Supplementary Material of Fast Proximal Linearized Alternating Direction Method of Multiplier with Parallel Splitting

Canyi Lu[1], Huan Li[2], Zhouchen Lin[2,3,*], Shuicheng Yan[1]
[1] Department of Electrical and Computer Engineering, National University of Singapore
[2] Key Laboratory of Machine Perception (MOE), School of EECS, Peking University
[3] Cooperative Medianet Innovation Center, Shanghai Jiaotong University
canyilu@gmail.com, lihuan_ss@126.com, zlin@pku.edu.cn, eleyans@nus.edu.sg

This documents provides the proof details of the convergence results of our proposed fast methods. First, in Section 1, we give some useful results which are useful for the convergence analysis of Fast PALM in Section 2 and Fast PL-ADMM-PS in Section 3.

1. Some Lemmas

Lemma 1 *[2] Let $g : \mathbb{R}^m \to \mathbb{R}$ be a continuously differentiable function with Lipschitz continuous gradient and Lipschitz constant L. Then, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$,*

$$g(\mathbf{x}) \leq g(\mathbf{y}) + \langle \mathbf{x} - \mathbf{y}, \nabla g(\mathbf{y}) \rangle + \frac{L}{2}||\mathbf{x} - \mathbf{y}||^2. \tag{1}$$

Lemma 2 *Given any $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{d} \in \mathbb{R}^m$, we have*

$$\langle \boldsymbol{a} - \boldsymbol{b}, \boldsymbol{a} - \boldsymbol{c} \rangle = \frac{1}{2} \left(\|\boldsymbol{a} - \boldsymbol{b}\|^2 + \|\boldsymbol{a} - \boldsymbol{c}\|^2 - \|\boldsymbol{b} - \boldsymbol{c}\|^2 \right). \tag{2}$$

$$\langle \boldsymbol{a} - \boldsymbol{b}, \boldsymbol{c} - \boldsymbol{d} \rangle = \frac{1}{2} \left(\|\boldsymbol{a} - \boldsymbol{d}\|^2 - \|\boldsymbol{a} - \boldsymbol{c}\|^2 - \|\boldsymbol{b} - \boldsymbol{d}\|^2 + \|\boldsymbol{b} - \boldsymbol{c}\|^2 \right). \tag{3}$$

Lemma 3 *Assume the sequences $\{a^{(k)}\}$ and $\{b^{(k)}\}$ satisfy $a^{(0)} = 1$, $0 < a^{(k+1)} - a^{(k)} \leq 1$ and $b^{(k)} > 0$. Then we have*

$$\sum_{k=0}^{K} a^{(k)}(b^{(k)} - b^{(k+1)}) \leq \sum_{k=0}^{K} b^{(k)}. \tag{4}$$

Proof. We deduce

$$\begin{aligned}
\sum_{k=0}^{K} a^{(k)}(b^{(k)} - b^{(k+1)}) &= a^{(0)}b^{(0)} + \sum_{k=0}^{K-1}(a^{(k+1)} - a^{(k)})b^{(k+1)} - a^{(K)}b^{(K+1)} \\
&\leq b^{(0)} + \sum_{k=0}^{K-1} b^{(k+1)} = \sum_{k=0}^{K} b^{(k)}.
\end{aligned}$$

■

Lemma 4 *Define the sequence $\{\theta^{(k)}\}$ as $\theta^{(0)} = 1$, $\frac{1-\theta^{(k+1)}}{(\theta^{(k+1)})^2} = \frac{1}{(\theta^{(k)})^2}$ and $\theta^{(k)} > 0$. Then we have the following properties*

$$\theta^{(k+1)} = \frac{-(\theta^{(k)})^2 + \sqrt{(\theta^{(k)})^4 + 4(\theta^{(k)})^2}}{2}, \tag{5}$$

*Corresponding author.

$$\sum_{k=0}^{K} \frac{1}{\theta^{(k)}} = \frac{1}{(\theta^{(K)})^2}, \tag{6}$$

$$0 < \frac{1}{\theta^{(k+1)}} - \frac{1}{\theta^{(k)}} < 1, \tag{7}$$

$$\theta^{(k)} \leq \frac{2}{k+2}, \tag{8}$$

and

$$\theta^{(k)} \leq 1. \tag{9}$$

Proof. From the definition of $\theta^{(k+1)}$, it is easy to get that $\theta^{(k+1)} = \frac{-(\theta^{(k)})^2+\sqrt{(\theta^{(k)})^4+4(\theta^{(k)})^2}}{2}$. This implies that $\theta^{(k)}$ is well defined for any $k \geq 0$. Furthermore, since $\frac{1}{\theta^{(k+1)}} = \frac{1}{(\theta^{(k+1)})^2} - \frac{1}{(\theta^{(k)})^2}$ and $\theta^{(0)} = 1$, we have

$$\sum_{k=0}^{K} \frac{1}{\theta^{(k)}} = \frac{1}{\theta^{(0)}} + \sum_{k=0}^{K-1} \frac{1}{\theta^{(k+1)}} = \frac{1}{\theta^{(0)}} + \sum_{k=0}^{K-1} \left(\frac{1}{(\theta^{(k+1)})^2} - \frac{1}{(\theta^{(k)})^2} \right) = \frac{1}{\theta^{(0)}} + \frac{1}{(\theta^{(K)})^2} - \frac{1}{(\theta^{(0)})^2} = \frac{1}{(\theta^{(K)})^2}. \tag{10}$$

From $\frac{1}{\theta^{(k+1)}} = \frac{1}{(\theta^{(k+1)})^2} - \frac{1}{(\theta^{(k)})^2}$, $\theta^{(k)} > 0$ and $\theta^{(k-1)} > 0$, we can easily get

$$\frac{1}{\theta^{(k+1)}} - \frac{1}{\theta^{(k)}} > 0, \tag{11}$$

and

$$\frac{1}{\theta^{(k+1)}} - \frac{1}{\theta^{(k)}} = \frac{1}{\theta^{(k+1)}} - \frac{\sqrt{1-\theta^{(k+1)}}}{\theta^{(k+1)}} = \frac{1-\sqrt{1-\theta^{(k+1)}}}{\theta^{(k+1)}} = \frac{1}{1+\sqrt{1-\theta^{(k+1)}}} < 1. \tag{12}$$

Next we proof $\theta^{(k)} \leq \frac{2}{k+2}$ by induction. First $\theta^{(0)} = 1 \leq \frac{2}{0+2}$. Now assume that $\theta^{(k)} \leq \frac{2}{k+2}$ and we prove $\theta^{(k+1)} \leq \frac{2}{k+3}$. We deduce

$$\begin{aligned}\theta^{(k+1)} &= \frac{-(\theta^{(k)})^2 + \sqrt{(\theta^{(k)})^4 + 4(\theta^{(k)})^2}}{2} = \frac{2(\theta^{(k)})^2}{(\theta^{(k)})^2 + \sqrt{(\theta^{(k)})^4 + 4(\theta^{(k)})^2}} \\ &= \frac{2}{1+\sqrt{1+\frac{4}{(\theta^{(k)})^2}}} \leq \frac{2}{1+\sqrt{1+(k+2)^2}} \leq \frac{2}{k+3}.\end{aligned}$$

So (8) holds. Note that $\theta^{(k)}$ is decreasing by (7) and $\theta^{(0)} = 1$, we have (9). The proof is completed. ∎

2. Convergence Analysis of Fast PALM

In this section, we give the convergence analysis of Fast PALM for solving the following problem

$$\min_{\mathbf{x}} f(\mathbf{x}), \quad s.t. \quad \mathcal{A}(\mathbf{x}) = \mathbf{b}, \tag{13}$$

where $f(\mathbf{x}) = g(\mathbf{x}) + h(\mathbf{x})$, both g and h are convex, and $g \in C^{1.1}$:

$$\|\nabla g(\mathbf{x}) - \nabla g(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y}. \tag{14}$$

For the completeness, we give the Fast PALM in Algorithm 1.

