JMLR: Workshop and Conference Proceedings 29:1–20, 2013 ACML 2013

Linearized Alternating Direction Method with Parallel Splitting and Adaptive Penalty for Separable Convex Programs in Machine Learning – Supplementary Materials

Risheng Liu RSLIU@DLUT.EDU.CN
Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology

Zhouchen Lin ZLIN@PKU.EDU.CN
Key Lab. of Machine Perception (MOE), School of EECS, Peking University

Zhixun Su ZXSU@DLUT.EDU.CN
School of Mathematical Sciences, Dalian University of Technology

Editor: Cheng Soon Ong and Tu Bao Ho

1. Introduction

Please note that the number of equations, propositions and theorems in supplemental materials are different from that in the manuscript.

The problem we are interested in is as follows:

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

where $\mathbf{x}_i$ and $\mathbf{b}$ could be either vectors or matrices, f_i is a closed proper convex function, and $\mathcal{A}_i : \mathbb{R}^{d_i} \to \mathbb{R}^m$ is a linear mapping. Without loss of generality, we may assume that none of the $\mathcal{A}_i$'s is a zero mapping, the solution to $\sum\limits_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i) = \mathbf{b}$ is non-unique, and the mapping $\mathcal{A}(\mathbf{x}_1, \cdots, \mathbf{x}_n) \equiv \sum\limits_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i)$ is onto[1].

We propose LADMPSAP to solve (1), which consists of the following steps:

1. Update $\mathbf{x}_i$'s in parallel:

$$\mathbf{x}_i^{k+1} = \operatorname*{argmin}_{\mathbf{x}_i} f_i(\mathbf{x}_i) + \frac{\sigma_i^{(k)}}{2} \left\| \mathbf{x}_i - \mathbf{u}_i^k \right\|^2, \quad i = 1, \cdots, n, \tag{2}$$

2. Update λ:

$$\lambda^{k+1} = \lambda^k + \beta_k \left(\sum_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i^{k+1}) - \mathbf{b} \right), \tag{3}$$

1. The latter two assumptions are equivalent to that the matrix $\mathbf{A} \equiv (\mathbf{A}_1 \ \cdots \ \mathbf{A}_n)$ is not full column rank but full row rank, where $\mathbf{A}_i$ is the matrix representation of $\mathcal{A}_i$.

3. Update β:

$$\beta_{k+1} = \min(\beta_{\max}, \rho\beta_k), \tag{4}$$

where $\sigma_i^{(k)} = \eta_i \beta_k$,

$$\mathbf{u}_i^k = \mathbf{x}_i^k - \mathcal{A}_i^*(\hat{\lambda}^k)/\sigma_i^{(k)}, \tag{5}$$

in which $\mathcal{A}_i^*$ is the adjoint operator of $\mathcal{A}_i$ and

$$\hat{\lambda}^k = \lambda^k + \beta_k \left(\sum_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i^k) - \mathbf{b} \right), \tag{6}$$

and

$$\rho = \begin{cases} \rho_0, & \text{if } \max\left(\left\{ \sqrt{\beta_k \sigma_i^{(k)}} \left\| \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \right\|, i = 1, \cdots, n \right\} \right) / \|\mathbf{b}\| < \varepsilon_2, \\ 1, & \text{otherwise}, \end{cases} \tag{7}$$

with $\rho_0 > 1$ being a constant and $0 < \varepsilon_2 \ll 1$ being a threshold.

The iteration terminates when the following two conditions are met:

$$\left\| \sum_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i^{k+1}) - \mathbf{b} \right\| / \|\mathbf{b}\| < \varepsilon_1, \tag{8}$$

$$\max\left(\left\{ \sqrt{\beta_k \sigma_i^{(k)}} \left\| \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \right\|, i = 1, \cdots, n \right\} \right) / \|\mathbf{b}\| < \varepsilon_2. \tag{9}$$

For more details, please refer to Lin et al. (2011).

2. Some Lemmas

Lemma 1 *The Kuhn-Karush-Tucker (KKT) condition of problem (1) is that there exists* $(\mathbf{x}_1^*, \cdots, \mathbf{x}_n^*, \lambda^*)$, *such that*

$$\sum_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i^*) = \mathbf{b}, \tag{10}$$

$$-\mathcal{A}_i^*(\lambda^*) \in \partial f_i(\mathbf{x}_i^*), \quad i = 1, \cdots, n, \tag{11}$$

where ∂f_i is the subgradient of f_i. The first is the feasibility condition and the second is the duality condition. Such $(\mathbf{x}_1^*, \cdots, \mathbf{x}_n^*, \lambda^*)$ is called a KKT point of problem (1).

Lemma 2

$$-\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) \in \partial f_i(\mathbf{x}_i^{k+1}), \quad i = 1, \cdots, n. \tag{12}$$

This can be easily proved by checking the optimality conditions of (2).

Lemma 3

$$\left\langle -\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) + \mathcal{A}_i^*(\lambda^*), \mathbf{x}_i^{k+1} - \mathbf{x}_i^* \right\rangle \geq 0, \quad i = 1, \cdots, n. \tag{13}$$

Proof We first cite a classic result on the monotonicity of subgradient mapping Rockafellar (1970): for any convex function f,

$$\langle \mathbf{p}_1 - \mathbf{p}_2, \mathbf{x}_1 - \mathbf{x}_2 \rangle \geq 0, \quad \forall \mathbf{p}_i \in \partial f(\mathbf{x}_i), i = 1, 2. \tag{14}$$

By choosing $\mathbf{x}_1 = \mathbf{x}_i^{k+1}$ and $\mathbf{x}_2 = \mathbf{x}_i^*$ and utilizing (12) and (11), we have (13). ■

Lemma 4

$$\begin{aligned}
& \beta_k \sum_{i=1}^{n} \sigma_i^{(k)} \left\| \mathbf{x}_i^{k+1} - \mathbf{x}_i^* \right\|^2 + \left\| \lambda^{k+1} - \lambda^* \right\|^2 && (15) \\
= \; & \beta_k \sum_{i=1}^{n} \sigma_i^{(k)} \left\| \mathbf{x}_i^{k} - \mathbf{x}_i^* \right\|^2 + \left\| \lambda^{k} - \lambda^* \right\|^2 && (16) \\
& -2\beta_k \sum_{i=1}^{n} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, -\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) + \mathcal{A}_i^*(\lambda^*) \right\rangle && (17) \\
& -\beta_k \sum_{i=1}^{n} \sigma_i^{(k)} \left\| \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \right\|^2 - \left\| \lambda^{k+1} - \lambda^k \right\|^2 && (18) \\
& -2\beta_k \sum_{i=1}^{n} \sigma_i^{(k)} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathbf{x}_i^k - \mathbf{u}_i^k \right\rangle && (19) \\
& +2 \left\langle \lambda^{k+1} - \lambda^k, \lambda^{k+1} \right\rangle. && (20)
\end{aligned}$$

Proof This can be easily checked. First, we add (17) and (19) to have

$$\begin{aligned}
& -2\beta_k \sum_{i=1}^{n} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, -\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) + \mathcal{A}_i^*(\lambda^*) \right\rangle && (21) \\
& -2\beta_k \sum_{i=1}^{n} \sigma_i^{(k)} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathbf{x}_i^k - \mathbf{u}_i^k \right\rangle && (22) \\
= \; & -2\beta_k \sum_{i=1}^{n} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathcal{A}_i^*(\lambda^*) \right\rangle + 2\beta_k \sum_{i=1}^{n} \sigma_i^{(k)} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \right\rangle && (23) \\
= \; & -2\beta_k \sum_{i=1}^{n} \left\langle \mathcal{A}_i(\mathbf{x}_i^{k+1} - \mathbf{x}_i^*), \lambda^* \right\rangle + 2\beta_k \sum_{i=1}^{n} \sigma_i^{(k)} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \right\rangle && (24) \\
= \; & -2 \left\langle \beta_k \sum_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i^{k+1} - \mathbf{x}_i^*), \lambda^* \right\rangle + 2\beta_k \sum_{i=1}^{n} \sigma_i^{(k)} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \right\rangle && (25) \\
= \; & -2 \left\langle \beta_k \left(\sum_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i^{k+1}) - \mathbf{b} \right), \lambda^* \right\rangle + 2\beta_k \sum_{i=1}^{n} \sigma_i^{(k)} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \right\rangle && (26) \\
= \; & -2 \left\langle \lambda^{k+1} - \lambda^k, \lambda^* \right\rangle + 2\beta_k \sum_{i=1}^{n} \sigma_i^{(k)} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \right\rangle, && (27)
\end{aligned}$$

where we have used (10) in (26). Then we apply the identity

$$2\left\langle \mathbf{a}_{k+1}-\mathbf{a}^*, \mathbf{a}_{k+1}-\mathbf{a}_k\right\rangle = \|\mathbf{a}_{k+1}-\mathbf{a}^*\|^2 - \|\mathbf{a}_k-\mathbf{a}^*\|^2 + \|\mathbf{a}_{k+1}-\mathbf{a}_k\|^2 \tag{28}$$

to see that (15)-(20) holds. ■

3. Proofs of Propositions and Theorems

Proposition 5 *(Proposition 1 in the manuscript)*

$$\begin{aligned}
& \beta_k \sum_{i=1}^n \sigma_i^{(k)} \|\mathbf{x}_i^{k+1}-\mathbf{x}_i^*\|^2 + \|\lambda^{k+1}-\lambda^*\|^2 && (29)\\
\leq\ & \beta_k \sum_{i=1}^n \sigma_i^{(k)} \|\mathbf{x}_i^{k}-\mathbf{x}_i^*\|^2 + \|\lambda^{k}-\lambda^*\|^2 && (30)\\
& -2\beta_k \sum_{i=1}^n \left\langle \mathbf{x}_i^{k+1}-\mathbf{x}_i^*, -\sigma_i^{(k)}(\mathbf{x}_i^{k+1}-\mathbf{u}_i^k) + \mathcal{A}_i^*(\lambda^*)\right\rangle && (31)\\
& -\beta_k \sum_{i=1}^n \left(\sigma_i^{(k)} - n\beta_k\|\mathcal{A}_i\|^2\right) \|\mathbf{x}_i^{k+1}-\mathbf{x}_i^k\|^2 && (32)\\
& -\|\lambda^k-\hat{\lambda}^k\|^2. && (33)
\end{aligned}$$

