9 Extension of the Optimal Gradient Method (First-Order Method, Accelerated Gradient Method, Fast Gradient Method) for the Min-Max Problems over Simple Closed Convex Sets

Suppose we are given Q a closed convex subset of $\mathbb{R}^n$, simple enough to have an easy projection onto it. *E.g.*, positive orthant, n-dimensional box, simplex, Euclidean ball, ellipsoids, *etc.*

Given $f_i \in \mathcal{S}^{1,1}_{\mu,L}(Q)$ $(i = 1, 2, \ldots, m)$, we define the following function $f : Q \to \mathbb{R}$,

$$f(\boldsymbol{x}) := \max_{1\leq i\leq m} f_i(\boldsymbol{x}) \qquad \text{for} \qquad \boldsymbol{x} \in Q. \tag{18}$$

This function is non-differentiable in general, but convex (see Theorem 5.6). We will see that the method discussed so far can be easily adapted for the following min-max-type convex optimization problem.

$$\left\{ \begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in Q, \end{array} \right. \tag{19}$$

where Q is a closed convex set with a simple structure, and $f(\boldsymbol{x})$ is defined as above.

For a given $\bar{\boldsymbol{x}} \in Q$, let us define the following linearization of $f(\boldsymbol{x})$ at $\bar{\boldsymbol{x}}$.

$$f(\bar{\boldsymbol{x}}; \boldsymbol{x}) := \max_{1\leq i\leq m} \left[f_i(\bar{\boldsymbol{x}}) + \langle \boldsymbol{\nabla f}_i(\bar{\boldsymbol{x}}), \boldsymbol{x} - \bar{\boldsymbol{x}} \rangle \right], \qquad \text{for } \boldsymbol{x} \in Q.$$

Lemma 9.1 Let $f_i \in \mathcal{S}^{1,1}_{\mu,L}(Q)$ $(i = 1, 2, \ldots, m)$ and $\bar{\boldsymbol{x}} \in Q$. For $\boldsymbol{x} \in Q$, we have

$$\begin{aligned} f(\boldsymbol{x}) &\geq f(\bar{\boldsymbol{x}}; \boldsymbol{x}) + \frac{\mu}{2}\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2^2, \\ f(\boldsymbol{x}) &\leq f(\bar{\boldsymbol{x}}; \boldsymbol{x}) + \frac{L}{2}\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2^2. \end{aligned}$$

Proof:

It follows from the properties of $f_i \in \mathcal{S}^{1,1}_{\mu,L}(Q)$. ∎

Theorem 9.2 A point $\boldsymbol{x}^* \in Q$ is an optimal solution of (19) with $f_i \in \mathcal{S}^{1,1}_{\mu,L}(Q)$ $(i = 1, 2, \ldots, m)$ if and only if

$$f(\boldsymbol{x}^*; \boldsymbol{x}) \geq f(\boldsymbol{x}^*; \boldsymbol{x}^*) = f(\boldsymbol{x}^*), \quad \forall \boldsymbol{x} \in Q.$$

Proof:

Indeed, if the inequality is true, it follows from Lemma 9.1 that

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*; \boldsymbol{x}) + \frac{\mu}{2}\|\boldsymbol{x} - \boldsymbol{x}^*\|_2^2 \geq f(\boldsymbol{x}^*) + \frac{\mu}{2}\|\boldsymbol{x} - \boldsymbol{x}^*\|_2^2 \geq f(\boldsymbol{x}^*), \quad \forall \boldsymbol{x} \in Q.$$

For the converse, let $\boldsymbol{x}^*$ be an optimal solution of the minimization problem (19). Assume by contradiction that there is a $\boldsymbol{x} \in Q$ such that $f(\boldsymbol{x}^*; \boldsymbol{x}) < f(\boldsymbol{x}^*)$.