It is worth mentioning that the definition of $\theta^{(k+1)}$ in (20) is equivalent to $\theta^{(0)} = 1$, $\frac{1-\theta^{(k+1)}}{(\theta^{(k+1)})^2} = \frac{1}{(\theta^{(k)})^2}$ and $\theta^{(k)} > 0$ in Lemma 4. Such a property will be used in the following analysis several times.

The anaysis of our algorithms is based on the following property:

Lemma 5 *[1] $\tilde{\mathbf{x}}$ is an optimal solution to (13) if and only if there exists $\alpha > 0$, such that*

$$f(\tilde{\mathbf{x}}) - f(\mathbf{x}^*) + \langle \lambda^*, \mathcal{A}(\tilde{\mathbf{x}}) - \mathbf{b} \rangle + \frac{\alpha}{2} \|\mathcal{A}(\tilde{\mathbf{x}}) - \boldsymbol{b}\|^2 = 0. \tag{15}$$

Initialize: $\mathbf{x}^0$, $\mathbf{z}^0$, $\boldsymbol{\lambda}^0$, $\beta^{(0)} = \theta^{(0)} = 1$.

for $k = 0, 1, 2, \cdots$ **do**

$$\begin{aligned}
\mathbf{y}^{k+1} &= (1-\theta^{(k)})\mathbf{x}^k + \theta^{(k)}\mathbf{z}^k; && (16)\\
\mathbf{z}^{k+1} &= \operatorname*{argmin}_{\mathbf{x}} g(\mathbf{y}^{k+1}) + \left\langle \nabla g(\mathbf{y}^{k+1}), \mathbf{x} - \mathbf{y}^{k+1} \right\rangle + h(\mathbf{x}) \\
&\quad + \left\langle \boldsymbol{\lambda}^k, \mathcal{A}(\mathbf{x}) - \mathbf{b} \right\rangle + \frac{\beta^{(k)}}{2}\|\mathcal{A}(\mathbf{x}) - \mathbf{b}\|^2 + \frac{L\theta^{(k)}}{2}\|\mathbf{x} - \mathbf{z}^k\|^2; && (17)\\
\mathbf{x}^{k+1} &= (1-\theta^{(k)})\mathbf{x}^k + \theta^{(k)}\mathbf{z}^{k+1}; && (18)\\
\boldsymbol{\lambda}^{k+1} &= \boldsymbol{\lambda}^k + \beta^{(k)}(\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}); && (19)\\
\theta^{(k+1)} &= \frac{-(\theta^{(k)})^2 + \sqrt{(\theta^{(k)})^4 + 4(\theta^{(k)})^2}}{2}; && (20)\\
\beta^{(k+1)} &= \frac{1}{\theta^{(k+1)}}. && (21)
\end{aligned}$$

end

Algorithm 1: Fast PALM Algorithm

Proposition 1 *In Algorithm 1, for any* $\mathbf{x}$*, we have*

$$\begin{aligned}
&\frac{1-\theta^{(k+1)}}{(\theta^{(k+1)})^2}\left(f(\mathbf{x}^{k+1}) - f(\mathbf{x})\right) - \frac{1}{\theta^{(k)}}\left\langle \mathcal{A}^T(\boldsymbol{\lambda}^{k+1}), \mathbf{x} - \mathbf{z}^{k+1} \right\rangle \\
\leq\ &\frac{1-\theta^{(k)}}{(\theta^{(k)})^2}\left(f(\mathbf{x}^k) - f(\mathbf{x})\right) + \frac{L}{2}\left(\|\mathbf{z}^k - \mathbf{x}\|^2 - \|\mathbf{z}^{k+1} - \mathbf{x}\|^2\right). \qquad (22)
\end{aligned}$$

Proof. From the optimality of $\mathbf{z}^{k+1}$ to (17), we have

$$\begin{aligned}
0 &\in \partial h(\mathbf{z}^{k+1}) + \nabla g(\mathbf{y}^{k+1}) + \mathcal{A}^T(\boldsymbol{\lambda}^k) + \beta^{(k)}\mathcal{A}^T(\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}) + L\theta^{(k)}(\mathbf{z}^{k+1} - \mathbf{z}^k) \\
&= \partial h(\mathbf{z}^{k+1}) + \nabla g(\mathbf{y}^{k+1}) + \mathcal{A}^T(\boldsymbol{\lambda}^{k+1}) + L\theta^{(k)}(\mathbf{z}^{k+1} - \mathbf{z}^k), \qquad (23)
\end{aligned}$$

where (23) uses (19). From the convexity of h, we have

$$h(\mathbf{x}) - h(\mathbf{z}^{k+1}) \geq \left\langle -\nabla g(\mathbf{y}^{k+1}) - \mathcal{A}^T(\boldsymbol{\lambda}^{k+1}) - L\theta^{(k)}(\mathbf{z}^{k+1} - \mathbf{z}^k), \mathbf{x} - \mathbf{z}^{k+1} \right\rangle. \qquad (24)$$