Proof We continue from (19)-(20). As $\sigma_i^{(k)}(\mathbf{x}_i^k-\mathbf{u}_i^k) = \mathcal{A}_i^*(\hat{\lambda}^k)$, we have

$$\begin{aligned}
& -2\beta_k \sum_{i=1}^n \sigma_i^{(k)} \left\langle \mathbf{x}_i^{k+1}-\mathbf{x}_i^*, \mathbf{x}_i^k-\mathbf{u}_i^k\right\rangle + 2\left\langle \lambda^{k+1}-\lambda^k, \lambda^{k+1}\right\rangle && (34)\\
=\ & -2\beta_k \sum_{i=1}^n \left\langle \mathcal{A}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i^*), \hat{\lambda}^k\right\rangle + 2\left\langle \lambda^{k+1}-\lambda^k, \lambda^{k+1}\right\rangle && (35)\\
=\ & -2\beta_k \left\langle \sum_{i=1}^n \mathcal{A}_i(\mathbf{x}_i^{k+1}) - \sum_{i=1}^n \mathcal{A}_i(\mathbf{x}_i^*), \hat{\lambda}^k\right\rangle + 2\left\langle \lambda^{k+1}-\lambda^k, \lambda^{k+1}\right\rangle && (36)\\
=\ & -2\left\langle \lambda^{k+1}-\lambda^k, \hat{\lambda}^k\right\rangle + 2\left\langle \lambda^{k+1}-\lambda^k, \lambda^{k+1}\right\rangle && (37)\\
=\ & 2\left\langle \lambda^{k+1}-\lambda^k, \lambda^{k+1}-\hat{\lambda}^k\right\rangle && (38)\\
=\ & \|\lambda^{k+1}-\lambda^k\|^2 + \|\lambda^{k+1}-\hat{\lambda}^k\|^2 - \|\lambda^k-\hat{\lambda}^k\|^2 && (39)\\
=\ & \|\lambda^{k+1}-\lambda^k\|^2 + \beta_k^2 \left\| \sum_{i=1}^n \mathcal{A}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k)\right\|^2 - \|\lambda^k-\hat{\lambda}^k\|^2 && (40)\\
\leq\ & \|\lambda^{k+1}-\lambda^k\|^2 + \beta_k^2 \left(\sum_{i=1}^n \|\mathcal{A}_i\|\|\mathbf{x}_i^{k+1}-\mathbf{x}_i^k\|\right)^2 - \|\lambda^k-\hat{\lambda}^k\|^2 && (41)\\
\leq\ & \|\lambda^{k+1}-\lambda^k\|^2 + n\beta_k^2 \sum_{i=1}^n \|\mathcal{A}_i\|^2\|\mathbf{x}_i^{k+1}-\mathbf{x}_i^k\|^2 - \|\lambda^k-\hat{\lambda}^k\|^2 && (42)
\end{aligned}$$

Plugging the above into (19)-(20), we have (29)-(33). ■

Proposition 6 *(Proposition 3 in the manuscript) Let* $\sigma_i^{(k)} = \eta_i \beta_k$, $i = 1, \cdots, n$. *If* $\{\beta_k\}$ *is non-decreasing,* $\eta_i > n\|\mathcal{A}_i\|^2$, $i = 1, \cdots, n$, *and* $(\mathbf{x}_1^*, \cdots, \mathbf{x}_n^*, \lambda^*)$ *is any KKT point of problem (1), then:*

1) $\left\{\sum\limits_{i=1}^{n} \eta_i \|\mathbf{x}_i^k - \mathbf{x}_i^*\|^2 + \beta_k^{-2}\|\lambda^k - \lambda^*\|^2\right\}$ *is nonnegative and non-increasing.*

2) $\|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\| \to 0$, $i = 1, \cdots, n$, $\beta_k^{-1}\|\lambda^k - \hat{\lambda}^k\| \to 0$.

3) $\sum\limits_{k=1}^{+\infty} \beta_k^{-1} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, -\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) + \mathcal{A}_i^*(\lambda^*) \right\rangle < +\infty$, $i = 1, \cdots, n$.

Proof We divide both sides of (29)-(33) by β_k^2 to have

$$
\begin{aligned}
& \sum_{i=1}^{n} \eta_i \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^*\|^2 + \beta_k^{-2}\|\lambda^{k+1} - \lambda^*\|^2 && (43)\\
\leq\ & \sum_{i=1}^{n} \eta_i \|\mathbf{x}_i^{k} - \mathbf{x}_i^*\|^2 + \beta_k^{-2}\|\lambda^{k} - \lambda^*\|^2 && (44)\\
& -2\beta_k^{-1} \sum_{i=1}^{n} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, -\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) + \mathcal{A}_i^*(\lambda^*) \right\rangle && (45)\\
& -\sum_{i=1}^{n} \left(\eta_i - n\|\mathcal{A}_i\|^2\right) \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2 && (46)\\
& -\beta_k^{-2}\|\lambda^k - \hat{\lambda}^k\|^2. && (47)
\end{aligned}
$$

Then by (13), $\eta_i > n\|\mathcal{A}_i\|^2$ and the non-decrement of $\{\beta_k\}$, we can easily obtain 1). Second, we sum both sides of (43)-(47) over k to have

$$
\begin{aligned}
& 2\sum_{k=0}^{+\infty} \beta_k^{-1} \sum_{i=1}^{n} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}^*, -\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) + \mathcal{A}_i^*(\lambda^*) \right\rangle && (48)\\
& +\sum_{i=1}^{n} \left(\eta_i - n\|\mathcal{A}_i\|^2\right) \sum_{k=0}^{+\infty} \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2 && (49)\\
& +\sum_{k=0}^{+\infty} \beta_k^{-2}\|\lambda^k - \hat{\lambda}^k\|^2 && (50)\\
\leq\ & \sum_{i=1}^{n} \eta_i \|\mathbf{x}_i^{0} - \mathbf{x}^*\|^2 + \beta_0^{-2}\|\lambda^{0} - \lambda^*\|^2. && (51)
\end{aligned}
$$

Then 2) and 3) can be easily deduced. ■

Theorem 7 *(Theorem 4 in the manuscript) If $\{\beta_k\}$ is non-decreasing and upper bounded, $\eta_i > n\|\mathcal{A}_i\|^2$, $i = 1, \cdots, n$, then the sequence $\{(\{\mathbf{x}_i^k\}, \lambda^k)\}$ generated by LADMPSAP converges to a KKT point of problem (1).*

The proof resembles that in Lin et al. (2011).

Proof By Proposition 6-1) and the boundedness of $\{\beta_k\}$, $\{(\mathbf{x}_1^k, \cdots, \mathbf{x}_n^k, \lambda^k)\}$ is bounded, hence has an accumulation point, say $(\mathbf{x}_1^{k_j}, \cdots, \mathbf{x}_n^{k_j}, \lambda^{k_j}) \to (\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty, \lambda^\infty)$. We accomplish the proof in two steps.

1. We first prove that $(\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty, \lambda^\infty)$ is a KKT point of problem (1).

By Proposition 6-2),

$$\sum_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i^k) - \mathbf{b} = \beta_k^{-1}(\hat{\lambda}^k - \lambda^k) \to 0.$$

So any accumulation point of $\{(\mathbf{x}_1^k, \cdots, \mathbf{x}_n^k)\}$ is a feasible solution.

Since $-\sigma_i^{(k_j-1)}(\mathbf{x}_i^{k_j} - \mathbf{u}_i^{k_j-1}) \in \partial f_i(\mathbf{x}_i^{k_j})$, we have

$$\begin{aligned}\sum_{i=1}^{n} f_i(\mathbf{x}_i^{k_j}) &\leq \sum_{i=1}^{n} f_i(\mathbf{x}_i^*) + \sum_{i=1}^{n} \left\langle \mathbf{x}_i^{k_j} - \mathbf{x}_i^*, -\sigma_i^{(k_j-1)}(\mathbf{x}_i^{k_j} - \mathbf{u}_i^{k_j-1}) \right\rangle \\ &= \sum_{i=1}^{n} f_i(\mathbf{x}_i^*) + \sum_{i=1}^{n} \left\langle \mathbf{x}_i^{k_j} - \mathbf{x}_i^*, -\eta_i\beta_{k_j-1}(\mathbf{x}_i^{k_j} - \mathbf{x}_i^{k_j-1}) - \mathcal{A}_i^*(\hat{\lambda}^{k_j-1}) \right\rangle.\end{aligned}$$

Let $j \to +\infty$. By observing Proposition 6-2) and the boundedness of $\{\beta_k\}$, we have

$$\begin{aligned}\sum_{i=1}^{n} f_i(\mathbf{x}_i^\infty) &\leq \sum_{i=1}^{n} f_i(\mathbf{x}_i^*) + \sum_{i=1}^{n} \langle \mathbf{x}_i^\infty - \mathbf{x}_i^*, -\mathcal{A}_i^*(\lambda^\infty) \rangle \\ &= \sum_{i=1}^{n} f_i(\mathbf{x}_i^*) - \sum_{i=1}^{n} \langle \mathcal{A}(\mathbf{x}_i^\infty - \mathbf{x}_i^*), \lambda^\infty \rangle \\ &= \sum_{i=1}^{n} f_i(\mathbf{x}_i^*) - \left\langle \sum_{i=1}^{n} \mathcal{A}(\mathbf{x}_i^\infty) - \mathbf{b}, \lambda^\infty \right\rangle \\ &= \sum_{i=1}^{n} f_i(\mathbf{x}_i^*).\end{aligned}$$

So we conclude that $(\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty)$ is an optimal solution to (1).

