The University of Texas at Austin
Department of Electrical and Computer Engineering

ECE 381V: Large-Scale Optimization II — Spring 2022

Lecture 16

Caramanis & Mokhtari Wednesday, March 23, 2022

Goal: In this lecture, we talk about functional constrained problem as well as the projected dual subgradient method and its convergence guarantees.

1 Dual representation

Consider the following constrained optimization problem with functional constraints.

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) \\ \text{s.t} \quad & f_i(\mathbf{x}) \leq 0 \qquad \text{for } i = 1, \dots, m, \\ & \mathbf{x} \in \mathcal{Q} \end{aligned} \tag{1}$$

where the following conditions hold:

- $\mathcal{Q} \subseteq \mathbb{E}$
- $f_0 : \mathbb{E} \to \mathbb{R}$ is convex.
- $f_i : \mathbb{E} \to \mathbb{R}$ are convex for $i = 1, \dots, m$.
- The optimal objective function value is finite and denoted by f^*, and the optimal solution set is non-empty and denoted by $\mathcal{X}^*$.
- Slater's condition holds: There exists $\hat{\mathbf{x}} \in \text{int}\mathcal{Q}$ such that $f_i(\mathbf{x}) < 0$ for $i = 1, \dots, m$.

Note that the Lagrangian of the above problem is given by

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) = f_0(\mathbf{x}) + \boldsymbol{\lambda}^\top \mathbf{f}(\mathbf{x})$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$ and $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))$. The dual function is also defined as

$$q(\boldsymbol{\lambda}) = \min_{\mathbf{x} \in \mathcal{Q}} \{\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})\} = \min_{\mathbf{x} \in \mathcal{Q}} \{f_0(\mathbf{x}) + \boldsymbol{\lambda}^\top \mathbf{f}(\mathbf{x})\}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^m$ and the dual problem is formally given by

$$\begin{aligned} \max \quad & q(\boldsymbol{\lambda}) \\ \text{s.t} \quad & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned} \tag{2}$$

Since we deal with a convex program and Slater's condition holds, then strong duality holds, i.e.,

$$f^* = q^*$$

An important result that we need in the analysis of iterative dual methods, is boundedness of the dual variables $\boldsymbol{\lambda}$. In the following theorem, we show that the superlevel sets of the dual objective function are bounded.

Theorem 1. *If the above conditions are satisfied, then there exists* $\hat{\mathbf{x}}$ *such that* $\hat{\mathbf{x}} \in int\mathcal{Q}$ *and* $f_i(\hat{\mathbf{x}}) < 0$ *for* $i = 1, \ldots, m$. *Then, for any* $\mu \in \mathbb{R}$ *we have that the elements of the super level set* $\mathcal{S}_\mu := \{\boldsymbol{\lambda} \in \mathbb{R}_+^m : q(\boldsymbol{\lambda}) \geq \mu\}$ *satisfy the following condition:*

$$\|\boldsymbol{\lambda}\|_2 \leq \frac{f_0(\hat{\mathbf{x}}) - \mu}{\min_{i=1,\ldots,m} \; -f_i(\hat{\mathbf{x}})}$$

Proof. For $\boldsymbol{\lambda} \in \mathcal{S}_\mu$ we have

$$\mu \leq q(\boldsymbol{\lambda}) \leq f_0(\hat{\mathbf{x}}) + \sum_{i=1}^{m} \lambda_i f_i(\hat{\mathbf{x}})$$

Hence,

$$-\sum_{i=1}^{m} \lambda_i f_i(\hat{\mathbf{x}}) \leq f_0(\hat{\mathbf{x}}) - \mu$$

Now, using the fact that $\lambda_i \geq 0$ and $f_i(\hat{\mathbf{x}}) < 0$ we have

$$(\min_{i=1,\ldots,m} -f_i(\hat{\mathbf{x}})) \sum_{i=1}^{m} \lambda_i \leq f_0(\hat{\mathbf{x}}) - \mu$$

Now since

$$\|\boldsymbol{\lambda}\|_2 = \sqrt{\sum_{i=1}^{m} \lambda_i^2} \leq \sum_{i=1}^{m} |\lambda_i| = \sum_{i=1}^{m} \lambda_i$$

the claim follows.