On the other hand,

$$\begin{aligned}
f(\mathbf{x}^{k+1}) &\leq g(\mathbf{y}^{k+1}) + \left\langle \nabla g(\mathbf{y}^{k+1}), \mathbf{x}^{k+1} - \mathbf{y}^{k+1} \right\rangle + \frac{L}{2}\|\mathbf{x}^{k+1} - \mathbf{y}^{k+1}\|^2 + h(\mathbf{x}^{k+1}) \quad (25)\\
&= g(\mathbf{y}^{k+1}) + \left\langle \nabla g(\mathbf{y}^{k+1}), (1-\theta^{(k)})\mathbf{x}^k + \theta^{(k)}\mathbf{z}^{k+1} - \mathbf{y}^{k+1} \right\rangle \\
&\quad + \frac{L}{2}\|(1-\theta^{(k)})\mathbf{x}^k + \theta^{(k)}\mathbf{z}^{k+1} - \mathbf{y}^{k+1}\|^2 + h\left((1-\theta^{(k)})\mathbf{x}^k + \theta^{(k)}\mathbf{z}^{k+1}\right) \quad (26)\\
&\leq (1-\theta^{(k)})\left(g(\mathbf{y}^{k+1}) + \left\langle \nabla g(\mathbf{y}^{k+1}), \mathbf{x}^k - \mathbf{y}^{k+1} \right\rangle + h(\mathbf{x}^k)\right)\\
&\quad + \theta^{(k)}\left(g(\mathbf{y}^{k+1}) + \left\langle \nabla g(\mathbf{y}^{k+1}), \mathbf{z}^{k+1} - \mathbf{y}^{k+1} \right\rangle + h(\mathbf{z}^{k+1})\right) + \frac{L(\theta^{(k)})^2}{2}\|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 \quad (27)\\
&= (1-\theta^{(k)})\left(g(\mathbf{y}^{k+1}) + \left\langle \nabla g(\mathbf{y}^{k+1}), \mathbf{x}^k - \mathbf{y}^{k+1} \right\rangle + h(\mathbf{x}^k)\right)\\
&\quad + \theta^{(k)}\left(g(\mathbf{y}^{k+1}) + \left\langle \nabla g(\mathbf{y}^{k+1}), \mathbf{x} - \mathbf{y}^{k+1} \right\rangle + \left\langle \nabla g(\mathbf{y}^{k+1}), \mathbf{z}^{k+1} - \mathbf{x} \right\rangle + h(\mathbf{z}^{k+1})\right)\\
&\quad + \frac{L(\theta^{(k)})^2}{2}\|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2\\
&\leq (1-\theta^{(k)})f(\mathbf{x}^k) + \theta^{(k)}\left(g(\mathbf{x}) + \left\langle \nabla g(\mathbf{y}^{k+1}), \mathbf{z}^{k+1} - \mathbf{x} \right\rangle + h(\mathbf{z}^{k+1})\right) + \frac{L(\theta^{(k)})^2}{2}\|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 \quad (28)\\
&\leq (1-\theta^{(k)})f(\mathbf{x}^k) + \theta^{(k)}\left(g(\mathbf{x}) + h(\mathbf{x}) + \left\langle \mathcal{A}^T(\boldsymbol{\lambda}^{k+1}) + L\theta^{(k)}(\mathbf{z}^{k+1} - \mathbf{z}^k), \mathbf{x} - \mathbf{z}^{k+1} \right\rangle\right) \quad (29)\\
&\quad + \frac{L(\theta^{(k)})^2}{2}\|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2\\
&= (1-\theta^{(k)})f(\mathbf{x}^k) + \theta^{(k)}f(\mathbf{x}) + \theta^{(k)}\left\langle \mathcal{A}^T(\boldsymbol{\lambda}^{k+1}), \mathbf{x} - \mathbf{z}^{k+1} \right\rangle - \frac{L(\theta^{(k)})^2}{2}\left(\|\mathbf{z}^{k+1} - \mathbf{x}\|^2 - \|\mathbf{z}^k - \mathbf{x}\|^2\right), \quad (30)
\end{aligned}$$

where (25) uses (1), (26) uses (18), (27) is from the convexity of h, (28) is from the convexity of g, (29) uses (24) and (30) uses (2). Rerangging the above inequality leads to

$$\begin{aligned}
&\left(f(\mathbf{x}^{k+1}) - f(\mathbf{x})\right) - \theta^{(k)}\left\langle \mathcal{A}^T(\boldsymbol{\lambda}^{k+1}), \mathbf{x} - \mathbf{z}^{k+1} \right\rangle\\
\leq\ & (1-\theta^{(k)})\left(f(\mathbf{x}^k) - f(\mathbf{x})\right) + \frac{L(\theta^{(k)})^2}{2}\left(\|\mathbf{z}^k - \mathbf{x}\|^2 - \|\mathbf{z}^{k+1} - \mathbf{x}\|^2\right). \quad (31)
\end{aligned}$$

Diving both sides of the above inequality by $(\theta^{(k)})^2$ leads to

$$\begin{aligned}
&\frac{1}{(\theta^{(k)})^2}\left(f(\mathbf{x}^{k+1}) - f(\mathbf{x})\right) - \frac{1}{\theta^{(k)}}\left\langle \mathcal{A}^T(\boldsymbol{\lambda}^{k+1}), \mathbf{x} - \mathbf{z}^{k+1} \right\rangle\\
\leq\ & \frac{1-\theta^{(k)}}{(\theta^{(k)})^2}\left(f(\mathbf{x}^k) - f(\mathbf{x})\right) + \frac{L}{2}\left(\|\mathbf{z}^k - \mathbf{x}\|^2 - \|\mathbf{z}^{k+1} - \mathbf{x}\|^2\right).
\end{aligned}$$

The proof is completed by using the property of $\theta^{(k)}$ in Lemma 4. ■

Proposition 2 *In Algorithm 1, the following result holds for any* $\boldsymbol{\lambda}$

$$\begin{aligned}
&\langle \mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}, \boldsymbol{\lambda} - \boldsymbol{\lambda}^{k+1} \rangle + \frac{\beta^{(k)}}{2}\|\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\|^2\\
=\ & \frac{1}{2\beta^{(k)}}\left(\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}\|^2 - \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}\|^2\right) \quad (32)
\end{aligned}$$

Proof. By using (19) and (2), we have

$$\begin{aligned}
&\langle \mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}, \boldsymbol{\lambda} - \boldsymbol{\lambda}^{k+1} \rangle\\
=\ & \frac{1}{\beta^{(k)}}\langle \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k, \boldsymbol{\lambda} - \boldsymbol{\lambda}^{k+1} \rangle\\
=\ & \frac{1}{2\beta^{(k)}}\left(\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}\|^2 - \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}\|^2 - \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2\right)\\
=\ & \frac{1}{2\beta^{(k)}}\left(\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}\|^2 - \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}\|^2\right) - \frac{\beta^{(k)}}{2}\|\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\|^2.
\end{aligned}$$

The proof is completed. ■

Theorem 1 *In Algorithm 1, for any $K > 0$, we have*

$$f(\mathbf{x}^{K+1}) - f(\mathbf{x}^*) + \left\langle \boldsymbol{\lambda}^*, \mathcal{A}(\mathbf{x}^{K+1}) - \mathbf{b} \right\rangle + \frac{1}{2}\|\mathcal{A}(\mathbf{x}^{K+1}) - \mathbf{b}\|^2 \tag{33}$$

$$\leq \frac{2}{(K+2)^2}\left(LD_{\mathbf{x}^*}^2 + D_{\lambda^*}^2\right). \tag{34}$$

Proof. Let $\mathbf{x} = \mathbf{x}^*$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ in (22) and (32). We have

$$\frac{1-\theta^{(k+1)}}{(\theta^{(k+1)})^2}\left(f(\mathbf{x}^{k+1}) - f(\mathbf{x}^*)\right) - \frac{1-\theta^{(k)}}{(\theta^{(k)})^2}\left(f(\mathbf{x}^{k}) - f(\mathbf{x}^*)\right) + \frac{1}{\theta^{(k)}}\left\langle \boldsymbol{\lambda}^*, \mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b} \right\rangle \tag{35}$$

$$\leq \frac{L}{2}\left(\|\mathbf{z}^k - \mathbf{x}^*\|^2 - \|\mathbf{z}^{k+1} - \mathbf{x}^*\|^2\right) + \frac{1}{\theta^{(k)}}\langle \boldsymbol{\lambda}^* - \boldsymbol{\lambda}^{k+1}, \mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\rangle \tag{36}$$

$$\leq \frac{L}{2}\left(\|\mathbf{z}^k - \mathbf{x}^*\|^2 - \|\mathbf{z}^{k+1} - \mathbf{x}^*\|^2\right) + \frac{1}{2\theta^{(k)}\beta^{(k)}}\left(\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 - \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|^2\right) - \frac{\beta^{(k)}}{2\theta^{(k)}}\|\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\|^2 \tag{37}$$

$$= \frac{L}{2}\left(\|\mathbf{z}^k - \mathbf{x}^*\|^2 - \|\mathbf{z}^{k+1} - \mathbf{x}^*\|^2\right) + \frac{1}{2}\left(\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 - \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|^2\right) - \frac{1}{2(\theta^{(k)})^2}\|\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\|^2, \tag{38}$$

where (36) uses the fact $\mathcal{A}(\mathbf{x}^*) = \mathbf{b}$, (37) uses (32) and (38) uses $\beta^{(k)} = \frac{1}{\theta^{(k)}}$.