Again by $-\sigma_i^{(k_j-1)}(\mathbf{x}_i^{k_j} - \mathbf{u}_i^{k_j-1}) \in \partial f_i(\mathbf{x}_i^{k_j})$ we have

$$\begin{aligned}f_i(\mathbf{x}) &\geq f_i(\mathbf{x}_i^{k_j}) + \left\langle \mathbf{x} - \mathbf{x}_i^{k_j}, -\sigma_i^{(k_j-1)}(\mathbf{x}_i^{k_j} - \mathbf{u}_i^{k_j-1}) \right\rangle \\ &= f_i(\mathbf{x}_i^{k_j}) + \left\langle \mathbf{x} - \mathbf{x}_i^{k_j}, -\eta_i\beta_{k_j-1}(\mathbf{x}_i^{k_j} - \mathbf{x}_i^{k_j-1}) - \mathcal{A}_i^*(\hat{\lambda}^{k_j-1}) \right\rangle.\end{aligned}$$

Fixing $\mathbf{x}$ and letting $j \to +\infty$, we see that

$$f_i(\mathbf{x}) \geq f_i(\mathbf{x}_i^\infty) + \langle \mathbf{x} - \mathbf{x}_i^\infty, -\mathcal{A}_i^*(\lambda^\infty) \rangle, \quad \forall \mathbf{x}.$$

So $-\mathcal{A}_i^*(\lambda^\infty) \in \partial f_i(\mathbf{x}_i^\infty)$, $i = 1, \cdots, n$. Thus $(\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty, \lambda^\infty)$ is a KKT point of problem (1).

2. We next prove that the whole sequence $\{(\mathbf{x}_1^k, \cdots, \mathbf{x}_n^k, \lambda^k)\}$ converges to $(\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty, \lambda^\infty)$.

By choosing $(\mathbf{x}_1^*, \cdots, \mathbf{x}_n^*, \lambda^*) = (\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty, \lambda^\infty)$ in Proposition 6, we have

$$\sum_{i=1}^{n} \eta_i \|\mathbf{x}_i^{k_j} - \mathbf{x}_i^\infty\|^2 + \beta_{k_j}^{-2} \|\lambda^{k_j} - \lambda^\infty\|^2 \to 0.$$

By Proposition 6-1), we readily have

$$\sum_{i=1}^{n} \eta_i \|\mathbf{x}_i^{k} - \mathbf{x}_i^\infty\|^2 + \beta_{k}^{-2} \|\lambda^{k} - \lambda^\infty\|^2 \to 0.$$

So $(\mathbf{x}_1^k, \cdots, \mathbf{x}_n^k, \lambda^k) \to (\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty, \lambda^\infty)$.

As $(\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty, \lambda^\infty)$ can be an arbitrary accumulation point of $\{(\mathbf{x}_1^k, \cdots, \mathbf{x}_n^k, \lambda^k)\}$, we conclude that $\{(\mathbf{x}_1^k, \cdots, \mathbf{x}_n^k, \lambda^k)\}$ converges to a KKT point of problem (1). ∎

Proposition 8 *(Proposition 5 in the manuscript) If $\{\beta_k\}$ is non-decreasing and unbounded, $\eta_i > n\|\mathcal{A}_i\|^2$, $\partial f_i(\mathbf{x})$ is bounded, $i = 1, \cdots, n$, then Proposition 6 holds and*

$$\beta_k^{-1} \lambda^k \to 0. \tag{52}$$

Proof As the conditions here are stricter than those in Proposition 6, Proposition 6 holds. Then we have that $\{\beta_k^{-1}\|\lambda^k - \lambda^*\|\}$ is bounded due to Proposition 6-1). So $\{\beta_k^{-1}\lambda^k\}$ is bounded due to $\beta_k^{-1}\|\lambda^k\| \le \beta_k^{-1}\|\lambda^k - \lambda^*\| + \beta_k^{-1}\|\lambda^*\|$. $\{\beta_k^{-1}\hat{\lambda}^k\}$ is also bounded thanks to Proposition 6-2).

We rewrite Lemma 2 as

$$-\eta_i(\mathbf{x}_i^{k+1} - \mathbf{x}_i^k) - \mathcal{A}_i^*(\beta_k^{-1}\hat{\lambda}^k) \in \beta_k^{-1}\partial f_i(\mathbf{x}_i^{k+1}), \quad i = 1, \cdots, n. \tag{53}$$

Then by the boundedness of $\partial f_i(x)$, the unboundedness of $\{\beta_k\}$ and Proposition 6-2), letting $k \to +\infty$, we have that

$$\mathcal{A}_i^*(\check{\lambda}^\infty) = 0, \quad i = 1, \cdots, n. \tag{54}$$

where $\check{\lambda}^\infty$ is any accumulation points of $\{\beta_k^{-1}\hat{\lambda}^k\}$, which is the same as that of $\{\beta_k^{-1}\lambda^k\}$ due to Proposition 6-2).

Recall that we have assumed that the mapping $\mathcal{A}(\mathbf{x}_1, \cdots, \mathbf{x}_n) \equiv \sum\limits_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i)$ is onto. So $\cap_{i=1}^n null(\mathcal{A}_i^*) = 0$. Therefore by (54), $\check{\lambda}^\infty = 0$. ∎

Theorem 9 *(Theorem 6 in the manuscript) If $\{\beta_k\}$ is non-decreasing and $\sum\limits_{k=1}^{+\infty} \beta_k^{-1} = +\infty$, $\eta_i > n\|\mathcal{A}_i\|^2$, $\partial f_i(\mathbf{x})$ is bounded, $i = 1, \cdots, n$, then the sequence $\{\mathbf{x}_i^k\}$ generated by LADMP-SAP converges to an optimal solution to (1).*

Proof When $\{\beta_k\}$ is bounded, the convergence has been proven in Theorem 1. In the following, we only focus on the case that $\{\beta_k\}$ is unbounded.

By Proposition 6-1), $\{(\mathbf{x}_1^k, \cdots, \mathbf{x}_n^k)\}$ is bounded, hence has at least one accumulation point $(\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty)$. By Proposition 6-2), $(\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty)$ is a feasible solution.

Since $\sum\limits_{k=1}^{+\infty} \beta_k^{-1} = +\infty$ and Proposition 6-3), there exists a subsequence $\{(\mathbf{x}_1^{k_j}, \cdots, \mathbf{x}_n^{k_j})\}$ such that

$$\left\langle \mathbf{x}_i^{k_j} - \mathbf{x}_i^*, -\sigma_i^{(k_j-1)}(\mathbf{x}_i^{k_j} - \mathbf{u}_i^{k_j-1}) + \mathcal{A}_i^*(\lambda^*) \right\rangle \to 0, \quad i = 1, \cdots, n. \tag{55}$$

As $\mathbf{p}_i^{k_j} \equiv -\sigma_i^{(k_j-1)}(\mathbf{x}_i^{k_j} - \mathbf{u}_i^{k_j-1}) \in \partial f_i(\mathbf{x}_i^{k_j})$ and ∂f_i is bounded, we may assume that

$$\mathbf{x}_i^k \to \mathbf{x}_i^\infty \quad \text{and} \quad \mathbf{p}_i^{k_j} \to \mathbf{p}_i^\infty.$$

It can be easily proven that

$$\mathbf{p}_i^\infty \in \partial f_i(\mathbf{x}_i^\infty).$$

Then letting $j \to \infty$ in (55), we have

$$\langle \mathbf{x}_i^\infty - \mathbf{x}_i^*, \mathbf{p}_i^\infty + \mathcal{A}_i^*(\lambda^*) \rangle = 0, \quad i = 1, \cdots, n. \tag{56}$$

Then by $\mathbf{p}_i^{k_j} \in \partial f_i(\mathbf{x}_i^{k_j})$,

$$\sum_{i=1}^n f_i(\mathbf{x}_i^{k_j}) \le \sum_{i=1}^n f_i(\mathbf{x}_i^*) + \sum_{i=1}^n \left\langle \mathbf{x}_i^{k_j} - \mathbf{x}_i^*, \mathbf{p}_i^{k_j} \right\rangle. \tag{57}$$

Letting $j \to \infty$ and making use of (56), we have

$$\begin{aligned} \sum_{i=1}^n f_i(\mathbf{x}_i^\infty) &\le \sum_{i=1}^n f_i(\mathbf{x}_i^*) + \sum_{i=1}^n \langle \mathbf{x}_i^\infty - \mathbf{x}_i^*, \mathbf{p}_i^\infty \rangle \\ &= \sum_{i=1}^n f_i(\mathbf{x}_i^*) - \sum_{i=1}^n \langle \mathbf{x}_i^\infty - \mathbf{x}_i^*, \mathcal{A}_i^*(\lambda^*) \rangle \\ &= \sum_{i=1}^n f_i(\mathbf{x}_i^*) - \sum_{i=1}^n \langle \mathcal{A}_i(\mathbf{x}_i^\infty - \mathbf{x}_i^*), \lambda^* \rangle \\ &= \sum_{i=1}^n f_i(\mathbf{x}_i^*). \end{aligned} \tag{58}$$

So together with the feasibility of $\{(\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty)\}$ we have that $\{(\mathbf{x}_1^{k_j}, \cdots, \mathbf{x}_n^{k_j})\}$ converges to an optimal solution $\{(\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty)\}$ to (1).