Therefore, from the definition of $f(\cdot)$, letting $j \in \{1, 2, \ldots, m\}$ (which temporarily we assume is unique) such that $f_j(\boldsymbol{x}^*) = \max_{1\leq i\leq m} f_i(\boldsymbol{x}^*)$, we have

$$f_i(\boldsymbol{x}^*) + \langle \boldsymbol{\nabla f}_i(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle < f_j(\boldsymbol{x}^*) \quad \text{for} \quad i = 1, 2, \ldots, m \tag{20}$$

Notice that $\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*) \in Q$ for $\alpha \in [0, 1]$ since Q is convex. Then, calling $\phi_i(\alpha) := f_i(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*))$, we have $\phi_i'(0) = \langle \boldsymbol{\nabla f}_i(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle$. Moreover, $\phi_i(0) = f_i(\boldsymbol{x}^*) < f_j(\boldsymbol{x}^*)$ for

$i = 1, 2, \ldots, m$, $i \neq j$, and $\phi_j(0) = f_j(\boldsymbol{x}^*) = f(\boldsymbol{x}^*)$ and $\langle \boldsymbol{\nabla f}_j(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle < 0$ from (20) for $i = j$. Therefore, there exists $\tilde{\alpha} > 0$ small enough such that

$$\phi_j(\tilde{\alpha}) = f_j(\boldsymbol{x}^* + \tilde{\alpha}(\boldsymbol{x} - \boldsymbol{x}^*)) < \phi_j(0) = f_j(\boldsymbol{x}^*)$$

and

$$\phi_i(\tilde{\alpha}) = f_i(\boldsymbol{x}^* + \tilde{\alpha}(\boldsymbol{x} - \boldsymbol{x}^*)) < f_j(\boldsymbol{x}^*) \text{ for } i = 1, 2, \ldots, m \ i \neq j.$$

Finally, we have $f(\boldsymbol{x}^* + \tilde{\alpha}(\boldsymbol{x} - \boldsymbol{x}^*)) = \max_{1 \leq i \leq m} f_i(\boldsymbol{x}^* + \tilde{\alpha}(\boldsymbol{x} - \boldsymbol{x}^*)) < f_j(\boldsymbol{x}^*) = \max_{1 \leq i \leq m} f_i(\boldsymbol{x}^*) = f(\boldsymbol{x}^*)$. Therefore, we arrived to a contradiction. In the case there exists j_1, j_2 such that $f(\boldsymbol{x}^*) = f_{j_1}(\boldsymbol{x}^*) = f_{j_2}(\boldsymbol{x}^*)$ and $f_{j_1}(\boldsymbol{x}^* + \tilde{\alpha}(\boldsymbol{x} - \boldsymbol{x}^*)) < f_{j_2}(\boldsymbol{x}^* + \tilde{\alpha}(\boldsymbol{x} - \boldsymbol{x}^*))$, we choose $j = j_2$ and still we have the same conclusion. ∎

Corollary 9.3 Let $\boldsymbol{x}^*$ be a minimum of a max-type function $f(\boldsymbol{x})$ over the set Q as (18). If $f_i \in \mathcal{S}^1_\mu(Q)$ $(i = 1, 2, \ldots, m)$, then

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \frac{\mu}{2} \|\boldsymbol{x} - \boldsymbol{x}^*\|_2^2, \quad \forall \boldsymbol{x} \in Q.$$

Proof:
From Lemma 9.1 and Theorem 9.2, we have for $\forall \boldsymbol{x} \in Q$,

$$\begin{aligned} f(\boldsymbol{x}) &\geq f(\boldsymbol{x}^*; \boldsymbol{x}) + \frac{\mu}{2} \|\boldsymbol{x} - \boldsymbol{x}^*\|_2^2 \\ &\geq f(\boldsymbol{x}^*; \boldsymbol{x}^*) + \frac{\mu}{2} \|\boldsymbol{x} - \boldsymbol{x}^*\|_2^2 = f(\boldsymbol{x}^*) + \frac{\mu}{2} \|\boldsymbol{x} - \boldsymbol{x}^*\|_2^2. \end{aligned}$$

∎

Lemma 9.4 Let $f_i \in \mathcal{S}^1_\mu(Q)$ for $(i = 1, 2, \ldots, m)$ with $\mu > 0$ and Q be a closed convex set. Then there is a unique solution $\boldsymbol{x}^*$ for the problem (19).