Summing (35)-(38) from $k = 0$ to K, we have

$$\frac{1-\theta^{(K+1)}}{(\theta^{(K+1)})^2}\left(f(\mathbf{x}^{K+1}) - f(\mathbf{x}^*)\right) - \frac{1-\theta^{(0)}}{(\theta^{(0)})^2}\left(f(\mathbf{x}^{0}) - f(\mathbf{x}^*)\right) + \sum_{k=0}^{K}\frac{1}{\theta^{(k)}}\left\langle \boldsymbol{\lambda}^*, \mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b} \right\rangle \tag{39}$$

$$\leq \frac{L}{2}\|\mathbf{z}^0 - \mathbf{x}^*\|^2 + \frac{1}{2}\|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|^2 - \sum_{k=0}^{K}\frac{1}{2(\theta^{(k)})^2}\|\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\|^2$$

$$\leq \frac{L}{2}\|\mathbf{z}^0 - \mathbf{x}^*\|^2 + \frac{1}{2}\|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|^2 - \sum_{k=0}^{K}\frac{1}{2\theta^{(k)}}\|\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\|^2, \tag{40}$$

where (40) uses (9). Also note that $\theta^{(0)} = 1$. So the second term of (39) disappears.

On the other hand, by the property of $\theta^{(k)}$ in Lemma 4 and (18), we have

$$\begin{aligned}
&\sum_{k=0}^{K}\frac{\mathbf{z}^{k+1}}{\theta^{(k)}} \\
= &\sum_{k=0}^{K}\left(\frac{1}{(\theta^{(k)})^2}\mathbf{x}^{k+1} - \frac{1-\theta^{(k)}}{(\theta^{(k)})^2}\mathbf{x}^k\right) \\
= &\sum_{k=0}^{K}\left(\frac{1-\theta^{(k+1)}}{(\theta^{(k+1)})^2}\mathbf{x}^{k+1} - \frac{1-\theta^{(k)}}{(\theta^{(k)})^2}\mathbf{x}^k\right) \\
= &\frac{1-\theta^{(K+1)}}{(\theta^{(K+1)})^2}\mathbf{x}^{K+1} - \frac{1-\theta^{(0)}}{(\theta^{(0)})^2}\mathbf{x}^0 \\
= &\frac{1-\theta^{(K+1)}}{(\theta^{(K+1)})^2}\mathbf{x}^{K+1} \\
= &\frac{1}{(\theta^{(K)})^2}\mathbf{x}^{K+1}.
\end{aligned} \tag{41}$$

So

$$\sum_{k=0}^{K}\frac{1}{\theta^{(k)}}\left\langle \boldsymbol{\lambda}^*, \mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b} \right\rangle = \frac{1}{(\theta^{(K)})^2}\left\langle \boldsymbol{\lambda}^*, \mathcal{A}(\mathbf{x}^{K+1}) - \mathbf{b} \right\rangle. \tag{42}$$

By the convexity of $\|\cdot\|^2$, we have

$$\begin{aligned}
& \sum_{k=0}^{K} \frac{1}{2\theta^{(k)}} \|\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\|^2 \\
= & \frac{1}{2(\theta^{(K)})^2} \sum_{k=0}^{K} \frac{\theta^{(K)})^2}{\theta^{(k)}} \|\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\|^2 \\
\geq & \frac{1}{2(\theta^{(K)})^2} \left\| (\theta^{(K)})^2 \mathcal{A}\left(\sum_{k=0}^{K} \frac{\mathbf{z}^{k+1}}{\theta^{(k)}} \right) - \mathbf{b} \right\|^2 \quad (43) \\
= & \frac{1}{2(\theta^{(K)})^2} \|\mathcal{A}(\mathbf{x}^{K+1}) - \mathbf{b}\|^2, \quad (44)
\end{aligned}$$

where (43) uses (6) and (44) uses (41).

Substituting (42) into (39) and (44) into (40) respectively, we obtain

$$\begin{aligned}
& \frac{1-\theta^{(K+1)}}{(\theta^{(K+1)})^2} \left(f(\mathbf{x}^{K+1}) - f(\mathbf{x}^*) \right) + \frac{1}{(\theta^{(K)})^2} \left\langle \boldsymbol{\lambda}^*, \mathcal{A}(\mathbf{x}^{K+1}) - \mathbf{b} \right\rangle + \frac{1}{2(\theta^{(K)})^2} \|\mathcal{A}(\mathbf{x}^{K+1}) - \mathbf{b}\|^2 \quad (45) \\
\leq & \frac{L}{2} \|\mathbf{z}^0 - \mathbf{x}^*\|^2 + \frac{1}{2} \|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|^2. \quad (46)
\end{aligned}$$

Multiplying (45) and (46) by $(\theta^{(K)})^2$ and using (8) leads to

$$\begin{aligned}
& f(\mathbf{x}^{K+1}) - f(\mathbf{x}^*) + \left\langle \boldsymbol{\lambda}^*, \mathcal{A}(\mathbf{x}^{K+1}) - \mathbf{b} \right\rangle + \frac{1}{2} \|\mathcal{A}(\mathbf{x}^{K+1}) - \mathbf{b}\|^2 \\
\leq & \frac{2}{(K+2)^2} \left(L\|\mathbf{z}^0 - \mathbf{x}^*\|^2 + \|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|^2 \right) \\
= & \frac{2}{(K+2)^2} \left(L D_{\mathbf{x}^*}^2 + D_{\lambda^*}^2 \right).
\end{aligned}$$

The proof is completed. ■

3. Convergence Analysis of Fast PL-ADMM-PS

In this section, we give the convergence analysis of Fast PL-ADMM-PS for solving the following problem

$$\min_{\mathbf{x}_1, \cdots, \mathbf{x}_n} \sum_{i=1}^{n} f_i(\mathbf{x}_i), \quad s.t. \quad \sum_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i) = \mathbf{b}, \quad (47)$$

where $f_i(\mathbf{x}_i) = g_i(\mathbf{x}_i) + h_i(\mathbf{x}_i)$, both g_i and h_i are convex, and $g_i \in C^{1.1}$. The whole procedure of Fast PL-ADMM-PS is shown in Algorithm 2.

Proposition 3 *In Algorithm 2, for any $\mathbf{x}_i$, we have*

$$\begin{aligned}
& \frac{1-\theta^{(k+1)}}{(\theta^{(k+1)})^2} \left(f_i(\mathbf{x}_i^{k+1}) - f_i(\mathbf{x}_i) \right) - \frac{1}{\theta^{(k)}} \left\langle \mathcal{A}_i^T(\hat{\boldsymbol{\lambda}}^{k+1}), \mathbf{x}_i - \mathbf{z}_i^{k+1} \right\rangle \\
\leq & \frac{1-\theta^{(k)}}{(\theta^{(k)})^2} \left(f_i(\mathbf{x}_i^k) - f_i(\mathbf{x}_i) \right) + \frac{L_i}{2} \left(\|\mathbf{z}_i^k - \mathbf{x}_i\|^2 - \|\mathbf{z}_i^{k+1} - \mathbf{x}_i\|^2 \right) \\
& + \frac{\beta^{(k)} \eta_i}{2\theta^{(k)}} \left(\|\mathbf{z}_i^k - \mathbf{x}_i\|^2 - \|\mathbf{z}_i^{k+1} - \mathbf{x}_i\|^2 - \|\mathbf{z}_i^{k+1} - \mathbf{z}_i^k\|^2 \right), \quad (53)
\end{aligned}$$

where

$$\hat{\boldsymbol{\lambda}}^{k+1} = \boldsymbol{\lambda}^k + \beta^{(k)} \left(\mathcal{A}(\mathbf{z}^k) - \mathbf{b} \right). \quad (54)$$