Finally, we set $\mathbf{x}_i^* = \mathbf{x}_i^\infty$ and λ^* be the corresponding Lagrange multiplier λ^∞ in Proposition 6. By Proposition 8, we have that

$$\sum_{i=1}^n \eta_i \|\mathbf{x}_i^{k_j} - \mathbf{x}_i^\infty\|^2 + \beta_{k_j}^{-2} \|\lambda^{k_j} - \lambda^\infty\|^2 \to 0.$$

By Proposition 6-1), we readily have

$$\sum_{i=1}^n \eta_i \|\mathbf{x}_i^k - \mathbf{x}_i^\infty\|^2 + \beta_k^{-2} \|\lambda^k - \lambda^\infty\|^2 \to 0.$$

So $(\mathbf{x}_1^k, \cdots, \mathbf{x}_n^k) \to (\mathbf{x}_1^\infty, \cdots, \mathbf{x}_n^\infty)$. ■

Theorem 10 *(Theorem 7 in the manuscript) If $\{\beta_k\}$ is non-decreasing, $\eta_i > n\|\mathcal{A}_i\|^2$, $\partial f_i(\mathbf{x})$ is bounded, $i = 1, \cdots, n$, then $\sum\limits_{k=1}^{+\infty} \beta_k^{-1} = +\infty$ is also the necessary condition for the global convergence of $\{\mathbf{x}_i^k\}$ generated by LADMPSAP to an optimal solution to (1).*

Proof We first prove that there exist linear mappings $\mathcal{B}_i$, $i = 1, \cdots, n$, such that $\mathcal{B}_i$'s are not all zeros and $\sum\limits_{i=1}^{n} \mathcal{B}_i A_i^* = 0$. Indeed, $\sum\limits_{i=1}^{n} \mathcal{B}_i A_i^* = 0$ is equivalent to

$$\sum_{i=1}^{n} \mathbf{B}_i \mathbf{A}_i^T = 0, \tag{59}$$

where $\mathbf{B}_i$ is the matrix representation of $\mathcal{B}_i$. (59) can be further written as

$$(\mathbf{A}_1 \ \cdots \ \mathbf{A}_n) \begin{pmatrix} \mathbf{B}_1^T \\ \vdots \\ \mathbf{B}_n^T \end{pmatrix} = 0. \tag{60}$$

Recall that we have assumed that the solution to $\sum\limits_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i) = \mathbf{b}$ is non-unique. So $(\mathbf{A}_1 \ \cdots \ \mathbf{A}_n)$ is not full column rank hence (60) has nonzero solutions. Thus there exist $\mathcal{B}_i$'s such that they are not all zeros and $\sum\limits_{i=1}^{n} \mathcal{B}_i A_i^* = 0$.

By Lemma 2,

$$-\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) \in \partial f_i(\mathbf{x}_i^{k+1}), \quad i = 1, \cdots, n. \tag{61}$$

As ∂f_i is bounded, $i = 1, \cdots, n$, so is

$$\sum_{i=1}^{n} \mathcal{B}_i(\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k)) = \beta_k(\mathbf{v}^{k+1} - \mathbf{v}^k), \tag{62}$$

where $\mathbf{v}^k = \phi(\mathbf{x}_1^k, \cdots, \mathbf{x}_n^k)$ and

$$\phi(\mathbf{x}_1, \cdots, \mathbf{x}_n) = \sum_{i=1}^{n} \eta_i \mathcal{B}_i(\mathbf{x}_i). \tag{63}$$

In (62) we have utilized $\sum\limits_{i=1}^{n} \mathcal{B}_i A_i^* = 0$ to cancel $\hat{\lambda}^k$, whose boundedness is uncertain.

Then we have that there exists a constant $C > 0$ such that

$$\|\mathbf{v}^{k+1} - \mathbf{v}^k\| \le C\beta_k^{-1}. \tag{64}$$

If $\sum_{k=1}^{+\infty} \beta_k^{-1} < +\infty$, then $\{\mathbf{v}^k\}$ is a Cauchy sequence, hence has a limit $\mathbf{v}^\infty$. Define $\mathbf{v}^* = \phi(\mathbf{x}_1^*, \cdots, \mathbf{x}_n^*)$, where $(\mathbf{x}_1^*, \cdots, \mathbf{x}_n^*)$ is any optimal solution. Then

$$\begin{aligned}
\|\mathbf{v}^\infty - \mathbf{v}^*\| &= \left\| \mathbf{v}^0 + \sum_{k=0}^{\infty} (\mathbf{v}^{k+1} - \mathbf{v}^k) - \mathbf{v}^* \right\| && (65)\\
&\geq \|\mathbf{v}^0 - \mathbf{v}^*\| - \sum_{k=0}^{\infty} \|\mathbf{v}^{k+1} - \mathbf{v}^k\| && (66)\\
&\geq \|\mathbf{v}^0 - \mathbf{v}^*\| - C \sum_{k=0}^{\infty} \beta_k^{-1}. && (67)
\end{aligned}$$

So if $(\mathbf{x}_1^0, \cdots, \mathbf{x}_n^0)$ is initialized badly such that

$$\|\mathbf{v}^0 - \mathbf{v}^*\| > C \sum_{k=0}^{\infty} \beta_k^{-1}, \tag{68}$$

then $\|\mathbf{v}^\infty - \mathbf{v}^*\| > 0$, which implies that $(\mathbf{x}_1^k, \cdots, \mathbf{x}_n^k)$ cannot converge to $(\mathbf{x}_1^*, \cdots, \mathbf{x}_n^*)$. Note that (68) is possible because ϕ is not a zero mapping given the conditions on $\mathcal{B}_i$. ■

Proposition 11 *(Proposition 8 in the manuscript) $\tilde{\mathbf{x}}$ is an optimal solution to (1) if and only if there exists $\alpha > 0$, such that*

$$f(\tilde{\mathbf{x}}) - f(\mathbf{x}^*) + \sum_{i=1}^{n} \langle \mathcal{A}_i^*(\lambda^*), \tilde{\mathbf{x}}_i - \mathbf{x}_i^* \rangle + \alpha \left\| \sum_{i=1}^{n} \mathcal{A}_i(\tilde{\mathbf{x}}_i) - \mathbf{b} \right\|^2 = 0. \tag{69}$$

Proof If $\tilde{\mathbf{x}}$ is optimal, it is easy to check that

$$f(\tilde{\mathbf{x}}) - f(\mathbf{x}^*) + \sum_{i=1}^{n} \langle \mathcal{A}_i^*(\lambda^*), \tilde{\mathbf{x}}_i - \mathbf{x}_i^* \rangle + \alpha \left\| \sum_{i=1}^{n} \mathcal{A}_i(\tilde{\mathbf{x}}_i) - \mathbf{b} \right\|^2 = 0. \tag{70}$$

holds.

Since $-\mathcal{A}_i^*(\lambda^*) \in \partial f_i(\mathbf{x}_i^*)$, we have

$$f(\tilde{\mathbf{x}}) - f(\mathbf{x}^*) + \sum_{i=1}^{n} \langle \mathcal{A}_i^*(\lambda^*), \tilde{\mathbf{x}}_i - \mathbf{x}_i^* \rangle \geq 0.$$

So if (70) holds, we have

$$\begin{aligned}
f(\tilde{\mathbf{x}}) - f(\mathbf{x}^*) + \sum_{i=1}^{n} \langle \mathcal{A}_i^*(\lambda^*), \tilde{\mathbf{x}}_i - \mathbf{x}_i^* \rangle &= 0, && (71)\\
\sum_{i=1}^{n} \mathcal{A}_i(\tilde{\mathbf{x}}_i) - \mathbf{b} &= 0. && (72)
\end{aligned}$$

With (72), we have

$$\sum_{i=1}^{n}\langle\mathcal{A}_i^*(\lambda^*),\tilde{\mathbf{x}}_i-\mathbf{x}_i^*\rangle=\sum_{i=1}^{n}\langle\lambda^*,\mathcal{A}_i(\tilde{\mathbf{x}}_i-\mathbf{x}_i^*)\rangle=\left\langle\lambda^*,\sum_{i=1}^{n}\mathcal{A}_i(\tilde{\mathbf{x}}_i-\mathbf{x}_i^*)\right\rangle=0. \tag{73}$$

So (71) reduces to $f(\tilde{\mathbf{x}})=f(\mathbf{x}^*)$. As $\tilde{\mathbf{x}}$ satisfies the feasibility condition, it is an optimal solution to (1). ■

Theorem 12 *(Theorem 10 in the manuscript) Define* $\bar{\mathbf{x}}^K=\sum\limits_{k=0}^{K}\gamma_k\mathbf{x}^{k+1}$, *where* $\gamma_k=\beta_k^{-1}/\sum\limits_{j=0}^{K}\beta_j^{-1}$. *Then*

$$\begin{aligned}&f(\bar{\mathbf{x}}^K)-f(\mathbf{x}^*)+\sum_{i=1}^{n}\left\langle\mathcal{A}_i^*(\lambda^*),\bar{\mathbf{x}}_i^K-\mathbf{x}_i^*\right\rangle\\&+\frac{\alpha\beta_0}{2}\left\|\sum_{i=1}^{n}\mathcal{A}_i(\bar{\mathbf{x}}_i^K)-\mathbf{b}\right\|^2\leq C_0/\left(2\sum_{k=0}^{K}\beta_k^{-1}\right),\end{aligned} \tag{74}$$

where $\alpha^{-1}=(n+1)\max\left(1,\left\{\frac{\|\mathcal{A}_i\|^2}{\eta_i-n\|\mathcal{A}_i\|^2},i=1,\cdots,n\right\}\right)$ *and* $C_0=\sum\limits_{i=1}^{n}\eta_i\left\|\mathbf{x}_i^0-\mathbf{x}_i^*\right\|^2+\beta_0^{-2}\left\|\lambda^0-\lambda^*\right\|^2$.

