CHAPTER 3. PROJECTED GRADIENT DESCENT

Hao Yuan **Kun Yuan**

October 10, 2023

1 Problem formulation

This chapter considers the following constrained problem

$$\min_{x \in \mathbb{R}^d} \quad f(x), \quad x \in \mathcal{X} \tag{1}$$

where $f(x)$ is a differentiable objective function and $\mathcal{X}$ is a closed convex subset of $\mathbb{R}^d$.

Notation. We introduce the following notations:

- Let $x^\star := \arg\min_{x \in \mathcal{X}}\{f(x)\}$ be the optimial solution to problem (1).
- Let $f^\star := \min_{x \in \mathcal{X}}\{f(x)\}$ be the optimal function value.

2 Projection onto closed convex sets

Lemma 2.1. Given a closed convex set $C \subseteq \mathbb{R}^d$, for any $x \in \mathbb{R}^d$, there exists a unique $z^* \in C$ such that $\|x - z^*\| \le \|x - z\|$ for any $z \in C$. The point z^* will be called the projection of x onto C, and will be denoted by $\mathcal{P}_C[x]$.

Proof. First, we show the existence of z^*. Fix $x \in \mathbb{R}^d$, and denote $\delta := \inf\{\|z - x\|, z \in C\}$. It is evident that $\delta \ge 0$. Now let $\{z_k\}_{k \ge 1}$ be a sequence of points in C such that

$$\|z_k - x\|^2 \le \delta^2 + \frac{1}{k}$$

for any k. We first show that $\{z_k\}_{k\geq 1}$ is a Cauchy sequence. Let $k, l > 0$ be arbitrary, and notice that, since C is convex, we have $\frac{1}{2}(z_k + z_l) \in C$, which implies

$$\|\frac{1}{2}(z_k + z_l) - x\|^2 \geq \delta^2. \tag{2}$$

Expanding this inequality leads to

$$\frac{1}{2}\langle z_k - x, z_l - x\rangle \geq \delta^2 - \frac{1}{4}\|z_k - x\|^2 - \frac{1}{4}\|z_l - x\|^2.$$

We now calculate $\|z_k - z_l\|^2$ and get

$$\begin{aligned}\|z_k - z_l\|^2 &= \|z_k - x\|^2 + \|z_l - x\|^2 - 2\langle z_k - x, z_l - x\rangle \\ &\leq 2(\|z_k - x\|^2 + \|z_l - x\|^2) - 4\delta^2 \leq \frac{2}{k} + \frac{2}{l}.\end{aligned}$$

where we used the inequality 2 in the second step. Thus for any $\epsilon > 0$, as long as $k, l \geq \lceil 4/\epsilon^2 \rceil$, we have $\|z_k - z_l\| \leq \epsilon$, showing that $\{z_k\}$ is a Cauchy sequence. By the completeness of $\mathbb{R}^d$, $z^* = \lim_{k\to\infty} z_k$ exists. Since C is closed, we have $z^* \in C$, and by the continuity of the norm function, we have

$$\|z^* - x\| = \lim_{k\to\infty} \|z_k - x\| = \delta,$$

which is less than or equal to $\|z - x\|$ for any $z \in C$, by the definition of δ.

Next, we show the uniqueness of z^*. Suppose both z_1^* and z_2^* satisfy $\|z_1^* - x\| = \|z_2^* - x\| = \delta$. Denote $\bar{z} = \frac{1}{2}(z_1^* + z_2^*)$, and we have

$$\delta^2 \leq \|\bar{z} - x\|^2 = \left\|\frac{1}{2}(z_1^* - x) + \frac{1}{2}(z_2^* - x)\right\|^2 = \frac{1}{2}\delta^2 + \frac{1}{2}\langle z_1^* - x, z_2^* - x\rangle,$$

which leads to

$$\langle z_1^* - x, z_2^* - x\rangle \geq \delta^2.$$

Consequently,

$$\|z_1^* - z_2^*\|^2 = \|z_1^* - x\|^2 + \|z_2^* - x\|^2 - 2\langle z_1^* - x, z_2^* - x\rangle \leq \delta^2 + \delta^2 - 2\delta^2 = 0,$$

implying that $z_1^* = z_2^*$. The proof is now complete. ☐

Lemma 2.2. Let $C \subseteq \mathbb{R}^d$ be a closed convex set, then for any $x \in \mathbb{R}^d$ and $y \in C$, we have $y = \mathcal{P}_C[x]$ if and only if $\langle z - y, x - y\rangle \leq 0$ for any $z \in C$.