Corollary 1. *Deonte the optimal solution set of the dual problem by* Λ^*. *Then, for any* $\boldsymbol{\lambda}^* \in \Lambda^*$ *we have*

$$\|\boldsymbol{\lambda}^*\|_2 \leq \frac{f_0(\hat{\mathbf{x}}) - f^*}{\min_{i=1,\ldots,m} \; -f_i(\hat{\mathbf{x}})}$$

□

1.1 Dual function subgradient

Note that the dual function is defined as

$$q(\boldsymbol{\lambda}) = \min_{\mathbf{x} \in \mathcal{Q}} \{f_0(\mathbf{x}) + \boldsymbol{\lambda}^\top \mathbf{f}(\mathbf{x})\}$$

Hence, if $\mathbf{x}^*(\boldsymbol{\lambda})$ is the minimizer of the above problem we have

$$q(\boldsymbol{\lambda}) = f_0(\mathbf{x}^*(\boldsymbol{\lambda})) + \boldsymbol{\lambda}^\top \mathbf{f}(\mathbf{x}^*(\boldsymbol{\lambda}))$$

Then, $\mathbf{f}(\mathbf{x}^*(\boldsymbol{\lambda}))$ is a subgradient of the dual function, i.e.,

$$-\mathbf{f}(\mathbf{x}^*(\boldsymbol{\lambda})) \in \partial(-q)(\boldsymbol{\lambda})$$

This can be easily verified as

$$\begin{aligned}
q(\boldsymbol{\theta}) &= \min_{\mathbf{x} \in \mathcal{Q}} \{f_0(\mathbf{x}) + \boldsymbol{\theta}^\top \mathbf{f}(\mathbf{x})\} \\
&\leq f_0(\mathbf{x}^*(\boldsymbol{\lambda})) + \boldsymbol{\theta}^\top \mathbf{f}(\mathbf{x}^*(\boldsymbol{\lambda})) \\
&= f_0(\mathbf{x}^*(\boldsymbol{\lambda})) + \boldsymbol{\lambda}^\top \mathbf{f}(\mathbf{x}^*(\boldsymbol{\lambda})) + (\boldsymbol{\theta} - \boldsymbol{\lambda})^\top \mathbf{f}(\mathbf{x}^*(\boldsymbol{\lambda})) \\
&= q(\boldsymbol{\lambda}) + (\boldsymbol{\theta} - \boldsymbol{\lambda})^\top \mathbf{f}(\mathbf{x}^*(\boldsymbol{\lambda}))
\end{aligned}$$

Hence, we have

$$-q(\boldsymbol{\theta}) \geq -q(\boldsymbol{\lambda}) + (-\mathbf{f}(\mathbf{x}^*(\boldsymbol{\lambda})))^\top (\boldsymbol{\theta} - \boldsymbol{\lambda})$$

2 The Dual Projected Subgradient Method

Note that the problem of maximizing q over $\boldsymbol{\lambda} \in \mathbb{R}_+^m$ is equivalent to the problem of minimizing $-q$ over $\boldsymbol{\lambda} \in \mathbb{R}_+^m$.
Therefore, we can simply use the projected subgradient method to minimize $-q$ over $\boldsymbol{\lambda} \in \mathbb{R}_+^m$, where its update is given by:

- At iteration k we are given $\boldsymbol{\lambda}_k \in \mathbb{R}_+^m$.
- To compute the dual function $-q$ subgradient at $\boldsymbol{\lambda}_k$ we first compute

  $$\mathbf{x}_k = \underset{\mathbf{x} \in \mathcal{Q}}{\operatorname{argmin}} \{f_0(\mathbf{x}) + \boldsymbol{\lambda}_k^\top \mathbf{f}(\mathbf{x})\}$$

  and then set the subgradient as $\mathbf{g}(\boldsymbol{\lambda}_k) = -\mathbf{f}(\mathbf{x}_k)$
- The new update is then

  $$\boldsymbol{\lambda}_{k+1} = \Pi_{\mathbb{R}_+^m}[\boldsymbol{\lambda}_k - \eta_k \mathbf{g}_k] = [\boldsymbol{\lambda}_k - \eta_k \mathbf{g}_k]_+ = [\boldsymbol{\lambda}_k + \eta_k \mathbf{f}(\mathbf{x}_k)]_+$$

The process stops when $\|\mathbf{f}(\mathbf{x}_k)\| = \mathbf{0}$. Why? Cause if it happens, we have reached the optimal solution!