Proof We first deduce

$$\begin{array}{cl}&\left\|\sum\limits_{i=1}^{n}\mathcal{A}_i(\mathbf{x}_i^{k+1})-\mathbf{b}\right\|^2\\=&\left\|\sum\limits_{i=1}^{n}\mathcal{A}_i(\mathbf{x}_i^{k})-\mathbf{b}+\sum\limits_{i=1}^{n}\mathcal{A}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k)\right\|^2\\\leq&\left(\left\|\sum\limits_{i=1}^{n}\mathcal{A}_i(\mathbf{x}_i^{k})-\mathbf{b}\right\|+\sum\limits_{i=1}^{n}\left\|\mathcal{A}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k)\right\|\right)^2\\\leq&(n+1)\left(\left\|\sum\limits_{i=1}^{n}\mathcal{A}_i(\mathbf{x}_i^{k})-\mathbf{b}\right\|^2+\sum\limits_{i=1}^{n}\|\mathcal{A}_i\|^2\left\|\mathbf{x}_i^{k+1}-\mathbf{x}_i^k\right\|^2\right)\\\leq&(n+1)\left(\beta_k^{-2}\left\|\lambda_k-\hat{\lambda}_k\right\|^2+\max\left\{\frac{\|\mathcal{A}_i\|^2}{\eta_i-n\left\|\mathcal{A}_i\right\|^2}\right\}\sum\limits_{i=1}^{n}\left(\eta_i-n\left\|\mathcal{A}_i\right\|^2\right)\left\|\mathbf{x}_i^{k+1}-\mathbf{x}_i^k\right\|^2\right)\\\leq&(n+1)\max\left\{1,\left\{\frac{\|\mathcal{A}_i\|^2}{\eta_i-n\left\|\mathcal{A}_i\right\|^2}\right\}\right\}\left(\beta_k^{-2}\left\|\lambda_k-\hat{\lambda}_k\right\|^2+\sum\limits_{i=1}^{n}\left(\eta_i-n\left\|\mathcal{A}_i\right\|^2\right)\left\|\mathbf{x}_i^{k+1}-\mathbf{x}_i^k\right\|^2\right)\\=&\alpha^{-1}\left(\beta_k^{-2}\left\|\lambda_k-\hat{\lambda}_k\right\|^2+\sum\limits_{i=1}^{n}\left(\eta_i-n\left\|\mathcal{A}_i\right\|^2\right)\left\|\mathbf{x}_i^{k+1}-\mathbf{x}_i^k\right\|^2\right).\end{array} \tag{75}$$

By Proposition 5, we have

$$\begin{aligned}
& \beta_k^{-1} \sum_{i=1}^n \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, -\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) + \mathcal{A}_i^*(\lambda^*) \right\rangle \\
& + \frac{1}{2} \sum_{i=1}^n \left(\eta_i - n \left\| \mathcal{A}_i \right\|^2 \right) \left\| \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \right\|^2 + \frac{1}{2} \beta_k^{-2} \| \lambda^k - \hat{\lambda}^k \|^2 \\
\leq \ & \frac{1}{2} \left(\sum_{i=1}^n \eta_i \left\| \mathbf{x}_i^k - \mathbf{x}_i^* \right\|^2 + \beta_k^{-2} \left\| \lambda^k - \lambda^* \right\|^2 \right) \\
& - \frac{1}{2} \left(\sum_{i=1}^n \eta_i \left\| \mathbf{x}_i^{k+1} - \mathbf{x}_i^* \right\|^2 + \beta_{k+1}^{-2} \left\| \lambda^{k+1} - \lambda^* \right\|^2 \right).
\end{aligned} \tag{76}$$

So by Lemma 2 and combining the above inequalities, we have

$$\begin{aligned}
& \beta_k^{-1} \left(f(\mathbf{x}^{k+1}) - f(\mathbf{x}^*) + \sum_{i=1}^n \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathcal{A}_i^*(\lambda^*) \right\rangle + \frac{\alpha \beta_0}{2} \left\| \sum_{i=1}^n \mathcal{A}_i(\mathbf{x}_i^{k+1}) - \mathbf{b} \right\|^2 \right) \\
\leq \ & \beta_k^{-1} \left(\sum_{i=1}^n \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, -\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) \right\rangle + \sum_{i=1}^n \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathcal{A}_i^*(\lambda^*) \right\rangle \right) \\
& + \frac{\alpha}{2} \left\| \sum_{i=1}^n \mathcal{A}_i(\mathbf{x}_i^{k+1}) - \mathbf{b} \right\|^2 \\
\leq \ & \beta_k^{-1} \sum_{i=1}^n \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, -\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) + \mathcal{A}_i^*(\lambda^*) \right\rangle \\
& + \frac{1}{2} \beta_k^{-2} \left\| \lambda_k - \hat{\lambda}_k \right\|^2 + \frac{1}{2} \sum_{i=1}^n \left(\eta_i - n \left\| \mathcal{A}_i \right\|^2 \right) \left\| \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \right\|^2 \\
\leq \ & \frac{1}{2} \left(\sum_{i=1}^n \eta_i \left\| \mathbf{x}_i^k - \mathbf{x}_i^* \right\|^2 + \beta_k^{-2} \left\| \lambda^k - \lambda^* \right\|^2 \right) \\
& - \frac{1}{2} \left(\sum_{i=1}^n \eta_i \left\| \mathbf{x}_i^{k+1} - \mathbf{x}_i^* \right\|^2 + \beta_{k+1}^{-2} \left\| \lambda^{k+1} - \lambda^* \right\|^2 \right).
\end{aligned} \tag{77}$$

Summing the above inequalities from $k = 0$ to $k = K$, and dividing both sides with $\sum_{k=0}^K \beta_k^{-1}$, we have

$$\sum_{k=0}^K \gamma_k f(\mathbf{x}^{k+1}) - f(\mathbf{x}^*) + \sum_{i=1}^n \left\langle \sum_{k=0}^K \gamma_k \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathcal{A}_i^*(\lambda^*) \right\rangle \tag{78}$$

$$+ \frac{\alpha \beta_0}{2} \sum_{k=0}^K \gamma_k \left\| \sum_{i=1}^n \mathcal{A}_i(\mathbf{x}_i^{k+1}) - \mathbf{b} \right\|^2 \tag{79}$$

$$\leq \frac{1}{2 \sum_{k=0}^K \beta_k^{-1}} \left(\sum_{i=1}^n \eta_i \left\| \mathbf{x}_i^0 - \mathbf{x}_i^* \right\|^2 + \beta_0^{-2} \left\| \lambda^0 - \lambda^* \right\|^2 \right). \tag{80}$$

Next, by the convexity of f and the squared F-norm $\|\cdot\|^2$, we have

$$\begin{aligned}
& f(\bar{\mathbf{x}}^K) - f(\mathbf{x}^*) + \sum_{i=1}^{n} \left\langle \bar{\mathbf{x}}_i^K - \mathbf{x}_i^*, \mathcal{A}_i^*(\lambda^*) \right\rangle + \frac{\alpha\beta_0}{2} \left\| \sum_{i=1}^{n} \mathcal{A}_i(\bar{\mathbf{x}}_i^K) - \mathbf{b} \right\|^2 && (81)\\
\leq\ & \sum_{k=0}^{K} \gamma_k f(\mathbf{x}^{k+1}) - f(\mathbf{x}^*) + \sum_{i=1}^{n} \left\langle \sum_{k=0}^{K} \gamma_k \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathcal{A}_i^*(\lambda^*) \right\rangle && (82)\\
& + \frac{\alpha\beta_0}{2} \sum_{k=0}^{K} \gamma_k \left\| \sum_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i^{k+1}) - \mathbf{b} \right\|^2 . && (83)
\end{aligned}$$

Combining (78)-(80) and (81)-(83), we have

$$\begin{aligned}
& f(\bar{\mathbf{x}}^K) - f(\mathbf{x}^*) + \sum_{i=1}^{n} \left\langle \bar{\mathbf{x}}_i^K - \mathbf{x}_i^*, \mathcal{A}_i^*(\lambda^*) \right\rangle + \frac{\alpha\beta_0}{2} \left\| \sum_{i=1}^{n} \mathcal{A}_i(\bar{\mathbf{x}}_i^K) - \mathbf{b} \right\|^2 && (84)\\
\leq\ & \frac{1}{2\sum_{k=0}^{K} \beta_k^{-1}} \left(\sum_{i=1}^{n} \eta_i \left\| \mathbf{x}_i^0 - \mathbf{x}_i^* \right\|^2 + \beta_0^{-2} \left\| \lambda^0 - \lambda^* \right\|^2 \right). && (85)
\end{aligned}$$

■

In real applications, we are often faced with convex programs with convex set constraints:

$$\min_{\mathbf{x}_1,\cdots,\mathbf{x}_n} \sum_{i=1}^{n} f_i(\mathbf{x}_i), \quad s.t. \quad \sum_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i) = \mathbf{b}, \quad \mathbf{x}_i \in X_i, i = 1, \cdots, n, \tag{86}$$

where $X_i \subseteq \mathbb{R}^{d_i}$ is a closed convex set. In this section, we assume that the projections onto X_i's are all easily computable. For many convex sets used in machine learning, such an assumption is valid, e.g., X_i's are nonnegative cones or positive semi-definite cones. In the following, we discuss how to solve (86) efficiently. For simplicity, we assume that $X_i \neq \mathbb{R}^{d_i}$, $\forall i$.

We introduce auxiliary variables $\mathbf{x}_{n+i}$ to convert $\mathbf{x}_i \in X_i$ into $\mathbf{x}_i = \mathbf{x}_{n+i}$ and $\mathbf{x}_{n+i} \in X_i$, $i = 1, \cdots, n$. Then (86) can be reformulated as an equivalent one:

$$\min_{\mathbf{x}_1,\cdots,\mathbf{x}_{2n}} \sum_{i=1}^{2n} f_i(\mathbf{x}_i), \quad s.t. \quad \sum_{i=1}^{2n} \hat{\mathcal{A}}_i(\mathbf{x}_i) = \hat{\mathbf{b}}, \tag{87}$$

where $f_{n+i}(\mathbf{x}) \equiv \chi_{X_i}(\mathbf{x}) = \begin{cases} 0, & \text{if } \mathbf{x} \in X_i, \\ +\infty, & \text{otherwise} \end{cases}$, is the characteristic function of X_i and

$$\begin{aligned}
\hat{\mathcal{A}}_i(\mathbf{x}_i) &= (\mathcal{A}_i(\mathbf{x}_i), 0, \cdots, \mathbf{x}_i, \cdots, 0)^T, \\
\hat{\mathcal{A}}_{n+i}(\mathbf{x}_{n+i}) &= (0, 0, \cdots, -\mathbf{x}_{n+i}, \cdots, 0)^T, \\
& \quad i = 1, \cdots, n, \\
\hat{\mathbf{b}} &= (\mathbf{b}, 0, \cdots, 0, \cdots, 0)^T.
\end{aligned} \tag{88}$$

The adjoint operator $\hat{\mathcal{A}}_i^*$ is

$$\begin{array}{ll} \hat{\mathcal{A}}_i^*(\mathbf{y}) = \mathcal{A}_i^*(\mathbf{y}_1) + \mathbf{y}_{i+1}, & i = 1, \cdots, n, \\ \hat{\mathcal{A}}_{n+i}^*(\mathbf{y}) = -\mathbf{y}_{i+1}, & \end{array} \tag{89}$$

where $\mathbf{y}_i$ is the i-th sub-vector of $\mathbf{y}$, partitioned according to the sizes of $\mathbf{b}$ and $\mathbf{x}_i$, $i = 1, \cdots, n$.