Initialize: $\mathbf{x}^0$, $\mathbf{z}^0$, $\boldsymbol{\lambda}^0$, $\theta^{(0)}=1$, fix $\beta^{(k)}=\beta$ for $k\geq 0$, $\eta_i > n\|\mathcal{A}_i\|^2$, $i=1,\cdots,n$,
for $k=0,1,2,\cdots$ **do**

// Update $\mathbf{y}_i$, $\mathbf{z}_i$, $\mathbf{x}_i$, $i=1,\cdots,n$, in parallel by

$$\mathbf{y}_i^{k+1} = (1-\theta^{(k)})\mathbf{x}_i^k + \theta^{(k)}\mathbf{z}_i^k; \tag{48}$$

$$\begin{aligned}\mathbf{z}_i^{k+1} = \operatorname*{argmin}_{\mathbf{x}_i} & \left\langle \nabla g_i(\mathbf{y}_i^{k+1}), \mathbf{x}_i\right\rangle + h_i(\mathbf{x}_i) + \left\langle \boldsymbol{\lambda}^k, \mathcal{A}_i(\mathbf{x}_i)\right\rangle + \left\langle \beta^{(k)}\mathcal{A}_i^T\left(\mathcal{A}(\mathbf{z}^k)-\mathbf{b}\right), \mathbf{x}_i\right\rangle \\ & + \frac{L(g_i)\theta^{(k)}+\beta^{(k)}\eta_i}{2}\|\mathbf{x}_i-\mathbf{z}_i^k\|^2;\end{aligned} \tag{49}$$

$$\mathbf{x}_i^{k+1} = (1-\theta^{(k)})\mathbf{x}_i^k + \theta^{(k)}\mathbf{z}_i^{k+1}; \tag{50}$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \beta^k\left(\mathcal{A}(\mathbf{z}^{k+1})-\mathbf{b}\right); \tag{51}$$

$$\theta^{(k+1)} = \frac{-(\theta^{(k)})^2+\sqrt{(\theta^{(k)})^4+4(\theta^{(k)})^2}}{2}. \tag{52}$$

end

Algorithm 2: Fast PL-ADMM-PS Algorithm

Proof. From the optimality of $\mathbf{z}_i^{k+1}$ to (49), we have

$$\begin{aligned} 0 &\in \partial h_i(\mathbf{z}_i^{k+1}) + \nabla g_i(\mathbf{y}_i^{k+1}) + \mathcal{A}_i^T(\boldsymbol{\lambda}^k) + \beta^{(k)}\mathcal{A}_i^T(\mathcal{A}(\mathbf{z}^k)-\mathbf{b}) + (L_i\theta^{(k)}+\beta^{(k)}\eta_i)(\mathbf{z}_i^{k+1}-\mathbf{z}_i^k) \\ &= \partial h_i(\mathbf{z}_i^{k+1}) + \nabla g_i(\mathbf{y}_i^{k+1}) + \mathcal{A}_i^T(\hat{\boldsymbol{\lambda}}^{k+1}) + (L_i\theta^{(k)}+\beta^{(k)}\eta_i)(\mathbf{z}_i^{k+1}-\mathbf{z}_i^k), \end{aligned} \tag{55}$$

where (55) uses (54). From the convexity of h_i, we have

$$h_i(\mathbf{x}_i) - h_i(\mathbf{z}_i^{k+1}) \geq \left\langle -\nabla g_i(\mathbf{y}_i^{k+1}) - \mathcal{A}_i^T(\hat{\boldsymbol{\lambda}}^{k+1}) - (L_i\theta^{(k)}+\beta^{(k)}\eta_i)(\mathbf{z}_i^{k+1}-\mathbf{z}_i^k), \mathbf{x}_i-\mathbf{z}_i^{k+1}\right\rangle. \tag{56}$$

On the other hand,

$$\begin{aligned}
f_i(\mathbf{x}_i^{k+1}) &\leq g_i(\mathbf{y}_i^{k+1}) + \left\langle \nabla g_i(\mathbf{y}_i^{k+1}), \mathbf{x}_i^{k+1}-\mathbf{y}_i^{k+1}\right\rangle + \frac{L_i}{2}\|\mathbf{x}_i^{k+1}-\mathbf{y}_i^{k+1}\|^2 + h_i(\mathbf{x}_i^{k+1}) && (57)\\
&= g_i(\mathbf{y}_i^{k+1}) + \left\langle \nabla g_i(\mathbf{y}_i^{k+1}), (1-\theta^{(k)})\mathbf{x}_i^k + \theta^{(k)}\mathbf{z}_i^{k+1}-\mathbf{y}_i^{k+1}\right\rangle \\
&\quad + \frac{L_i}{2}\|(1-\theta^{(k)})\mathbf{x}_i^k + \theta^{(k)}\mathbf{z}_i^{k+1}-\mathbf{y}_i^{k+1}\|^2 + h_i\left((1-\theta^{(k)})\mathbf{x}_i^k + \theta^{(k)}\mathbf{z}_i^{k+1}\right) && (58)\\
&\leq (1-\theta^{(k)})\left(g_i(\mathbf{y}_i^{k+1}) + \left\langle \nabla g_i(\mathbf{y}_i^{k+1}), \mathbf{x}_i^k-\mathbf{y}_i^{k+1}\right\rangle + h_i(\mathbf{x}_i^k)\right) \\
&\quad + \theta^{(k)}\left(g_i(\mathbf{y}_i^{k+1}) + \left\langle \nabla g_i(\mathbf{y}_i^{k+1}), \mathbf{z}_i^{k+1}-\mathbf{y}_i^{k+1}\right\rangle + h_i(\mathbf{z}_i^{k+1})\right) + \frac{L_i(\theta^{(k)})^2}{2}\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2 && (59)\\
&= (1-\theta^{(k)})\left(g_i(\mathbf{y}_i^{k+1}) + \left\langle \nabla g_i(\mathbf{y}_i^{k+1}), \mathbf{x}_i^k-\mathbf{y}_i^{k+1}\right\rangle + h_i(\mathbf{x}_i^k)\right) \\
&\quad + \theta^{(k)}\left(g_i(\mathbf{y}_i^{k+1}) + \left\langle \nabla g_i(\mathbf{y}_i^{k+1}), \mathbf{x}_i-\mathbf{y}_i^{k+1}\right\rangle + \left\langle \nabla g_i(\mathbf{y}_i^{k+1}), \mathbf{z}_i^{k+1}-\mathbf{x}_i\right\rangle + h_i(\mathbf{z}_i^{k+1})\right) \\
&\quad + \frac{L_i(\theta^{(k)})^2}{2}\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2 && (60)\\
&\leq (1-\theta^{(k)})f_i(\mathbf{x}_i^k) + \theta^{(k)}\left(g_i(\mathbf{x}_i) + \left\langle \nabla g_i(\mathbf{y}_i^{k+1}), \mathbf{z}_i^{k+1}-\mathbf{x}_i\right\rangle + h_i(\mathbf{z}_i^{k+1})\right) + \frac{L_i(\theta^{(k)})^2}{2}\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2 && (61)\\
&\leq (1-\theta^{(k)})f_i(\mathbf{x}_i^k) + \theta^{(k)}\left(g_i(\mathbf{x}_i) + h_i(\mathbf{x}_i) + \left\langle \mathcal{A}_i^T(\hat{\boldsymbol{\lambda}}^{k+1}) + (L_i\theta^{(k)}+\beta^{(k)}\eta_i)(\mathbf{z}_i^{k+1}-\mathbf{z}_i^k), \mathbf{x}_i-\mathbf{z}_i^{k+1}\right\rangle\right) \\
&\quad + \frac{L_i(\theta^{(k)})^2}{2}\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2 && (62)\\
&= (1-\theta^{(k)})f_i(\mathbf{x}_i^k) + \theta^{(k)}f_i(\mathbf{x}_i) + \theta^{(k)}\left\langle \mathcal{A}_i^T(\hat{\boldsymbol{\lambda}}^{k+1}), \mathbf{x}_i-\mathbf{z}_i^{k+1}\right\rangle \\
&\quad - \frac{L_i(\theta^{(k)})^2+\theta^{(k)}\beta^{(k)}\eta_i}{2}\left(\|\mathbf{z}_i^{k+1}-\mathbf{x}_i\|^2 - \|\mathbf{z}_i^k-\mathbf{x}_i\|^2 + \|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2\right) + \frac{L_i(\theta^{(k)})^2}{2}\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2, && (63)
\end{aligned}$$