Proposition 1. *Suppose* $\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{Q}} \{f_0(\mathbf{x}) + \hat{\boldsymbol{\lambda}}^\top \mathbf{f}(\mathbf{x})\}$ *and* $\|\mathbf{f}(\hat{\mathbf{x}})\| = 0$*. Then,* $\hat{\mathbf{x}}$ *is an optimal solution.*

Proof. It is trivial to show that $\hat{\mathbf{x}}$ is a feasible point of the problem and $\mathbf{f}(\hat{\mathbf{x}}) \leq \mathbf{0}$. Now, for any other feasible point $\mathbf{x}$ of our main problem we have

$$\begin{aligned} f_0(\mathbf{x}) &\geq f_0(\mathbf{x}) + \hat{\boldsymbol{\lambda}}^\top \mathbf{f}(\mathbf{x}) \\ &\geq \min_{\mathbf{x} \in \mathcal{Q}} \{f_0(\mathbf{x}) + \hat{\boldsymbol{\lambda}}^\top \mathbf{f}(\mathbf{x})\} \\ &= f_0(\hat{\mathbf{x}}) + \hat{\boldsymbol{\lambda}}^\top \mathbf{f}(\hat{\mathbf{x}}) \\ &\geq f_0(\hat{\mathbf{x}}) \end{aligned}$$

leading to the optimality of $\hat{\mathbf{x}}$. □

3 Convergence Analysis

In this section, we analyze the dual projected subgradient method for the case that the stepsize is selected as $\eta_k = \frac{1}{\|\mathbf{f}(\mathbf{x}_k)\|_2 \sqrt{k+1}}$.

Theorem 2. *Consider the optimization problem in* (1) *and the update of dual projected subgradient methods with stepsize* $\eta_k = \frac{1}{\|\mathbf{f}(\mathbf{x}_k)\|_2 \sqrt{k+1}}$*. If for any point* $\mathbf{x} \in \mathcal{Q}$ *we have* $\|\mathbf{f}(\mathbf{x})\| \leq G$*, then for any* $\rho > 0$ *we have*

$$f_0(\tilde{\mathbf{x}}_k) - f_0(\mathbf{x}^*) + \rho \|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\|_2 \leq G \frac{(\|\boldsymbol{\lambda}_0\| + \rho)^2 + 1 + \log(k+1)}{2\sqrt{k+1}} \tag{3}$$

where $\tilde{\mathbf{x}}_k = \sum_{i=0}^k \frac{\eta_i}{\sum_{i=0}^k \eta_i} \mathbf{x}_i$.

Proof. Consider $\hat{\boldsymbol{\lambda}} \in \mathbb{R}^m_+$. Based, on the update of the algorithm we have

$$\begin{aligned}
\|\boldsymbol{\lambda}_{k+1} - \hat{\boldsymbol{\lambda}}\|_2^2 &= \left\| [\boldsymbol{\lambda}_k + \eta_k \mathbf{f}(\mathbf{x}_k)]_+ - [\hat{\boldsymbol{\lambda}}]_+ \right\|_2^2 \\
&\leq \left\| \boldsymbol{\lambda}_k + \eta_k \mathbf{f}(\mathbf{x}_k) - \hat{\boldsymbol{\lambda}} \right\|_2^2 \\
&= \left\| \boldsymbol{\lambda}_k - \hat{\boldsymbol{\lambda}} \right\|_2^2 + 2\eta_k \mathbf{f}(\mathbf{x}_k)^\top (\boldsymbol{\lambda}_k - \hat{\boldsymbol{\lambda}}) + \eta_k^2 \|\mathbf{f}(\mathbf{x}_k)\|_2^2
\end{aligned}$$