Theorem 13 *(Theorem 11 in the manuscript) For problem (87), if $\{\beta_k\}$ is non-decreasing and upper bounded and η_i's are chosen as $\eta_i > n\|\mathcal{A}_i\|^2 + 2$ and $\eta_{n+i} > 2$, $i = 1, \cdots, n$, then the sequence $\{(\{\mathbf{x}_i^k\}, \lambda^k)\}$ generated by LADMPSAP converges to a KKT point of problem (87).*

We only need to prove the following proposition. Then by the same technique for proving Theorem 7, we can prove Theorem 13.

Proposition 14

$$\beta_k \sum_{i=1}^{2n} \sigma_i^{(k)} \|\mathbf{x}_i^{k+1} - \mathbf{x}^*\|^2 + \|\lambda^{k+1} - \lambda^*\|^2 \tag{90}$$

$$\leq \beta_k \sum_{i=1}^{n} \sigma_i^{(k)} \|\mathbf{x}_i^{k} - \mathbf{x}^*\|^2 + \|\lambda^{k} - \lambda^*\|^2 \tag{91}$$

$$-2\beta_k \sum_{i=1}^{2n} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}^*, -\sigma_i^{(k)}(\mathbf{x}_i^{k+1} - \mathbf{u}_i^k) + \hat{\mathcal{A}}_i^*(\lambda^*) \right\rangle \tag{92}$$

$$-\beta_k \sum_{i=1}^{n} \left(\sigma_i^{(k)} - \beta_k(n\|\mathcal{A}_i\|^2 + 2)\right) \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2 \tag{93}$$

$$-\beta_k \sum_{i=n+1}^{2n} \left(\sigma_i^{(k)} - 2\beta_k\right) \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2 \tag{94}$$

$$-\|\lambda^k - \hat{\lambda}^k\|^2. \tag{95}$$

Proof We continue from (40):

$$-2\beta_k \sum_{i=1}^{2n} \sigma_i^{(k)} \left\langle \mathbf{x}_i^{k+1} - \mathbf{x}_i^*, \mathbf{x}_i^k - \mathbf{u}_i^k \right\rangle + 2 \left\langle \lambda^{k+1} - \lambda^k, \lambda^{k+1} \right\rangle \tag{96}$$

$$= \|\lambda^{k+1} - \lambda^k\|^2 + \beta_k^2 \left\| \sum_{i=1}^{2n} \hat{\mathcal{A}}_i(\mathbf{x}_i^{k+1} - \mathbf{x}_i^k) \right\|^2 - \|\lambda^k - \hat{\lambda}^k\|^2 \tag{97}$$

$$= \|\lambda^{k+1} - \lambda^k\|^2 + \beta_k^2 \left\| \sum_{i=1}^{n} \mathcal{A}_i(\mathbf{x}_i^{k+1} - \mathbf{x}_i^k) \right\|^2 \tag{98}$$

$$+\beta_k^2 \sum_{i=1}^{n} \left\| (\mathbf{x}_i^{k+1} - \mathbf{x}_i^k) - (\mathbf{x}_{n+i}^{k+1} - \mathbf{x}_{n+i}^k) \right\|^2 - \|\lambda^k - \hat{\lambda}^k\|^2 \tag{99}$$

$$\leq \|\lambda^{k+1} - \lambda^k\|^2 + n\beta_k^2 \sum_{i=1}^{n} \|\mathcal{A}_i\|^2 \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2 \tag{100}$$

$$+2\beta_k^2 \sum_{i=1}^{n} \left(\|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|^2 + \|\mathbf{x}_{n+i}^{k+1} - \mathbf{x}_{n+i}^k)\|^2 \right) - \|\lambda^k - \hat{\lambda}^k\|^2. \tag{101}$$

Then we can have (90)-(95). ■

4. Solving Latent LRR via APG

When using APG Beck and Teboulle (2009), the latent LRR problem has to be reformulated into an unconstrained optimization problem as follows:

$$\min_{\mathbf{Z},\mathbf{L},\mathbf{E}} \gamma(\|\mathbf{Z}\|_* + \|\mathbf{L}\|_* + \mu\|\mathbf{E}\|_1) + \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{Z} - \mathbf{L}\mathbf{X} - \mathbf{E}\|^2, \tag{102}$$

where $\gamma > 0$ is a parameter that controls the closeness between (102) and the original latent LRR problem.

The updating schemes of APG are as follows:

$$\bar{\mathbf{E}}^k = \mathbf{E}^k + \frac{t_{k-1} - 1}{t_k}(\mathbf{E}^k - \mathbf{E}^{k-1}), \tag{103a}$$

$$\bar{\mathbf{Z}}^k = \mathbf{Z}^k + \frac{t_{k-1} - 1}{t_k}(\mathbf{Z}^k - \mathbf{Z}^{k-1}), \tag{103b}$$

$$\bar{\mathbf{L}}^k = \mathbf{L}^k + \frac{t_{k-1} - 1}{t_k}(\mathbf{L}^k - \mathbf{L}^{k-1}), \tag{103c}$$

$$\mathbf{E}^{k+1} = \arg\min_{\mathbf{E}} \mu\gamma\|\mathbf{E}\|_1 + \frac{\tau}{2} \left\| \mathbf{E} - \left(\bar{\mathbf{E}}^k - \frac{1}{2\tau} \frac{\partial}{\partial \mathbf{E}} \|\mathbf{X} - \mathbf{X}\mathbf{Z} - \mathbf{L}\mathbf{X} - \mathbf{E}\|^2 \Big|_{\begin{pmatrix} \bar{\mathbf{E}}^k \\ \bar{\mathbf{Z}}^k \\ \bar{\mathbf{L}}^k \end{pmatrix}} \right) \right\|^2, \tag{103d}$$

$$\mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}} \gamma\|\mathbf{Z}\|_* + \frac{\tau}{2}\left\|\mathbf{Z} - \left(\bar{\mathbf{Z}}^k - \frac{1}{2\tau}\frac{\partial}{\partial \mathbf{Z}}\|\mathbf{X}-\mathbf{XZ}-\mathbf{LX}-\mathbf{E}\|^2\Big|_{\begin{pmatrix}\bar{\mathbf{E}}^k\\ \bar{\mathbf{Z}}^k\\ \bar{\mathbf{L}}^k\end{pmatrix}}\right)\right\|^2, \quad (103e)$$

$$\mathbf{L}^{k+1} = \arg\min_{\mathbf{L}} \gamma\|\mathbf{L}\|_* + \frac{\tau}{2}\left\|\mathbf{L} - \left(\bar{\mathbf{L}}^k - \frac{1}{2\tau}\frac{\partial}{\partial \mathbf{L}}\|\mathbf{X}-\mathbf{XZ}-\mathbf{LX}-\mathbf{E}\|^2\Big|_{\begin{pmatrix}\bar{\mathbf{E}}^k\\ \bar{\mathbf{Z}}^k\\ \bar{\mathbf{L}}^k\end{pmatrix}}\right)\right\|^2, \quad (103f)$$

$$t_{k+1} = \frac{1+\sqrt{4t_k^2+1}}{2}, \quad (103g)$$

where τ is a Lipschitz constant such that

$$\left\| \frac{1}{2}\begin{pmatrix}\frac{\partial}{\partial \mathbf{E}}\|\mathbf{X}-\mathbf{XZ}-\mathbf{LX}-\mathbf{E}\|^2\\ \frac{\partial}{\partial \mathbf{Z}}\|\mathbf{X}-\mathbf{XZ}-\mathbf{LX}-\mathbf{E}\|^2\\ \frac{\partial}{\partial \mathbf{L}}\|\mathbf{X}-\mathbf{XZ}-\mathbf{LX}-\mathbf{E}\|^2\end{pmatrix}\Bigg|_{\begin{pmatrix}\mathbf{E}_1\\ \mathbf{Z}_1\\ \mathbf{L}_1\end{pmatrix}} - \frac{1}{2}\begin{pmatrix}\frac{\partial}{\partial \mathbf{E}}\|\mathbf{X}-\mathbf{XZ}-\mathbf{LX}-\mathbf{E}\|^2\\ \frac{\partial}{\partial \mathbf{Z}}\|\mathbf{X}-\mathbf{XZ}-\mathbf{LX}-\mathbf{E}\|^2\\ \frac{\partial}{\partial \mathbf{L}}\|\mathbf{X}-\mathbf{XZ}-\mathbf{LX}-\mathbf{E}\|^2\end{pmatrix}\Bigg|_{\begin{pmatrix}\mathbf{E}_2\\ \mathbf{Z}_2\\ \mathbf{L}_2\end{pmatrix}} \right\|$$

$$\leq \quad \tau\left\|\begin{pmatrix}\mathbf{E}_1\\ \mathbf{Z}_1\\ \mathbf{L}_1\end{pmatrix} - \begin{pmatrix}\mathbf{E}_2\\ \mathbf{Z}_2\\ \mathbf{L}_2\end{pmatrix}\right\|, \quad \forall \begin{pmatrix}\mathbf{E}_1\\ \mathbf{Z}_1\\ \mathbf{L}_1\end{pmatrix} \text{ and } \begin{pmatrix}\mathbf{E}_2\\ \mathbf{Z}_2\\ \mathbf{L}_2\end{pmatrix}. \quad (104)$$

By the inequality $\|\mathbf{AB}\| \leq \|\mathbf{A}\|_2\|\mathbf{B}\|$, we can prove that $\tau \geq \sqrt{3\max(1,\|\mathbf{X}\|_2^2)(1+2\|\mathbf{X}\|_2^2)}$ ensures the correctness of (104), where $\|\mathbf{A}\|_2 = \sigma_{\max}(\mathbf{A})$ is the largest singular value of $\mathbf{A}$.