Proof:
Left for exercise. ∎

Definition 9.5 Let $f_i \in \mathcal{C}^1(Q)$ $(i = 1, 2, \ldots, m)$, Q a closed convex set, $\bar{\boldsymbol{x}} \in Q$, and $\gamma > 0$. Denote by

$$\begin{aligned} \boldsymbol{x}_f(\bar{\boldsymbol{x}}; \gamma) &:= \arg\min_{\boldsymbol{y} \in Q} \left[f(\bar{\boldsymbol{x}}; \boldsymbol{y}) + \frac{\gamma}{2} \|\boldsymbol{y} - \bar{\boldsymbol{x}}\|_2^2 \right], \\ \boldsymbol{g}_f(\bar{\boldsymbol{x}}; \gamma) &:= \gamma(\bar{\boldsymbol{x}} - \boldsymbol{x}_f(\bar{\boldsymbol{x}}; \gamma)). \end{aligned}$$

We call $\boldsymbol{g}_f(\bar{\boldsymbol{x}}; \gamma)$ the *gradient mapping of max-type function f on Q*. Observe that due to Lemma 9.4, $\boldsymbol{x}_f(\bar{\boldsymbol{x}}; \gamma)$ exists and it is uniquely defined.

Theorem 9.6 Let $f_i \in \mathcal{S}^{1,1}_{\mu,L}(Q)$ $(i = 1, 2, \ldots, m)$, $\gamma \geq L$, $\gamma > 0$, Q a closed convex set, and $\bar{\boldsymbol{x}} \in Q$. Then

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}_f(\bar{\boldsymbol{x}}; \gamma)) + \langle \boldsymbol{g}_f(\bar{\boldsymbol{x}}; \gamma), \boldsymbol{x} - \bar{\boldsymbol{x}} \rangle + \frac{1}{2\gamma} \|\boldsymbol{g}_f(\bar{\boldsymbol{x}}; \gamma)\|_2^2 + \frac{\mu}{2} \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2^2, \quad \forall \boldsymbol{x} \in Q.$$

Proof: Let us use the following notation: $\boldsymbol{x}_f := \boldsymbol{x}_f(\bar{\boldsymbol{x}}; \gamma)$ and $\boldsymbol{g}_f := \boldsymbol{g}_f(\bar{\boldsymbol{x}}; \gamma)$.
From Lemma 9.1 and Corollary 9.3 (taking $f(\boldsymbol{x})$ in there as $f(\bar{\boldsymbol{x}}; \boldsymbol{x}) + \frac{\gamma}{2} \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2^2$), we have $\forall \boldsymbol{x} \in Q$,

$$
\begin{aligned}
f(\boldsymbol{x}) - \frac{\mu}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_2^2 &\geq f(\bar{\boldsymbol{x}};\boldsymbol{x}) \\
&= f(\bar{\boldsymbol{x}};\boldsymbol{x}) + \frac{\gamma}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_2^2 - \frac{\gamma}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_2^2 \\
&\geq f(\bar{\boldsymbol{x}};\boldsymbol{x}_f) + \frac{\gamma}{2}\|\boldsymbol{x}_f-\bar{\boldsymbol{x}}\|_2^2 + \frac{\gamma}{2}\|\boldsymbol{x}-\boldsymbol{x}_f\|_2^2 - \frac{\gamma}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_2^2 \\
&= f(\bar{\boldsymbol{x}};\boldsymbol{x}_f) + \frac{\gamma}{2}\|\boldsymbol{x}_f-\bar{\boldsymbol{x}}\|_2^2 + \frac{\gamma}{2}\langle \bar{\boldsymbol{x}}-\boldsymbol{x}_f, 2\boldsymbol{x}-\boldsymbol{x}_f-\bar{\boldsymbol{x}}\rangle \\
&= f(\bar{\boldsymbol{x}};\boldsymbol{x}_f) + \frac{\gamma}{2}\|\boldsymbol{x}_f-\bar{\boldsymbol{x}}\|_2^2 + \frac{\gamma}{2}\langle \bar{\boldsymbol{x}}-\boldsymbol{x}_f, 2(\boldsymbol{x}-\bar{\boldsymbol{x}})+\bar{\boldsymbol{x}}-\boldsymbol{x}_f\rangle \\
&= f(\bar{\boldsymbol{x}};\boldsymbol{x}_f) + \frac{\gamma}{2}\|\boldsymbol{x}_f-\bar{\boldsymbol{x}}\|_2^2 + \langle \boldsymbol{g}_f, \boldsymbol{x}-\bar{\boldsymbol{x}}\rangle + \frac{1}{2\gamma}\|\boldsymbol{g}_f\|_2^2 \\
&\geq f(\boldsymbol{x}_f) + \langle \boldsymbol{g}_f, \boldsymbol{x}-\bar{\boldsymbol{x}}\rangle + \frac{1}{2\gamma}\|\boldsymbol{g}_f\|_2^2,
\end{aligned}
$$