Now by summing up both sides we obtain

$$\|\boldsymbol{\lambda}_{k+1} - \hat{\boldsymbol{\lambda}}\|_2^2 \leq \left\| \boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}} \right\|_2^2 + 2\sum_{i=0}^{k} \eta_i \mathbf{f}(\mathbf{x}_i)^\top (\boldsymbol{\lambda}_i - \hat{\boldsymbol{\lambda}}) + \sum_{i=0}^{k} \eta_i^2 \|\mathbf{f}(\mathbf{x}_i)\|_2^2$$

Now note that

$$f_0(\mathbf{x}^*) \geq f_0(\mathbf{x}^*) + \boldsymbol{\lambda}_i^\top \mathbf{f}(\mathbf{x}^*) \geq f_0(\mathbf{x}_i) + \boldsymbol{\lambda}_i^\top \mathbf{f}(\mathbf{x}_i)$$

where the second inequality follows from the fact that

$$\mathbf{x}_i = \operatorname*{argmin}_{\mathbf{x} \in \mathcal{Q}} \{ f_0(\mathbf{x}) + \boldsymbol{\lambda}_i^\top \mathbf{f}(\mathbf{x}) \}$$

Considering the above bounds we have

$$\|\boldsymbol{\lambda}_{k+1} - \hat{\boldsymbol{\lambda}}\|_2^2 \leq \left\| \boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}} \right\|_2^2 + 2\sum_{i=0}^{k} \eta_i (f_0(\mathbf{x}^*) - f_0(\mathbf{x}_i)) - 2\sum_{i=0}^{k} \eta_i \mathbf{f}(\mathbf{x}_i)^\top \hat{\boldsymbol{\lambda}} + \sum_{i=0}^{k} \eta_i^2 \|\mathbf{f}(\mathbf{x}_i)\|_2^2$$

Hence,

$$2\sum_{i=0}^{k} \eta_i (f_0(\mathbf{x}_i) - f_0(\mathbf{x}^*)) + 2\sum_{i=0}^{k} \eta_i \mathbf{f}(\mathbf{x}_i)^\top \hat{\boldsymbol{\lambda}} \leq \left\| \boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}} \right\|_2^2 + \sum_{i=0}^{k} \eta_i^2 \|\mathbf{f}(\mathbf{x}_i)\|_2^2$$

Let's divide both sides by $2\sum_{i=0}^{k} \eta_i$ to obtain

$$\sum_{i=0}^{k} \frac{\eta_i}{\sum_{i=0}^{k} \eta_i} (f_0(\mathbf{x}_i) - f_0(\mathbf{x}^*)) + \sum_{i=0}^{k} \frac{\eta_i}{\sum_{i=0}^{k} \eta_i} \mathbf{f}(\mathbf{x}_i)^\top \hat{\boldsymbol{\lambda}} \leq \frac{\left\| \boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}} \right\|_2^2 + \sum_{i=0}^{k} \eta_i^2 \|\mathbf{f}(\mathbf{x}_i)\|_2^2}{2\sum_{i=0}^{k} \eta_i}$$

Now by the convexity of the functions f_0 and $f_1, \ldots, f_m$ we have

$$f_0(\tilde{\mathbf{x}}_k) - f_0(\mathbf{x}^*) + \mathbf{f}(\tilde{\mathbf{x}}_k)^\top \hat{\boldsymbol{\lambda}} \leq \frac{\left\| \boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}} \right\|_2^2 + \sum_{i=0}^{k} \eta_i^2 \|\mathbf{f}(\mathbf{x}_i)\|_2^2}{2\sum_{i=0}^{k} \eta_i}$$

where $\tilde{\mathbf{x}}_k = \sum_{i=0}^{k} \frac{\eta_i}{\sum_{i=0}^{k} \eta_i} \mathbf{x}_i$. Now if we set $\hat{\boldsymbol{\lambda}}$ as

$$\hat{\boldsymbol{\lambda}} := \begin{cases} \rho \frac{[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+}{\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\|_2}, & \text{if} \quad [\mathbf{f}(\tilde{\mathbf{x}}_k)]_+ \neq \mathbf{0} \\ \mathbf{0} & \text{if} \quad [\mathbf{f}(\tilde{\mathbf{x}}_k)]_+ = \mathbf{0} \end{cases}$$

we obtain that

$$f_0(\tilde{\mathbf{x}}_k) - f_0(\mathbf{x}^*) + \|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\|_2 \leq \frac{\left\| \boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}} \right\|_2^2 + \sum_{i=0}^{k} \eta_i^2 \|\mathbf{f}(\mathbf{x}_i)\|_2^2}{2\sum_{i=0}^{k} \eta_i}$$