where (57) uses (1), (58) uses (50), (59) is from the convexity of h_i, (61) is from the convexity of g_i, (62) uses (56) and (63) uses (3). Rerangging the above inequality leads to

$$\begin{aligned} & \left(f_i(\mathbf{x}_i^{k+1}) - f_i(\mathbf{x}_i)\right) - \theta^{(k)} \left\langle \mathcal{A}_i^T(\hat{\boldsymbol{\lambda}}^{k+1}), \mathbf{x}_i - \mathbf{z}_i^{k+1} \right\rangle \\ \leq \ & (1-\theta^{(k)}) \left(f_i(\mathbf{x}_i^k) - f_i(\mathbf{x}_i)\right) + \frac{L_i(\theta^{(k)})^2}{2} \left(\|\mathbf{z}_i^k - \mathbf{x}_i\|^2 - \|\mathbf{z}_i^{k+1} - \mathbf{x}_i\|^2\right) \\ & + \frac{\theta^{(k)}\beta^{(k)}\eta_i}{2} \left(\|\mathbf{z}_i^k - \mathbf{x}_i\|^2 - \|\mathbf{z}_i^{k+1} - \mathbf{x}_i\|^2 - \|\mathbf{z}_i^{k+1} - \mathbf{z}_i^k\|^2\right) \end{aligned} \tag{64}$$

Dividing both sides of the above inequality by $(\theta^{(k)})^2$ leads to

$$\begin{aligned} & \frac{1}{(\theta^{(k)})^2} \left(f_i(\mathbf{x}_i^{k+1}) - f_i(\mathbf{x}_i)\right) - \frac{1}{\theta^{(k)}} \left\langle \mathcal{A}_i^T(\hat{\boldsymbol{\lambda}}^{k+1}), \mathbf{x}_i - \mathbf{z}_i^{k+1} \right\rangle \\ \leq \ & \frac{1-\theta^{(k)}}{(\theta^{(k)})^2} \left(f_i(\mathbf{x}_i^k) - f_i(\mathbf{x}_i)\right) + \frac{L_i}{2} \left(\|\mathbf{z}_i^k - \mathbf{x}_i\|^2 - \|\mathbf{z}_i^{k+1} - \mathbf{x}_i\|^2\right) \\ & + \frac{\beta^{(k)}\eta_i}{2\theta^{(k)}} \left(\|\mathbf{z}_i^k - \mathbf{x}_i\|^2 - \|\mathbf{z}_i^{k+1} - \mathbf{x}_i\|^2 - \|\mathbf{z}_i^{k+1} - \mathbf{z}_i^k\|^2\right) \end{aligned}$$

The proof is completed by using $\frac{1-\theta^{(k+1)}}{(\theta^{(k+1)})^2} = \frac{1}{(\theta^{(k)})^2}$. ■

Proposition 4 *In Algorithm 2, the following result holds for any* $\boldsymbol{\lambda}$

$$\begin{aligned} & \langle \mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}, \boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}^{k+1} \rangle + \frac{\beta^{(k)}\alpha}{2} \|\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\|^2 \\ \leq \ & \frac{1}{2\beta^{(k)}} \left(\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}\|^2 - \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}\|^2\right) + \frac{\beta^{(k)}}{2} \sum_{i=1}^n \eta_i \|\mathbf{z}_i^{k+1} - \mathbf{z}_i^k\|^2, \end{aligned} \tag{65}$$

where $\alpha = \min\left\{\frac{1}{n+1}, \left\{\frac{\eta_i - n\|\mathcal{A}_i\|^2}{(n+1)\|\mathcal{A}_i\|^2}, i = 1, \cdots, n\right\}\right\}$.

Proof. By using (51) and (3), we have

$$\begin{aligned} & \langle \mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}, \boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}^{k+1} \rangle \\ = \ & \frac{1}{\beta^{(k)}} \langle \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k, \boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}^{k+1} \rangle \\ = \ & \frac{1}{2\beta^{(k)}} \left(\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^k\|^2 - \|\boldsymbol{\lambda} - \boldsymbol{\lambda}^{k+1}\|^2\right) - \frac{1}{2\beta^{(k)}} \left(\|\hat{\boldsymbol{\lambda}}^{k+1} - \boldsymbol{\lambda}^k\|^2 - \|\boldsymbol{\lambda}^{k+1} - \hat{\boldsymbol{\lambda}}^{k+1}\|^2\right). \end{aligned} \tag{66}$$

Now, consider the last two terms in the above inequality. We deduce

$$
\begin{aligned}
& \frac{1}{2\beta^{(k)}}\left(\|\hat{\boldsymbol{\lambda}}^{k+1}-\boldsymbol{\lambda}^k\|^2-\|\boldsymbol{\lambda}^{k+1}-\hat{\boldsymbol{\lambda}}^{k+1}\|^2\right) \\
= & \frac{\beta^{(k)}}{2}\left(\left\|\sum_{i=1}^n \mathcal{A}_i(\mathbf{z}_i^k)-\mathbf{b}\right\|^2-\left\|\sum_{i=1}^n \mathcal{A}_i(\mathbf{z}_i^{k+1}-\mathbf{z}_i^k)\right\|^2\right) \\
\geq & \frac{\beta^{(k)}}{2}\left(\left\|\sum_{i=1}^n \mathcal{A}_i(\mathbf{z}_i^k)-\mathbf{b}\right\|^2-\sum_{i=1}^n n\|\mathcal{A}_i\|^2\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2\right) \\
= & \frac{\beta^{(k)}}{2}\left(\left\|\sum_{i=1}^n \mathcal{A}_i(\mathbf{z}_i^k)-\mathbf{b}\right\|^2+\sum_{i=1}^n \frac{\eta_i-n\|\mathcal{A}_i\|^2}{\|\mathcal{A}_i\|^2}\|\mathcal{A}_i\|^2\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2-\sum_{i=1}^n \eta_i\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2\right) \\
\geq & \frac{\beta^{(k)}}{2}\left(\alpha(n+1)\left(\left\|\sum_{i=1}^n \mathcal{A}_i(\mathbf{z}_i^k)-\mathbf{b}\right\|^2+\sum_{i=1}^n\|\mathcal{A}_i(\mathbf{z}_i^{k+1}-\mathbf{z}_i^k)\|^2\right)-\sum_{i=1}^n \eta_i\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2\right) \\
\geq & \frac{\beta^{(k)}}{2}\left(\alpha\left\|\sum_{i=1}^n \mathcal{A}_i(\mathbf{z}_i^{k+1})-\mathbf{b}\right\|^2-\sum_{i=1}^n \eta_i\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2\right) \quad (67)\\
= & \frac{\beta^{(k)}\alpha}{2}\|\mathcal{A}(\mathbf{z}^{k+1})-\mathbf{b}\|^2-\frac{\beta^{(k)}}{2}\sum_{i=1}^n \eta_i\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2 \quad (68)
\end{aligned}
$$