where the last inequality is due to the fact that $\gamma \geq L$. ∎

Now, we are ready to define our estimated sequence. Assume that $f_i \in \mathcal{S}^{1,1}_{\mu,L}(Q)$ $(i = 1, 2, \ldots, m)$ possible with $\mu = 0$ (which means that $f_i \in \mathcal{F}^{1,1}_L(Q)$), $\boldsymbol{x}_0 \in Q$, and $\gamma_0 > 0$. Define

$$
\begin{aligned}
\phi_0(\boldsymbol{x}) &:= f(\boldsymbol{x}_0) + \frac{\gamma_0}{2}\|\boldsymbol{x}-\boldsymbol{x}_0\|_2^2, \\
\phi_{k+1}(\boldsymbol{x}) &:= (1-\alpha_k)\phi_k(\boldsymbol{x}) + \alpha_k\left[f(\boldsymbol{x}_f(\boldsymbol{y}_k;L)) + \frac{1}{2L}\|\boldsymbol{g}_f(\boldsymbol{y}_k;L)\|_2^2 + \langle \boldsymbol{g}_f(\boldsymbol{y}_k;L), \boldsymbol{x}-\boldsymbol{y}_k\rangle \right. \\
&\qquad \left. + \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{y}_k\|_2^2\right],
\end{aligned}
$$

for the sequences $\{\alpha_k\}_{k=0}^{\infty}$ and $\{\boldsymbol{y}_k\}_{k=0}^{\infty}$ which will be defined later.

Similarly to the previous subsection, we can prove that $\{\phi_k(\boldsymbol{x})\}_{k=0}^{\infty}$ can be written in the form

$$
\phi_k(\boldsymbol{x}) = \phi_k^* + \frac{\gamma_k}{2}\|\boldsymbol{x}-\boldsymbol{v}_k\|_2^2
$$

for $\phi_0^* = f(\boldsymbol{x}_0)$, $\boldsymbol{v}_0 = \boldsymbol{x}_0$:

$$
\begin{aligned}
\gamma_{k+1} &= (1-\alpha_k)\gamma_k + \alpha_k\mu \\
\boldsymbol{v}_{k+1} &= \frac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k\boldsymbol{v}_k + \alpha_k\mu\boldsymbol{y}_k - \alpha_k\boldsymbol{g}_f(\boldsymbol{y}_k;L)], \\
\phi_{k+1}^* &= (1-\alpha_k)\phi_k^* + \alpha_k f(\boldsymbol{x}_f(\boldsymbol{y}_k;L)) + \left(\frac{\alpha_k}{2L} - \frac{\alpha_k^2}{2\gamma_{k+1}}\right)\|\boldsymbol{g}_f(\boldsymbol{y}_k;L)\|_2^2 \\
&\quad + \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\frac{\mu}{2}\|\boldsymbol{y}_k-\boldsymbol{v}_k\|_2^2 + \langle \boldsymbol{g}_f(\boldsymbol{y}_k;L), \boldsymbol{v}_k-\boldsymbol{y}_k\rangle\right).
\end{aligned}
$$