Proof. We first prove that if $y = \mathcal{P}_C[x]$ then $\langle z - y, x - y\rangle \leq 0$ for any $z \in C$. To this end, we fix x and y and suppose there exists $z_0 \in C$ such that $\langle z_0 - y, x - y\rangle > 0$, it then follows that $z_0 \neq y$. Set $z = y + t(z_0 - y)$, it holds that

$$\begin{aligned}\|x - z\|^2 - \|x - y\|^2 &= \|x - y - t(z_0 - y)\|^2 - \|x - y\|^2 \\ &= \|z_0 - y\|^2 t^2 - 2t\langle x - y, z_0 - y\rangle \\ &= \|z_0 - y\|^2 t\left(t - \frac{2\langle x - y, z_0 - y\rangle}{\|z_0 - y\|^2}\right). \end{aligned} \tag{3}$$

Defining $t^* := \min\left\{1, \frac{\langle x-y, z_0-y\rangle}{\|z_0-y\|^2}\right\}$, we have $0 < t^* \leq 1$ and $z^* := y + t^*(z_0 - y) = (1 - t^*)y + t^* z_0 \in C$. Substituting $0 < t^* \leq \frac{\langle x-y, z_0-y\rangle}{\|z_0-y\|^2}$ to (3), we have $\|x - z^*\| < \|x - y\|$, which conflicts with $y = \mathcal{P}_C[x]$.

Next we prove if $y = \mathcal{P}_C[x]$ then $\langle z - y, x - y\rangle \leq 0$ for any $z \in C$ then $y = \mathcal{P}_C[x]$. We notice that $\|x - z\|^2 = \|(x - y) - (z - y)\|^2 = \|x - y\|^2 + \|z - y\|^2 - 2\langle x - y, z - y\rangle$ for any $z \in C$. With $\langle x - y, z - y\rangle \leq 0$, we have $\|x - z\| \geq \|x - y\|$. Since z is arbitrary, we conclude that $y = \mathcal{P}_C[x]$. $\square$

Lemma 2.3. Let $C \subseteq \mathbb{R}^d$ be a closed convex set, then $\|\mathcal{P}_C[x] - \mathcal{P}_C[y]\| \leq \|x - y\|$ for any $x, y \in \mathbb{R}^d$.

Proof. We first notice that

$$\begin{aligned}\|x - y\|^2 &= \|(x - \mathcal{P}_C[x] + \mathcal{P}_C[y] - y) + (\mathcal{P}_C[x] - \mathcal{P}_C[y])\|^2 \\ &= \|x - \mathcal{P}_C[x] + \mathcal{P}_C[y] - y\|^2 + \|\mathcal{P}_C[x] - \mathcal{P}_C[y]\|^2 \\ &\qquad + 2\langle x - \mathcal{P}_C[x] + \mathcal{P}_C[y] - y, \mathcal{P}_C[x] - \mathcal{P}_C[y]\rangle. \\ &= \|x - \mathcal{P}_C[x] + \mathcal{P}_C[y] - y\|^2 + \|\mathcal{P}_C[x] - \mathcal{P}_C[y]\|^2 \\ &\qquad - 2\langle y - \mathcal{P}_C[y], \mathcal{P}_C[x] - \mathcal{P}_C[y]\rangle - 2\langle x - \mathcal{P}_C[x], \mathcal{P}_C[y] - \mathcal{P}_C[x]\rangle. \end{aligned} \tag{4}$$

From Lemma 2.2, we know

$$\langle y - \mathcal{P}_C[y], \mathcal{P}_C[x] - \mathcal{P}_C[y]\rangle \leq 0, \qquad \langle x - \mathcal{P}_C[x], \mathcal{P}_C[y] - \mathcal{P}_C[x]\rangle \leq 0.$$

Substituting the above inequalities to (4), we reach $\|x - y\|^2 \geq \|\mathcal{P}_C[x] - \mathcal{P}_C[y]\|^2$. $\square$

3 Examples of projections

- Box: $C = [\eta_1, \eta_2]^N$, $\forall x \in \mathbb{R}^N$, $(\mathcal{P}_C[x])_i = (\max\{\eta_1, \min\{x_i, \eta_2\}\})$, $i = 1, \cdots, N$.
- Hyperplane: $C = \{x \mid u^\top x = \eta, u \in \mathbb{R}^N, \eta \in \mathbb{R}\}$, $\forall x \in \mathbb{R}^N$, $\mathcal{P}_C[x] = x + \frac{\eta - u^\top x}{\|u\|_2^2} u$.