Now using the fact that $\left\|\boldsymbol{\lambda}_0 - \hat{\boldsymbol{\lambda}}\right\| \leq \|\boldsymbol{\lambda}_0\| + \|\hat{\boldsymbol{\lambda}}\|$, and considering the fact that $\|\hat{\boldsymbol{\lambda}}\| \leq \rho$ we can show that

$$f_0(\tilde{\mathbf{x}}_k) - f_0(\mathbf{x}^*) + \|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\|_2 \leq \frac{(\|\boldsymbol{\lambda}_0\| + \rho)^2 + \sum_{i=0}^k \eta_i^2 \|\mathbf{f}(\mathbf{x}_i)\|_2^2}{2\sum_{i=0}^k \eta_i}$$

Now by replacing η_i by its expression and using the fact that $\|\mathbf{f}(\mathbf{x}_k)\|_2 \leq G$ we have

$$f_0(\tilde{\mathbf{x}}_k) - f_0(\mathbf{x}^*) + \|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\|_2 \leq \frac{(\|\boldsymbol{\lambda}_0\| + \rho)^2 + \sum_{i=0}^k \frac{1}{i+1}}{2\sum_{i=0}^k \frac{1}{\|\mathbf{f}(\mathbf{x}_i)\|_2\sqrt{i+1}}} \leq G\frac{(\|\boldsymbol{\lambda}_0\| + \rho)^2 + \sum_{i=0}^k \frac{1}{i+1}}{2\sum_{i=0}^k \frac{1}{\sqrt{i+1}}}$$

Now using the fact that for any positive a we have

$$\frac{a + \sum_{i=0}^k \frac{1}{i+1}}{\sum_{i=0}^k \frac{1}{\sqrt{i+1}}} \leq \frac{a + 1 + \log(k+1)}{\sqrt{k+1}}$$

we obtain that

$$f_0(\tilde{\mathbf{x}}_k) - f_0(\mathbf{x}^*) + \rho\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\|_2 \leq G\frac{(\|\boldsymbol{\lambda}_0\| + \rho)^2 + 1 + \log(k+1)}{2\sqrt{k+1}}$$

□

Note that the above result immediately implies that for any $\rho > 0$ that

$$f_0(\tilde{\mathbf{x}}_k) - f_0(\mathbf{x}^*) \leq \mathcal{O}(1/\sqrt{k})$$

However, the same conclusion **does not hold** for $\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\|_2$, as $f_0(\tilde{\mathbf{x}}_k) - f_0(\mathbf{x}^*)$ could be possibly negative. This issue can be addressed by proper selection of the free parameter ρ, the result in Corollary 1, and the following lemma.

Lemma 1. *Suppose for point $\tilde{\mathbf{x}}$ we have*

$$f_0(\tilde{\mathbf{x}}_k) - f_0(\mathbf{x}^*) + \rho\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\|_2 \leq \delta$$

if $\rho \geq 2\|\boldsymbol{\lambda}^*\|$, where $\boldsymbol{\lambda}^*$ is an optimal solution of the dual problem, then we have

$$\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\| \leq \frac{\delta}{(\rho - \|\boldsymbol{\lambda}^*\|_2)} \leq \frac{2\delta}{\rho}$$

Proof. Consider the function

$$v(\mathbf{u}) = \min_{\mathbf{x}\in\mathcal{Q}}\{f_0(\mathbf{x}),\ \mathbf{f}(\mathbf{x}) \leq \mathbf{u}\}$$

