left(\eta^2 - \frac{2\eta}{\mu + L} \right) \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2 + \left(1 - \frac{2\eta\mu L}{\mu + L} \right) \|\mathbf{x} - \mathbf{y}\|_2^2 \\
&= \left(\frac{\mu - L}{\mu + L} \right)^2 \|\mathbf{x} - \mathbf{y}\|_2^2,
\end{aligned}$$

which means S is a contraction mapping. Then we let $T(\mathbf{x}) = \mathcal{P}_{\mathcal{C}}(S(\mathbf{x}))$, which immediately follows that

$$\|T(\mathbf{x}) - T(\mathbf{y})\|_2 \leq \|S(\mathbf{x}) - S(\mathbf{y})\|_2 \leq \left(\frac{\mu - L}{\mu + L} \right) \|\mathbf{x} - \mathbf{y}\|_2$$

Therefore, projected gradient descent algorithm maintains the same convergence rate for constrained problems as for unconstrained problems. □

3 Convex and Smooth Constrained Optimization

Lemma 2. *Let $\mathcal{C} \in \mathbb{R}^d$ be closed and convex, for any $\mathbf{x}, \mathbf{y} \in \mathcal{C}$, $\mathbf{x}^+ = \mathcal{P}_{\mathcal{C}}(\mathbf{x} - \frac{1}{L} \nabla f(\mathbf{x}))$ and $g_{\mathcal{C}}(\mathbf{x}) = L(\mathbf{x} - \mathbf{x}^+)$. Then*

$$\nabla f(\mathbf{x})^\top (\mathbf{x}^+ - \mathbf{y}) \leq g_{\mathcal{C}}(\mathbf{x})^\top (\mathbf{x}^+ - \mathbf{y}).$$

Proof. By applying Property 2, we immediately obtain

$$\left\langle \mathbf{x} - \frac{1}{L} \nabla f(\mathbf{x}) - \mathbf{x}^+, \mathbf{y} - \mathbf{x}^+ \right\rangle \leq 0$$

$$\left\langle g_{\mathcal{C}}(\mathbf{x}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}^+ \right\rangle \leq 0.$$

Thus we finish the proof. □

Lemma 3. *Suppose f is convex and L-smooth. For any $\mathbf{x}, \mathbf{y} \in \mathcal{C}$, let $\mathbf{x}^+ = \mathcal{P}_{\mathcal{C}}(\mathbf{x} - \frac{1}{L} \nabla f(\mathbf{x}))$ and $g_{\mathcal{C}}(\mathbf{x}) = L(\mathbf{x} - \mathbf{x}^+)$. Then*

$$f(\mathbf{y}) \geq f(\mathbf{x}^+) + g_{\mathcal{C}}(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{1}{2L} \|g_{\mathcal{C}}(\mathbf{x})\|_2^2.$$

Proof. It follows that

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}^+) &= f(\mathbf{y}) - f(\mathbf{x}) - (f(\mathbf{x}^+) - f(\mathbf{x})) \\ &\geq \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) - \left(\nabla f(\mathbf{x})^\top (\mathbf{x}^+ - \mathbf{x}) + \frac{L}{2} \|\mathbf{x}^+ - \mathbf{x}\|_2^2 \right) \\ &= \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}^+) - \frac{L}{2} \|\mathbf{x}^+ - \mathbf{x}\|_2^2 \\ &\geq g_{\mathcal{C}}(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}^+) - \frac{L}{2} \|\mathbf{x}^+ - \mathbf{x}\|_2^2 \\ &= g_{\mathcal{C}}(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + g_{\mathcal{C}}(\mathbf{x})^\top (\mathbf{x} - \mathbf{x}^+) - \frac{L}{2} \|\mathbf{x}^+ - \mathbf{x}\|_2^2 \\ &= g_{\mathcal{C}}(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{1}{2L} \|g_{\mathcal{C}}(\mathbf{x})\|_2^2, \end{aligned}$$

where the last inequality uses Lemma 2 mentioned above, and we finish the proof. □

Theorem 2. *Let f be convex and L-smooth. If $\eta_t \equiv \eta = \frac{1}{L}$, then PGD obeys*

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

Proof. For the constrained case, we aims to replace $\nabla f(\mathbf{x})$ in the unconstrained case by

$$g_{\mathcal{C}}(\mathbf{x}) = L(\mathbf{x} - \mathcal{P}_{\mathcal{C}}(\mathbf{x} - \nabla f(\mathbf{x}))),$$

we have $g_{\mathcal{C}}(\mathbf{x}_t) = L(\mathbf{x}_t - \mathbf{x}_{t+1})$. Then by applying Lemma 3 and setting $\mathbf{x} = \mathbf{y} = \mathbf{x}_t$, we obtain

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}_t) - \frac{1}{2L} \|g_{\mathcal{C}}(\mathbf{x}_t)\|_2^2,$$

which means we have a guarantee of descent. Applying again and setting $\mathbf{x} = \mathbf{x}_t, \mathbf{y} = \mathbf{x}_*$, we obtain

$$f(\mathbf{x}_*) \geq f(\mathbf{x}_{t+1}) + g_{\mathcal{C}}(\mathbf{x}_t)^\top (\mathbf{x}_* - \mathbf{x}_t) + \frac{1}{2L} \|g_{\mathcal{C}}(\mathbf{x}_t)\|_2^2,$$

then rearrange these terms, we get

$$\begin{aligned} f(\mathbf{x}_{t+1}) &\leq f(\mathbf{x}_*) - g_{\mathcal{C}}(\mathbf{x}_t)^\top (\mathbf{x}_* - \mathbf{x}_t) - \frac{1}{2L} \|g_{\mathcal{C}}(\mathbf{x}_t)\|_2^2 \\ &= f(\mathbf{x}_*) - \frac{1}{2L} \left[2L g_{\mathcal{C}}(\mathbf{x}_t)^\top (\mathbf{x}_* - \mathbf{x}_t) + g_{\mathcal{C}}(\mathbf{x}_t)^\top g_{\mathcal{C}}(\mathbf{x}_t) \right] \\ &= f(\mathbf{x}_*) - \frac{1}{2L} \langle g_{\mathcal{C}}(\mathbf{x}_t), 2L(\mathbf{x}_* - \mathbf{x}_t) + g_{\mathcal{C}}(\mathbf{x}_t) \rangle \\ &= f(\mathbf{x}_*) - \frac{L}{2} \langle \mathbf{x}_t - \mathbf{x}_{t+1}, 2\mathbf{x}_* - \mathbf{x}_t - \mathbf{x}_{t+1} \rangle \\ &= f(\mathbf{x}_*) - \frac{L}{2} \left[\|\mathbf{x}_* - \mathbf{x}_{t+1}\|_2^2 - \|\mathbf{x}_* - \mathbf{x}_t\|_2^2 \right]. \end{aligned}$$

Then by telescoping the last inequality, we finally obtain

$$\begin{aligned} f(\mathbf{x}_t) - f(\mathbf{x}_*) &\leq \frac{1}{t} \sum_{i=0}^{t-1} f(\mathbf{x}_{i+1}) - f(\mathbf{x}_*) \\ &\leq \frac{L}{2t} \left[\|\mathbf{x}_* - \mathbf{x}_0\|_2^2 - \|\mathbf{x}_* - \mathbf{x}_{t+1}\|_2^2 \right] \\ &\leq \frac{L\|\mathbf{x}_* - \mathbf{x}_0\|_2^2}{2t}. \end{aligned}$$

Thus we finish the proof. □