4 Projected gradient descent

For optimization problem 1, given any arbitrary initialization variable $x_0 \in \mathcal{X}$, projected gradient descent iterates as follows

$$
\begin{aligned}
y_{k+1} &= x_k - \gamma \nabla f(x_k), && \text{(5a)}\\
x_{k+1} &= \mathcal{P}_{\mathcal{X}}[y_{k+1}], \quad \forall\, k = 0, 1, 2, \cdots && \text{(5b)}
\end{aligned}
$$

where γ is the learning rate.

5 Convergence analysis

5.1 Smooth and generally convex problem

Lemma 5.1. Suppose $f(x)$ is L-smooth. If $\gamma = \frac{1}{L}$, then the sequence generated by projected gradient descent (5) with arbitrary $x_0 \in \mathcal{X}$ satisfies

$$
f(x_{k+1}) \le f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{L}{2}\|y_{k+1} - x_{k+1}\|^2, \quad k = 0, 1, 2, \cdots
$$

Proof. Since $f(x)$ is L-smooth, it holds that

$$
\begin{aligned}
f(x_{k+1}) &\le f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2\\
&\overset{(5a)}{=} f(x_k) - L\langle y_{k+1} - x_k, x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2\\
&= f(x_k) - \frac{L}{2}(\|y_{k+1} - x_k\|^2 + \|x_{k+1} - x_k\|^2 - \|y_{k+1} - x_{k+1}\|^2) + \frac{L}{2}\|x_{k+1} - x_k\|^2\\
&= f(x_k) - \frac{L}{2}\|y_{k+1} - x_k\|^2 + \frac{L}{2}\|y_{k+1} - x_{k+1}\|^2\\
&= f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{L}{2}\|y_{k+1} - x_{k+1}\|^2.
\end{aligned}
$$

□

Lemma 5.2. Suppose $f(x)$ is L-smooth. If $\gamma = \frac{1}{L}$, then the sequence generated by projected gradient descent (5) with arbitrary $x_0 \in \mathcal{X}$ satisfies

$$
f(x_{k+1}) \le f(x_k) - \frac{L}{2}\|x_{k+1} - x_k\|^2, \quad k = 0, 1, 2, \cdots
$$

Proof. From Lemma 2.2, we have

$$
\begin{aligned}
& \mathcal{P}_{\mathcal{X}}[x_k - \gamma \nabla f(x_k)] = x_{k+1}\\
\Rightarrow\ & \langle (x_k - \gamma \nabla f(x_k)) - x_{k+1}, x_k - x_{k+1} \rangle \le 0\\
\Rightarrow\ & \|x_{k+1} - x_k\|^2 + \langle \gamma \nabla f(x_k), x_{k+1} - x_k \rangle \le 0\\
\Rightarrow\ & \langle \nabla f(x_k), x_{k+1} - x_k \rangle \le -\frac{1}{\gamma}\|x_{k+1} - x_k\|^2 = -L\|x_{k+1} - x_k\|^2 && (6)
\end{aligned}
$$

Since $f(x)$ is L-smooth, we have

$$\begin{aligned} f(x_{k+1}) &\le f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\ &\overset{(6)}{=} f(x_k) - \frac{L}{2}\|x_{k+1} - x_k\|^2 \end{aligned} \tag{7}$$

□

Theorem 5.3. Suppose $f(x)$ is L-smooth. If $\gamma = \frac{1}{L}$, then the sequence generated by projected gradient descent (5) with arbitrary $x_0 \in \mathcal{X}$ satisfies

$$f(x_K) - f(x^\star) \le \frac{L}{2K}\|x_0 - x^\star\|^2, \quad K > 0.$$

Proof. First, we have

$$\langle \nabla f(x_k), x_k - x^\star \rangle = \frac{1}{2\gamma}(\gamma^2\|\nabla f(x_k)\|^2 + \|x_k - x^\star\|^2 - \|y_{k+1} - x^\star\|^2). \tag{8}$$