Indeed, it can be easily verified that $v(\mathbf{0}) = f^*$ and $v(\mathbf{u}) \leq f^*$ when $\mathbf{u} \geq \mathbf{0}$. Now note that $\boldsymbol{\lambda}^*$ is an optimal solution for the dual problem, therefore we have

$$\begin{aligned}
v(\mathbf{0}) - (\boldsymbol{\lambda}^*)^\top\mathbf{u} &= f^* - (\boldsymbol{\lambda}^*)^\top\mathbf{u} \\
&= \min_{\mathbf{x}\in\mathcal{Q}}\{f_0(\mathbf{x}) + \mathbf{f}(\mathbf{x})^\top\boldsymbol{\lambda}^*\} - (\boldsymbol{\lambda}^*)^\top\mathbf{u} \\
&\leq f_0(\hat{\mathbf{x}}) + (\boldsymbol{\lambda}^*)^\top\mathbf{f}(\hat{\mathbf{x}}) - (\boldsymbol{\lambda}^*)^\top\mathbf{u} \\
&\leq f_0(\hat{\mathbf{x}})
\end{aligned}$$

where $\hat{\mathbf{x}}$ is any point in $\mathcal{Q}$ such that $\mathbf{f}(\hat{\mathbf{x}}) \leq \mathbf{u}$. Hence, for any point in $\mathcal{Q}$ such that $\mathbf{f}(\hat{\mathbf{x}}) \leq \mathbf{u}$ we have

$$v(\mathbf{0}) - (\boldsymbol{\lambda}^*)^\top \mathbf{u} \leq f_0(\hat{\mathbf{x}})$$

By minimizing the right hand side with respect to $\hat{\mathbf{x}}$ we obtain

$$v(\mathbf{0}) - (\boldsymbol{\lambda}^*)^\top \mathbf{u} \leq v(\mathbf{u})$$

Therefore,

$$v(\mathbf{u}) \geq v(\mathbf{0}) - (\boldsymbol{\lambda}^*)^\top \mathbf{u}$$

which implies that $-\boldsymbol{\lambda}^* \in \partial v(\mathbf{0})$.

Now by setting $\mathbf{u} = [\mathbf{f}(\tilde{\mathbf{x}}_k)]_+$ we obtain that

$$\begin{aligned}
(\rho - \|\boldsymbol{\lambda}^*\|_2)\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\| &= \rho\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\| - \|\boldsymbol{\lambda}^*\|_2\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\| \\
&\leq \rho\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\| - (\boldsymbol{\lambda}^*)^\top [\mathbf{f}(\tilde{\mathbf{x}}_k)]_+ \\
&\leq \rho\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\| + v([\mathbf{f}(\tilde{\mathbf{x}}_k)]_+) - v(\mathbf{0}) \\
&\leq \rho\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\| + f_0(\tilde{\mathbf{x}}_k) - f^* \\
&\leq \delta.
\end{aligned}$$

Now, since $\rho \geq 2\|\boldsymbol{\lambda}^*\|_2$ we have

$$\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\| \leq \frac{\delta}{(\rho - \|\boldsymbol{\lambda}^*\|_2)} \leq \frac{2\delta}{\rho}$$

and the claim follows. □

Considering the above results we obtain that if we select

$$\rho = 2\left(\frac{f_0(\hat{\mathbf{x}}) - f^*}{\min_{i=1,\ldots,m} \; -f_i(\hat{\mathbf{x}})}\right) \geq 2\|\boldsymbol{\lambda}^*\|_2$$

then we can conclude that

$$\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\| \leq \frac{1}{\sqrt{k+1}}$$

Corollary 2. *Define* $\alpha := \left(\frac{f_0(\hat{\mathbf{x}}) - f^*}{\min_{i=1,\ldots,m} \; -f_i(\hat{\mathbf{x}})}\right)$. *Then, by setting* $\rho = 2\alpha$ *in* (3) *and considering the result of above lemma we obtain*

$$f_0(\tilde{\mathbf{x}}_k) - f_0(\mathbf{x}^*) \leq G\frac{(\|\boldsymbol{\lambda}_0\| + 2\alpha)^2 + 1 + \log(k+1)}{2\sqrt{k+1}}$$

and

$$\|[\mathbf{f}(\tilde{\mathbf{x}}_k)]_+\|_2 \leq \frac{G}{\alpha}\frac{(\|\boldsymbol{\lambda}_0\| + 2\alpha)^2 + 1 + \log(k+1)}{2\sqrt{k+1}}$$