The proof is completed by substituting (68) into (66). ■

Theorem 2 *In Algorithm 2, for any $K>0$, we have*

$$
\begin{aligned}
& f(\mathbf{x}^{K+1})-f(\mathbf{x}^*)+\left\langle\boldsymbol{\lambda}^*,\mathcal{A}(\mathbf{x}^{K+1})-\mathbf{b}\right\rangle+\frac{\beta\alpha}{2}\left\|\mathcal{A}(\mathbf{x}^{K+1})-\mathbf{b}\right\|^2 \\
\leq & \frac{2L_{\max}D_{\mathbf{x}^*}^2}{(K+2)^2}+\frac{2\beta\eta_{\max}D_X^2}{K+2}+\frac{2D_\Lambda^2}{\beta(K+2)},
\end{aligned} \quad (69)
$$

where $\alpha=\min\left\{\frac{1}{n+1},\left\{\frac{\eta_i-n\|\mathcal{A}_i\|^2}{2(n+1)\|\mathcal{A}_i\|^2},i=1,\cdots,n\right\}\right\}$, $L_{\max}=\max\{L_i,i=1,\cdots,n\}$ *and* $\eta_{\max}=\max\{\eta_i,i=1,\cdots,n\}$.

Proof. Let $\mathbf{x}_i=\mathbf{x}_i^*$ and $\boldsymbol{\lambda}=\boldsymbol{\lambda}^*$ in (53) and (65). We have

$$
\frac{1-\theta^{(k+1)}}{(\theta^{(k+1)})^2}\sum_{i=1}^n\left(f_i(\mathbf{x}_i^{k+1})-f_i(\mathbf{x}_i^*)\right)-\frac{1-\theta^{(k)}}{(\theta^{(k)})^2}\sum_{i=1}^n\left(f_i(\mathbf{x}_i^k)-f_i(\mathbf{x}_i^*)\right)+\frac{1}{\theta^{(k)}}\sum_{i=1}^n\left\langle\boldsymbol{\lambda}^*,\mathcal{A}_i(\mathbf{z}_i^{k+1})-\mathbf{b}\right\rangle \quad (70)
$$

$$
\begin{aligned}
\leq & \frac{1}{2}\sum_{i=1}^n L_i\left(\|\mathbf{z}_i^k-\mathbf{x}_i^*\|^2-\|\mathbf{z}_i^{k+1}-\mathbf{x}_i^*\|^2\right)+\frac{\beta^{(k)}}{2\theta^{(k)}}\sum_{i=1}^n\eta_i\left(\|\mathbf{z}_i^k-\mathbf{x}_i^*\|^2-\|\mathbf{z}_i^{k+1}-\mathbf{x}_i^*\|^2-\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2\right) \\
& \frac{1}{\theta^{(k)}}\langle\boldsymbol{\lambda}^*-\hat{\boldsymbol{\lambda}}^{k+1},\mathcal{A}(\mathbf{z}^{k+1})-\mathbf{b}\rangle \quad (71)\\
\leq & \frac{1}{2}\sum_{i=1}^n L_i\left(\|\mathbf{z}_i^k-\mathbf{x}_i^*\|^2-\|\mathbf{z}_i^{k+1}-\mathbf{x}_i^*\|^2\right)+\frac{\beta^{(k)}}{2\theta^{(k)}}\sum_{i=1}^n\eta_i\left(\|\mathbf{z}_i^k-\mathbf{x}_i^*\|^2-\|\mathbf{z}_i^{k+1}-\mathbf{x}_i^*\|^2-\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2\right) \\
& +\frac{1}{2\theta^{(k)}\beta^{(k)}}\left(\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^*\|^2-\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^*\|^2\right)+\frac{\beta^{(k)}}{2\theta^{(k)}}\sum_{i=1}^n\eta_i\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|^2-\frac{\beta^{(k)}\alpha}{2\theta^{(k)}}\|\mathcal{A}(\mathbf{z}^{k+1})-\mathbf{b}\|^2 \quad (72)\\
= & \frac{1}{2}\sum_{i=1}^n L_i\left(\|\mathbf{z}_i^k-\mathbf{x}_i^*\|^2-\|\mathbf{z}_i^{k+1}-\mathbf{x}_i^*\|^2\right)+\frac{\beta^{(k)}}{2\theta^{(k)}}\sum_{i=1}^n\eta_i\left(\|\mathbf{z}_i^k-\mathbf{x}_i^*\|^2-\|\mathbf{z}_i^{k+1}-\mathbf{x}_i^*\|^2\right) \\
& +\frac{1}{2\theta^{(k)}\beta^{(k)}}\left(\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^*\|^2-\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^*\|^2\right)-\frac{\beta^{(k)}\alpha}{2\theta^{(k)}}\|\mathcal{A}(\mathbf{z}^{k+1})-\mathbf{b}\|^2 \quad (73)
\end{aligned}
$$

where (71) uses the fact $\mathcal{A}(\mathbf{x}^*) = \mathbf{b}$ and (72) uses (65).

Summing (70)-(73) from $k = 0$ to K and fixing $\beta^{(k)} = \beta > 0$, we have

$$\begin{aligned}
& \frac{1-\theta^{(K+1)}}{(\theta^{(K+1)})^2}\sum_{i=1}^n \left(f_i(\mathbf{x}_i^{K+1}) - f_i(\mathbf{x}_i^*)\right) - \frac{1-\theta^{(0)}}{(\theta^{(0)})^2}\sum_{i=1}^n \left(f_i(\mathbf{x}_i^{0}) - f_i(\mathbf{x}_i^*)\right) + \sum_{k=0}^K \frac{1}{\theta^{(k)}}\left\langle \boldsymbol{\lambda}^*, \mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\right\rangle \qquad (74)\\
\leq\ & \frac{1}{2}\sum_{i=1}^n L_i\|\mathbf{z}_i^0 - \mathbf{x}_i^*\|^2 + \sum_{k=0}^K \frac{1}{2\theta^{(k)}}\sum_{i=1}^n \beta\eta_i\left(\|\mathbf{z}_i^k - \mathbf{x}_i^*\|^2 - \|\mathbf{z}_i^{k+1} - \mathbf{x}_i^*\|^2\right)\\
& + \sum_{k=0}^K \frac{1}{2\theta^{(k)}\beta}\left(\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 - \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|^2\right) - \sum_{k=0}^K \frac{\beta\alpha}{2\theta^{(k)}}\|\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\|^2\\
\leq\ & \frac{1}{2}\sum_{i=1}^n L_i\|\mathbf{z}_i^0 - \mathbf{x}_i^*\|^2 + \frac{1}{2}\sum_{k=0}^K\sum_{i=1}^n \beta\eta_i\|\mathbf{z}_i^k - \mathbf{x}_i^*\|^2 + \frac{1}{2\beta}\sum_{k=0}^K \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 - \sum_{k=0}^K \frac{\beta\alpha}{2\theta^{(k)}}\|\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\|^2, \qquad (75)\\
\leq\ & \frac{1}{2}\left(L_{\max}D_{\mathbf{x}^*}^2 + K\beta\eta_{\max}D_{\mathbf{X}}^2 + \frac{1}{\beta}KD_{\boldsymbol{\Lambda}} - \beta\alpha\sum_{k=0}^K \frac{1}{\theta^{(k)}}\|\mathcal{A}(\mathbf{z}^{k+1}) - \mathbf{b}\|^2\right) \qquad (76)
\end{aligned}$$

where (75) uses (4). Also note that $\theta^{(0)} = 1$. So the second term of (74) disappears.