The pseudo code of the APG approach for latent LRR, with the continuation technique (step 8 in Algorithm 1, which reduces γ gradually), is provided in Algorithm 1.

For the parameters of APG, we follow the suggestions in Lin et al. (2009) to set $\gamma_0 = 0.99\|\mathbf{X}\|_2$, $\gamma_{\min} = 10^{-10}$, and $\theta = 0.9$.

5. Solving Latent LRR via Naive ADM

Naive ADM is to introduce auxiliary variables to the original latent LRR model, such that each subproblem can have a closed form solution. The equivalent problem to solve is:

$$\min_{\mathbf{J},\mathbf{S},\mathbf{E},\mathbf{Z},\mathbf{L}} \|\mathbf{J}\|_* + \|\mathbf{S}\|_* + \mu\|\mathbf{E}\|_1, \quad s.t. \quad \mathbf{X} = \mathbf{XZ}+\mathbf{LX}+\mathbf{E}, \mathbf{Z}=\mathbf{J}, \mathbf{L}=\mathbf{S}. \quad (105)$$

Naive ADM operates on the augmented Lagrangian function of problem (105):

$$\|\mathbf{J}\|_* + \|\mathbf{S}\|_* + \mu\|\mathbf{E}\|_1 \quad (106)$$

$$+\langle \mathbf{Y}_1, \mathbf{X}-\mathbf{XZ}-\mathbf{LX}-\mathbf{E}\rangle + \langle \mathbf{Y}_2, \mathbf{Z}-\mathbf{J}\rangle + \langle \mathbf{Y}_3, \mathbf{L}-\mathbf{S}\rangle \quad (107)$$

$$+\frac{\beta}{2}\left(\|\mathbf{X}-\mathbf{XZ}-\mathbf{LX}-\mathbf{E}\|^2 + \|\mathbf{Z}-\mathbf{J}\|^2 + \|\mathbf{L}-\mathbf{S}\|^2\right), \quad (108)$$

Algorithm 1 APG for Latent LRR

Input: Observation matrix $\mathbf{X}$ and parameter $\mu > 0$.

Set Parameter Values: $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $\gamma_0 \gg \gamma_{\min} > 0$, $\theta \in (0,1)$, and $\tau = \sqrt{3\max(1, \|\mathbf{X}\|_2^2)(1+2\|\mathbf{X}\|_2^2)}$.

Initialize: Set $\mathbf{E}^0 = \mathbf{E}^{-1} = \mathbf{0}$, $\mathbf{Z}^0 = \mathbf{Z}^{-1} = \mathbf{0}$, $t_0 = t_{-1} = 1$, and $k \leftarrow 0$.

while not converged **do**

Step 1: Update $\bar{\mathbf{E}}^k = \mathbf{E}^k + \frac{t_{k-1}-1}{t_k}(\mathbf{E}^k - \mathbf{E}^{k-1})$.

Step 2: Update $\bar{\mathbf{Z}}^k = \mathbf{Z}^k + \frac{t_{k-1}-1}{t_k}(\mathbf{Z}^k - \mathbf{Z}^{k-1})$.

Step 3: Update $\bar{\mathbf{L}}^k = \mathbf{L}^k + \frac{t_{k-1}-1}{t_k}(\mathbf{L}^k - \mathbf{L}^{k-1})$.

Step 4: Update $\mathbf{E}^{k+1} = \mathcal{S}_{\frac{\mu\gamma_k}{\tau}}(\bar{\mathbf{E}}^k + \frac{1}{\tau}(\mathbf{X} - \mathbf{X}\bar{\mathbf{Z}}^k - \bar{\mathbf{L}}^k\mathbf{X} - \bar{\mathbf{E}}^k))$, where $\mathcal{S}$ is the shrinkage operator.

Step 5: Update $\mathbf{Z}^{k+1} = \mathbf{U}_Z\mathcal{S}_{\frac{\gamma_k}{\tau}}(\mathbf{\Sigma}_Z)\mathbf{V}_Z^T$, where $\mathbf{U}_Z\mathbf{\Sigma}_Z\mathbf{V}_Z^T$ is the SVD of $\bar{\mathbf{Z}}^k + \frac{1}{\tau}\mathbf{X}^T(\mathbf{X} - \mathbf{X}\bar{\mathbf{Z}}^k - \bar{\mathbf{L}}^k\mathbf{X} - \bar{\mathbf{E}}^k)$.

Step 6: Update $\mathbf{L}^{k+1} = \mathbf{U}_L\mathcal{S}_{\frac{\gamma_k}{\tau}}(\mathbf{\Sigma}_L)\mathbf{V}_L^T$, where $\mathbf{U}_L\mathbf{\Sigma}_L\mathbf{V}_L^T$ is the SVD of $\bar{\mathbf{L}}^k + \frac{1}{\tau}(\mathbf{X} - \mathbf{X}\bar{\mathbf{Z}}^k - \bar{\mathbf{L}}^k\mathbf{X} - \bar{\mathbf{E}}^k)\mathbf{X}^T$.

Step 7: Update $t_{k+1} = \frac{1+\sqrt{4t_k^2+1}}{2}$.

Step 8: Update $\gamma_{k+1} = \max(\gamma_{\min}, \theta\gamma_k)$.

Step 9: Check the convergence conditions:
$\frac{\|\mathbf{X}\mathbf{Z}^{k+1}+\mathbf{L}^{k+1}\mathbf{X}+\mathbf{E}^{k+1}-\mathbf{X}\|}{\|\mathbf{X}\|} \le \varepsilon_1$ and $\max\left(\frac{\|\mathbf{Z}^{k+1}-\mathbf{Z}^k\|}{\|\mathbf{X}\|}, \frac{\|\mathbf{L}^{k+1}-\mathbf{L}^k\|}{\|\mathbf{X}\|}, \frac{\|\mathbf{E}^{k+1}-\mathbf{E}^k\|}{\|\mathbf{X}\|}\right) \le \varepsilon_2$.
If they are satisfied, break.

Step 10: $k \leftarrow k+1$.

end while

where $\mathbf{Y}_1$, $\mathbf{Y}_2$, and $\mathbf{Y}_3$ are Lagrange multipliers.

The updating schemes of naive ADM are as follows:

$$\mathbf{J}^{k+1} = \underset{\mathbf{J}}{\operatorname{argmin}} \|\mathbf{J}\|_* + \frac{\beta_k}{2}\|\mathbf{Z}^k - \mathbf{J} + \mathbf{Y}_2^k/\beta_k\|^2, \tag{109a}$$

$$\mathbf{S}^{k+1} = \underset{\mathbf{S}}{\operatorname{argmin}} \|\mathbf{S}\|_* + \frac{\beta_k}{2}\|\mathbf{L}^k - \mathbf{S} + \mathbf{Y}_3^k/\beta_k\|^2, \tag{109b}$$

$$\mathbf{E}^{k+1} = \underset{\mathbf{E}}{\operatorname{argmin}}\, \mu\|\mathbf{E}\|_1 + \frac{\beta_k}{2}\|\mathbf{X} - \mathbf{X}\mathbf{Z}^k - \mathbf{L}^k\mathbf{X} - \mathbf{E} + \mathbf{Y}_1^k/\beta_k\|^2, \tag{109c}$$

$$\mathbf{Z}^{k+1} = \underset{\mathbf{Z}}{\operatorname{argmin}} \frac{\beta_k}{2}\left(\|\mathbf{X} - \mathbf{X}\mathbf{Z} - \mathbf{L}^k\mathbf{X} - \mathbf{E}^{k+1} + \mathbf{Y}_1^k/\beta_k\|^2 + \|\mathbf{Z} - \mathbf{J}^{k+1} + \mathbf{Y}_2^k/\beta_k\|^2\right), \tag{109d}$$

$$\mathbf{L}^{k+1} = \underset{\mathbf{L}}{\operatorname{argmin}} \frac{\beta_k}{2}\left(\|\mathbf{X} - \mathbf{X}\mathbf{Z}^{k+1} - \mathbf{L}\mathbf{X} - \mathbf{E}^{k+1} + \mathbf{Y}_1^k/\beta_k\|^2 + \|\mathbf{L} - \mathbf{S}^{k+1} + \mathbf{Y}_3^k/\beta_k\|^2\right), \tag{109e}$$

$$\mathbf{Y}_1^{k+1} = \mathbf{Y}_1^k + \beta_k(\mathbf{X} - \mathbf{X}\mathbf{Z}^{k+1} - \mathbf{L}^{k+1}\mathbf{X} - \mathbf{E}^{k+1}), \tag{109f}$$

$$\mathbf{Y}_2^{k+1} = \mathbf{Y}_2^k + \beta_k(\mathbf{Z}^{k+1} - \mathbf{J}^{k+1}), \tag{109g}$$

$$\mathbf{Y}_3^{k+1} = \mathbf{Y}_3^k + \beta_k(\mathbf{L}^{k+1} - \mathbf{S}^{k+1}), \tag{109h}$$

$$\beta_{k+1} = \min(\beta_{\max}, \rho\beta_k), \tag{109i}$$

where $\rho > 1$ and $\beta_{\max} > 0$ are constants.

The pseudo code of the naive ADM approach for latent LRR is provided in Algorithm 2.

