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

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

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

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

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 11.6.

Therefore, if we choose

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

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 I for the Optimal Gradient Method for the Min-Max Problem	
Step 0:	Choose $\boldsymbol{x}_0 \in \mathbb{R}^n$, $\alpha_0 \in (0,1)$ 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 $f_i'(\boldsymbol{y}_k)$ $(i = 1, 2, \ldots, m)$.
Step 2:	Set $\boldsymbol{x}_{k+1} := \boldsymbol{x}_f(\boldsymbol{y}_k; L) := \arg\min\limits_{\boldsymbol{x} \in Q} \left[\max\limits_{i=1,2,\ldots,m} f_i(\boldsymbol{y}_k) + \langle f_i'(\boldsymbol{y}_k), \boldsymbol{x} - \boldsymbol{y}_k \rangle + \frac{\alpha_k(\alpha_k L-\mu)}{2(1-\alpha_k)} \Vert \boldsymbol{x} - \boldsymbol{y}_k \Vert_2^2 \right]$.
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 algorithm is exactly the same as the previous ones, but it is necessary to solve a convex program in Step 2 for each iteration.

12 Applications for Optimization Problems with Convex Constraints

The techniques developed in the previous section can be applied to solve the following smooth convex problem:

$$\begin{cases} \text{minimize} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, \quad (i = 1, 2, \ldots, m) \\ & \boldsymbol{x} \in Q, \end{cases} \tag{17}$$

where $f_i \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$ $(i = 0, 1, \ldots, m)$ with $\mu > 0$ and Q is a closed convex subset of $\mathbb{R}^n$.

Let us introduce the *parametric max-type function*:

$$f(t; \boldsymbol{x}) = \max\{f_0(\boldsymbol{x}) - t; f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})\}, \quad t \in \mathbb{R}, \quad \boldsymbol{x} \in Q,$$

in order to define the function

$$f^*(t) = \min_{\boldsymbol{x} \in Q} f(t; \boldsymbol{x}). \tag{18}$$

Since the components of the parametric max-type function $f(t; \cdot)$ are strongly convex in $\boldsymbol{x}$, (18) has a unique solution $\boldsymbol{x}^*(t)$ for any $t \in \mathbb{R}$ due to Lemma 11.4.

Lemma 12.1 Let t^* be an optimal value of the problem (17). Then

$$\begin{aligned} f^*(t) \leq 0, \quad & \text{for all} \quad t \geq t^*, \\ f^*(t) > 0, \quad & \text{for all} \quad t < t^*, \end{aligned}$$

Proof: Let $\boldsymbol{x}^*$ be a solution of (17). Consider first the case $t \geq t^*$. Then

$$\begin{aligned} f^*(t) &\leq f(t; \boldsymbol{x}^*) = \max\{f_0(\boldsymbol{x}^*) - t, f_1(\boldsymbol{x}^*), f_2(\boldsymbol{x}^*), \ldots, f_m(\boldsymbol{x}^*)\} \\ &= \max\{t^* - t, f_1(\boldsymbol{x}^*), f_2(\boldsymbol{x}^*), \ldots, f_m(\boldsymbol{x}^*)\} \leq 0. \end{aligned}$$

Now, let $t < t^*$. Additionally, suppose that $f^*(t) \leq 0$. Then, there exists an $\boldsymbol{y} \in Q$ such that

$$f_0(\boldsymbol{y}) \leq t < t^*, \quad f_i(\boldsymbol{y}) \leq 0, \quad (i = 1, 2, \ldots, m).$$

Thus, t^* can not be an optimal value of the problem. ∎

We conclude that the root of the function $f^*(t)$ corresponds to the optimal value of (17).

Lemma 12.2 For any $t_1 < t_2$ and $\delta \geq 0$, we have

$$f^*(t_1 - \delta) \geq f^*(t_1) + \frac{\delta}{t_2 - t_1}(f^*(t_1) - f^*(t_2)).$$

Proof: Let $\alpha = \frac{\delta}{t_2 - t_1 + \delta} \in [0, 1]$. Then, the statement can be rewritten as:

$$\begin{aligned} f^*(t_1 - \delta) &\geq \left(\frac{t_2 - t_1 + \delta}{t_2 - t_1}\right) f^*(t_1) - \frac{\delta}{t_2 - t_1} f^*(t_2) \\ \frac{\delta}{t_2 - t_1 + \delta} f^*(t_2) + \frac{t_2 - t_1}{t_2 - t_1 + \delta} f^*(t_1 - \delta) &\geq f^*(t_1) \\ \alpha f^*(t_2) + (1 - \alpha) f^*(t_1 - \delta) &\geq f^*(\alpha t_2 + (1 - \alpha)(t_1 - \delta)), \end{aligned}$$

but the convexity of $f^*(t)$ is a consequence of Theorem 6.5. ∎

Please check "Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, (Kluwer Academic Publishers, Boston, 2004)" for further details.

12.1 Further Reading

There are other variants of the method which are considered more efficient in general. For instance,

- Yu. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical Programming* **103** (2005), pp. 127–152. The algorithm described in page 150. In there, there is a typo: τ_k in c) must be α_{k+1}.
- P. Tseng, "Approximation accuracy, gradient methods, and error bound for structured convex optimization," *Mathematical Programming, Series B* **125** (2010), pp. 125–295. The three Accelerated Proximal Gradient (APG) methods, described in page 274, although it covers a more general case, it can be applied to (15).
- Yu. Nesterov, "A method of solving a convex programming problem with convergence rate $OC(1/k^2)$," *Soviet Mathematics Doklady*, **27** (1984), pp. 372–376. This is the classical result.

The most interesting case is when $f(\boldsymbol{x})$ is non-differentiable or it is a composite type of function. Classical algorithms includes the one in the first reference or

- Yu. Nesterov, "Primal-dual subgradient methods for convex problems," *Mathematical Programming, Series B*, **120** (2009), pp. 221-259.
- A. Beck and M. Teboulle, "Mirror descent and nonlinear projected subgradient methods for convex optimization," *Operations Research Letters*, **31** (2003), pp. 167–175.