Now, $\phi_0^* \geq f(\boldsymbol{x}_0)$. Assuming that $\phi_k^* \geq f(\boldsymbol{x}_k)$,

$$
\begin{aligned}
\phi_{k+1}^* &\geq (1-\alpha_k)f(\boldsymbol{x}_k) + \alpha_k f(\boldsymbol{x}_f(\boldsymbol{y}_k;L)) + \left(\frac{\alpha_k}{2L} - \frac{\alpha_k^2}{2\gamma_{k+1}}\right)\|\boldsymbol{g}_f(\boldsymbol{y}_k;L)\|_2^2 \\
&\quad + \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\langle \boldsymbol{g}_f(\boldsymbol{y}_k;L), \boldsymbol{v}_k-\boldsymbol{y}_k\rangle \\
&\geq f(\boldsymbol{x}_f(\boldsymbol{y}_k;L)) + \left(\frac{1}{2L} - \frac{\alpha_k^2}{2\gamma_{k+1}}\right)\|\boldsymbol{g}_f(\boldsymbol{y}_k;L)\|_2^2 \\
&\quad + (1-\alpha_k)\left\langle \boldsymbol{g}_f(\boldsymbol{y}_k;L), \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(\boldsymbol{v}_k-\boldsymbol{y}_k) + \boldsymbol{x}_k-\boldsymbol{y}_k\right\rangle + \frac{(1-\alpha_k)\mu}{2}\|\boldsymbol{x}_k-\boldsymbol{y}_k\|_2^2,
\end{aligned}
$$

where the last inequality follows from Theorem 9.6.

Therefore, if we choose

$$\begin{array}{rcl} \boldsymbol{x}_{k+1} & = & \boldsymbol{x}_f(\boldsymbol{y}_k; L), \\ L\alpha_k^2 & = & (1-\alpha_k)\gamma_k + \alpha_k\mu, \\ \gamma_{k+1} & := & L\alpha_k^2, \\ \boldsymbol{y}_k & = & \dfrac{1}{\gamma_k + \alpha_k\mu}(\alpha_k\gamma_k\boldsymbol{v}_k + \gamma_{k+1}\boldsymbol{x}_k), \end{array}$$

we obtain $\phi_{k+1}^* \geq f(\boldsymbol{x}_{k+1})$ as desired.

Hereafter, we assume that $L > \mu$ to exclude the trivial case $L = \mu$ with finished in one iteration.

Constant Step Scheme for the Optimal Gradient Method for the Min-Max Problem	
Step 0:	Choose $\boldsymbol{x}_0 \in Q$, $\alpha_0 \in (0,1)$ such that $\frac{\alpha_0(\alpha_0 L-\mu)}{1-\alpha_0} > 0$, $\mu \leq \frac{\alpha_0(\alpha_0 L-\mu)}{1-\alpha_0} \leq L$, set $\boldsymbol{y}_0 := \boldsymbol{x}_0$, $k := 0$.
Step 1:	Compute $f_i(\boldsymbol{y}_k)$ and $\boldsymbol{\nabla f}_i(\boldsymbol{y}_k)$ $(i = 1, 2, \ldots, m)$.
Step 2:	Set $\boldsymbol{x}_{k+1} := \boldsymbol{x}_f(\boldsymbol{y}_k; L) := \arg\min_{\boldsymbol{x}\in Q} \left[\max_{i=1,2,\ldots,m} f_i(\boldsymbol{y}_k) + \langle \boldsymbol{\nabla f}_i(\boldsymbol{y}_k), \boldsymbol{x} - \boldsymbol{y}_k \rangle + \frac{\alpha_k(\alpha_k L-\mu)}{2(1-\alpha_k)} \|\boldsymbol{x} - \boldsymbol{y}_k\|_2^2 \right].$
Step 3:	Compute $\alpha_{k+1} \in (0,1)$ from the equation $\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + \frac{\mu}{L}\alpha_{k+1}$.
Step 4:	Set $\beta_k := \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2 + \alpha_{k+1}}$.
Step 5:	Set $\boldsymbol{y}_{k+1} := \boldsymbol{x}_{k+1} + \beta_k(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)$, $k := k+1$ and go to Step 1.

The rate of converge of this method is exactly the same as Theorem 8.6 for $\gamma_0 := \alpha_0(\alpha_0 L - \mu)/(1-\alpha_0)$, but we need to solve a convex program in Step 2 for each iteration, and it can turn the method computationally expensive.

9.1 Exercises

1. Prove Lemma 9.4.