From Theorem 2.2, it holds that

$$\langle y_{k+1} - x_{k+1}, x^\star - x_{k+1} \rangle \le 0,$$

which leads to

$$\|x_{k+1} - x^\star\|^2 + \|y_{k+1} - x_{k+1}\|^2 \le \|y_{k+1} - x^\star\|^2. \tag{9}$$

Substituting 9 into 8, we have

$$\langle \nabla f(x_k), x_k - x^\star \rangle \le \frac{1}{2\gamma}(\gamma^2\|\nabla f(x_k)\|^2 + \|x_k - x^\star\|^2 - \|x_{k+1} - x^\star\|^2 - \|y_{k+1} - x_{k+1}\|^2). \tag{10}$$

Then,

$$\begin{aligned} &\sum_{k=0}^{K-1}(f(x_k) - f(x^\star)) \\ \le &\sum_{k=0}^{K-1}\langle \nabla f(x_k), x_k - x^\star \rangle \\ \le &\frac{1}{2L}\sum_{k=0}^{K-1}\|\nabla f(x_k)\|^2 + \frac{L}{2}\|x_0 - x^\star\|^2 - \frac{L}{2}\sum_{k=0}^{K-1}\|y_{k+1} - x_{k+1}\|^2. \end{aligned} \tag{11}$$

From Lemma 5.1, we have

$$\frac{1}{2L}\sum_{k=0}^{K-1}\|\nabla f(x_k)\|^2 \le \sum_{k=0}^{K-1}(f(x_k)-f(x_{k+1})+\frac{L}{2}\|y_{k+1}-x_{k+1}\|^2)$$
$$= f(x_0)-f(x_K)+\frac{L}{2}\sum_{k=0}^{K-1}\|y_{k+1}-x_{k+1}\|^2.$$

Plugging this into 11, we have

$$\sum_{k=1}^{K}(f(x_k)-f(x^\star)) \le \frac{L}{2}\|x_0-x^\star\|^2.$$

Using the Lemma 5.2, we complete the proof. □

5.2 Smooth and strongly convex problem

Theorem 5.4. Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a closed convex set, $f:\mathcal{X}\to\mathbb{R}$ be differentiable, L-smooth and μ-strongly convex. If $\gamma=\frac{1}{L}$, projected gradient descent 5 with arbitrary $x_0\in\mathcal{X}$ satisfies

$$\|x_K-x^\star\|^2 \le (1-\frac{\mu}{L})^K\|x_0-x^\star\|^2, \quad K>0.$$

Proof.

$$\begin{aligned}
& f(x_k)-f(x^*) \\
\le & \langle\nabla f(x_k), x_k-x^\star\rangle-\frac{\mu}{2}\|x_k-x^\star\|^2 \\
\le & \frac{1}{2\gamma}(\gamma^2\|\nabla f(x_k)\|^2+\|x_k-x^\star\|^2-\|x_{k+1}-x^\star\|^2-\|y_{k+1}-x_{k+1}\|^2)-\frac{\mu}{2}\|x_k-x^\star\|^2, \quad (12)
\end{aligned}$$

where the first inequality is from the strongly convex property and the last inequality is from 10. 12 leads to

$$\|x_{k+1}-x^\star\|^2 \le 2\gamma(f(x^\star)-f(x_k))+\gamma^2\|\nabla f(x_k)\|^2-\|y_{k+1}-x_{k+1}\|^2+(1-\gamma\mu)\|x_k-x^\star\|^2. \quad (13)$$

Using Lemma 5.1 and Lemma 5.2, we have

$$f(x^\star)-f(x_k) \le f(x_{k+1})-f(x_k) \le -\frac{1}{2L}\|\nabla f(x_k)\|^2+\frac{L}{2}\|y_{k+1}-x_{k+1}\|^2. \quad (14)$$

Substituting 14 into 13, we have

$$\|x_{k+1}-x^\star\|^2 \le (1-\frac{\mu}{L})\|x_k-x^\star\|^2,$$

which completes the proof. □

Remark: In Theorem 5.4, if we suppose f is differentiable, L-smooth and μ-strongly convex on $\mathbb{R}^d$, then the result can be strengthened as

$$\|x_K - x^\star\| \leq (1 - \frac{\mu}{L})^K \|x_0 - x^\star\|, \quad K > 0.$$

The proof can be found in chap 4 Theorem 5.7.