Note that (42) and (44) also holds here. Substituting (42) into (74) and (44) into (76) respectively and using $\frac{1-\theta^{(K+1)}}{(\theta^{(K+1)})^2} = \frac{1}{(\theta^{(K)})^2}$, we obtain

$$\begin{aligned}
& \frac{1}{(\theta^{(K)})^2}\sum_{i=1}^n \left(f_i(\mathbf{x}_i^{K+1}) - f_i(\mathbf{x}_i^*)\right) + \frac{1}{(\theta^{(K)})^2}\left\langle \boldsymbol{\lambda}^*, \mathcal{A}(\mathbf{x}^{K+1}) - \mathbf{b}\right\rangle + \frac{\beta\alpha}{2(\theta^{(K)})^2}\|\mathcal{A}(\mathbf{x}^{K+1}) - \mathbf{b}\|^2\\
\leq\ & \frac{1}{2}\left(L_{\max}D_{\mathbf{x}^*}^2 + K\beta\eta_{\max}D_X^2 + \frac{1}{\beta}KD_\Lambda\right).
\end{aligned}$$

The proof is completed by multiplying both sides of the above inequality with $\theta^{(K)}$ and using (8). ■

Theorem 3 *Assume the mapping $\mathcal{A}(\mathbf{x}_1, \cdots, \mathbf{x}_n) = \sum_{i=1}^n \mathcal{A}_i(\mathbf{x}_i)$ is onto*[1]*, the sequence $\{\mathbf{z}^k\}$ is bounded, $\partial h(\mathbf{x})$ and $\nabla g(\mathbf{x})$ are bounded if $\mathbf{x}$ is bounded, then $\{\mathbf{x}^k\}$, $\{\mathbf{y}^k\}$ and $\{\boldsymbol{\lambda}^k\}$ are bounded.*

Proof. Assume $\|\mathbf{z}^k\| \leq C_1$ for all k and $\|\mathbf{x}^0\| \leq C_1$. Then from (50) we can easily get $\|\mathbf{x}^k\| \leq C_1$ for all k. Then from (48) we have $\|\mathbf{y}^k\| \leq C_1$ for all k. Assume $\|\partial h(\mathbf{x})\| \leq C_2$ and $\|\nabla g(\mathbf{x})\| \leq C_2$ if $\|\mathbf{x}\| \leq C_1$. Then from (55), we have

$$0 \in \partial h(\mathbf{z}^{k+1}) + \nabla g(\mathbf{y}^{k+1}) + \mathcal{A}^T(\boldsymbol{\lambda}^k) + \beta^{(k)}\mathcal{A}^T(\mathcal{A}(\mathbf{z}^k) - \mathbf{b}) + \begin{bmatrix} (L_1\theta^{(k)} + \beta^{(k)}\eta_1)(\mathbf{z}_1^{k+1} - \mathbf{z}_1^k)\\ \vdots\\ (L_i\theta^{(k)} + \beta^{(k)}\eta_i)(\mathbf{z}_i^{k+1} - \mathbf{z}_i^k)\\ \vdots\\ (L_n\theta^{(k)} + \beta^{(k)}\eta_n)(\mathbf{z}_n^{k+1} - \mathbf{z}_n^k)\end{bmatrix},$$

and

$$-\mathcal{A}\mathcal{A}^T(\boldsymbol{\lambda}^k) \in \mathcal{A}\left(\partial h(\mathbf{z}^{k+1}) + \nabla g(\mathbf{y}^{k+1}) + \beta^{(k)}\mathcal{A}^T(\mathcal{A}(\mathbf{z}^k) - \mathbf{b}) + \begin{bmatrix} (L_1\theta^{(k)} + \beta^{(k)}\eta_1)(\mathbf{z}_1^{k+1} - \mathbf{z}_1^k)\\ \vdots\\ (L_i\theta^{(k)} + \beta^{(k)}\eta_i)(\mathbf{z}_i^{k+1} - \mathbf{z}_i^k)\\ \vdots\\ (L_n\theta^{(k)} + \beta^{(k)}\eta_n)(\mathbf{z}_n^{k+1} - \mathbf{z}_n^k)\end{bmatrix}\right).$$

[1]This assumption is equivalent to that the matrix $A \equiv (A_1, \cdots, A_n)$ is of full row rank, where A_i is the matrix representation of $\mathcal{A}_i$.

So we have

$$
\begin{aligned}
\|\boldsymbol{\lambda}^k\| &\leq \left\|(\mathcal{A}\mathcal{A}^T)^{-1}\mathcal{A}\left(\partial h(\mathbf{z}^{k+1}) + \nabla g(\mathbf{y}^{k+1}) + \beta^{(k)}\mathcal{A}^T(\mathcal{A}\mathbf{z}^k - \mathbf{b}) + \begin{bmatrix} (L_1\theta^{(k)} + \beta^{(k)}\eta_1)(\mathbf{z}_1^{k+1} - \mathbf{z}_1^k) \\ \vdots \\ (L_i\theta^{(k)} + \beta^{(k)}\eta_i)(\mathbf{z}_i^{k+1} - \mathbf{z}_i^k) \\ \vdots \\ (L_n\theta^{(k)} + \beta^{(k)}\eta_n)(\mathbf{z}_n^{k+1} - \mathbf{z}_n^k) \end{bmatrix}\right)\right\| \\
&\leq \|(\mathcal{A}\mathcal{A}^T)^{-1}\mathcal{A}\| \left(\|\partial h(\mathbf{z}^{k+1})\| + \|\nabla g(\mathbf{y}^{k+1})\| + \|\beta^{(k)}\mathcal{A}^T\mathcal{A}\mathbf{z}^k\| + \|\beta^{(k)}\mathcal{A}^T\mathbf{b}\| + (L_{\max}\theta^{(k)} + \beta^{(k)}\eta_{\max})(\|\mathbf{z}^{k+1}\| + \|\mathbf{z}^k\|)\right) \\
&\leq \|(\mathcal{A}\mathcal{A}^T)^{-1}\mathcal{A}\| \left(2C_2 + \beta^{(k)}\|\mathcal{A}^T\mathcal{A}\|C_1 + \beta^{(k)}\|\mathcal{A}^T\mathbf{b}\| + 2(L_{\max}\theta^{(k)} + \beta^{(k)}\eta_{\max})C_1\right).
\end{aligned}
$$

for all k. ■

References

[1] Zhouchen Lin, Risheng Liu, and Huan Li. Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning. In *Machine Learning*, 2014.

[2] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer, 2004.