Algorithm 2 Naive ADM for Latent LRR

Input: Observation matrix $\mathbf{X}$ and parameter $\mu > 0$.
Set Parameter Values: $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $0 < \beta_0 \ll 1 \ll \beta_{\max}$, and $\rho > 1$.
Initialize: Set $\mathbf{J}^0 = \mathbf{S}^0 = \mathbf{E}^0 = \mathbf{L}^0 = \mathbf{S}^0 = \mathbf{0}$ and $k \leftarrow 0$.
while not converged **do**
Step 1: Update $\mathbf{J}^{k+1} = \mathbf{U}_Z \mathcal{S}_{\frac{1}{\beta_k}}(\mathbf{\Sigma}_Z)\mathbf{V}_Z^T$, where $\mathbf{U}_Z\mathbf{\Sigma}_Z\mathbf{V}_Z^T$ is the SVD of $\mathbf{Z}^k + \mathbf{Y}_2^k/\beta_k$ and $\mathcal{S}$ is the shrinkage operator.
Step 2: Update $\mathbf{S}^{k+1} = \mathbf{U}_L \mathcal{S}_{\frac{1}{\beta_k}}(\mathbf{\Sigma}_L)\mathbf{V}_L^T$, where $\mathbf{U}_L\mathbf{\Sigma}_L\mathbf{V}_L^T$ is the SVD of $\mathbf{L}^k + \mathbf{Y}_3^k/\beta_k$.
Step 3: Update $\mathbf{E}^{k+1} = \mathcal{S}_{\frac{\mu}{\beta_k}}(\mathbf{X} - \mathbf{X}\mathbf{Z}^k - \mathbf{L}^k\mathbf{X} + \mathbf{Y}_1^k/\beta_k)$.
Step 4: Update $\mathbf{Z}^{k+1} = (\mathbf{I}+\mathbf{X}^T\mathbf{X})^{-1}[\mathbf{X}^T(\mathbf{X}-\mathbf{L}^k\mathbf{X}-\mathbf{E}^{k+1}+\mathbf{Y}_1^k/\beta_k)+\mathbf{J}^{k+1}-\mathbf{Y}_2^k/\beta_k]$.
Step 5: Update $\mathbf{L}^{k+1} = [(\mathbf{X}-\mathbf{X}\mathbf{Z}^{k+1}-\mathbf{E}^{k+1}+\mathbf{Y}_1^k/\beta_k)\mathbf{X}^T+\mathbf{S}^{k+1}-\mathbf{Y}_3^k/\beta_k](\mathbf{I}+\mathbf{X}\mathbf{X}^T)^{-1}$.
Step 6: Update $\mathbf{Y}_1^{k+1} = \mathbf{Y}_1^k + \beta_k(\mathbf{X} - \mathbf{X}\mathbf{Z}^{k+1} - \mathbf{L}^{k+1}\mathbf{X} - \mathbf{E}^{k+1})$.
Step 7: Update $\mathbf{Y}_2^{k+1} = \mathbf{Y}_2^k + \beta_k(\mathbf{Z}^{k+1} - \mathbf{J}^{k+1})$.
Step 8: Update $\mathbf{Y}_3^{k+1} = \mathbf{Y}_3^k + \beta_k(\mathbf{L}^{k+1} - \mathbf{S}^{k+1})$.
Step 9: Update $\beta_{k+1} = \min(\beta_{\max}, \rho\beta_k)$.
Step 10: Check the convergence conditions:
$\frac{\|\mathbf{X}\mathbf{Z}^{k+1}+\mathbf{L}^{k+1}\mathbf{X}+\mathbf{E}^{k+1}-\mathbf{X}\|}{\|\mathbf{X}\|} \le \varepsilon_1$ and $\max\left(\frac{\|\mathbf{Z}^{k+1}-\mathbf{Z}^k\|}{\|\mathbf{X}\|}, \frac{\|\mathbf{L}^{k+1}-\mathbf{L}^k\|}{\|\mathbf{X}\|}, \frac{\|\mathbf{E}^{k+1}-\mathbf{E}^k\|}{\|\mathbf{X}\|}\right) \le \varepsilon_2$.
If they are satisfied, break.
Step 11: $k \leftarrow k + 1$.
end while

Naive ADM has been described in Liu and Yan (2011) and was called inexact ALM therein. The parameters of naive ADM are the same as those in Liu and Yan (2011): $\beta_0 = 10^{-6}$, $\beta_{\max} = 10^6$, and $\rho = 1.1$.

6. Solving Latent LRR via LADMGB and Naive LADM

LADMGB He and Yuan (2013) consists of two steps. The first step is update the variables by LADM (prediction step). The second step is to update the variables by Gaussian back substitution (correction step), which updates the variables in reverse order.

The updating schemes of LADMGB are as follows:

LADM step:

$$\tilde{\mathbf{Z}}^k = \operatorname*{argmin}_{\mathbf{Z}} \|\mathbf{Z}\|_* + \frac{\beta\eta_Z}{2}\|\mathbf{Z} - \mathbf{Z}^k + \mathbf{X}^T(\lambda^k + \beta(\mathbf{X}\mathbf{Z}^k + \mathbf{L}^k\mathbf{X} + \mathbf{E}^k - \mathbf{X})/(\beta\eta_Z)\|^2, \tag{110a}$$

$$\tilde{\mathbf{L}}^k = \operatorname*{argmin}_{\mathbf{L}} \|\mathbf{L}\|_* + \frac{\beta\eta_L}{2}\|\mathbf{L} - \mathbf{L}^k + (\lambda^k + \beta(\mathbf{X}\tilde{\mathbf{Z}}^k + \mathbf{L}^k\mathbf{X} + \mathbf{E}^k - \mathbf{X})\mathbf{X}^T/(\beta\eta_L)\|^2, \tag{110b}$$

$$\tilde{\mathbf{E}}^k = \underset{\mathbf{E}}{\operatorname{argmin}}\, \mu\|\mathbf{E}\|_1 + \frac{\beta}{2}\|\mathbf{E} - \mathbf{E}^k + (\lambda^k + \beta(\mathbf{X}\tilde{\mathbf{Z}}^k + \tilde{\mathbf{L}}^k\mathbf{X} + \mathbf{E}^k - \mathbf{X})/\beta)\|^2, \tag{110c}$$

$$\tilde{\lambda}^k = \lambda^k + \beta(\mathbf{X}\tilde{\mathbf{Z}}^k + \tilde{\mathbf{L}}^k\mathbf{X} + \tilde{\mathbf{E}}^k - \mathbf{X}). \tag{110d}$$

GB step:

$$\lambda^{k+1} = \lambda^k + \alpha(\tilde{\lambda}^k - \lambda^k), \tag{111a}$$

$$\mathbf{E}^{k+1} = \mathbf{E}^k + \alpha(\tilde{\mathbf{E}}^k - \mathbf{E}^k), \tag{111b}$$

$$\mathbf{L}^{k+1} = \mathbf{L}^k - (\mathbf{E}^{k+1} - \mathbf{E}^k)\mathbf{X}^T/\eta_L + \alpha(\tilde{\mathbf{L}}^k - \mathbf{L}^k), \tag{111c}$$

$$\mathbf{Z}^{k+1} = \mathbf{Z}^k - \mathbf{X}^T(\mathbf{L}^{k+1} - \mathbf{L}^k)\mathbf{X}/\eta_Z - \mathbf{X}^T(\mathbf{E}^{k+1} - \mathbf{E}^k)/\eta_Z + \alpha(\tilde{\mathbf{Z}}^k - \mathbf{Z}^k). \tag{111d}$$

The pseudo code of the LADMGB approach for latent LRR is provided in Algorithm 3.

Algorithm 3 LADMGB for Latent LRR

Input: Observation matrix $\mathbf{X}$ and parameter $\mu > 0$.
Set Parameter Values: $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $\alpha \in (0,1)$, $\eta_Z = \|\mathbf{X}\|_2^2$, and $\beta > 0$.
Initialize: Set $\mathbf{L}^0 = \mathbf{S}^0 = \mathbf{E}^0 = \mathbf{0}$ and $k \leftarrow 0$.
while not converged **do**

Step 1: Update $\tilde{\mathbf{Z}}^k = \mathbf{U}_Z\mathcal{S}_{\frac{1}{\beta_k\eta_Z}}(\mathbf{\Sigma}_Z)\mathbf{V}_Z^T$, where $\mathbf{U}_Z\mathbf{\Sigma}_Z\mathbf{V}_Z^T$ is the SVD of $\mathbf{Z}^k - \mathbf{X}^T(\lambda^k + \beta(\mathbf{X}\mathbf{Z}^k + \mathbf{L}^k\mathbf{X} + \mathbf{E}^k - \mathbf{X})/(\beta\eta_Z)$ and $\mathcal{S}$ is the shrinkage operator.
Step 2: Update $\tilde{\mathbf{L}}^k = \mathbf{U}_L\mathcal{S}_{\frac{1}{\beta_k\eta_L}}(\mathbf{\Sigma}_L)\mathbf{V}_L^T$, where $\mathbf{U}_L\mathbf{\Sigma}_L\mathbf{V}_L^T$ is the SVD of $\mathbf{L}^k - (\lambda^k + \beta(\mathbf{X}\tilde{\mathbf{Z}}^k + \mathbf{L}^k\mathbf{X} + \mathbf{E}^k - \mathbf{X})\mathbf{X}^T/(\beta\eta_L)$.
Step 3: Update $\tilde{\mathbf{E}}^k = \mathcal{S}_{\frac{\mu}{\beta_k}}(\mathbf{E}^k - (\lambda^k + \beta(\mathbf{X}\tilde{\mathbf{Z}}^k + \tilde{\mathbf{L}}^k\mathbf{X} + \mathbf{E}^k - \mathbf{X})/\beta))$.
Step 4: Update $\tilde{\lambda}^k = \lambda^k + \beta(\mathbf{X}\tilde{\mathbf{Z}}^k + \tilde{\mathbf{L}}^k\mathbf{X} + \tilde{\mathbf{E}}^k - \mathbf{X})$.
Step 5: Update $\lambda^{k+1} = \lambda^k + \alpha(\tilde{\lambda}^k - \lambda^k)$.
Step 6: Update $\mathbf{E}^{k+1} = \mathbf{E}^k + \alpha(\tilde{\mathbf{E}}^k - \mathbf{E}^k)$.
Step 7: Update $\mathbf{L}^{k+1} = \mathbf{L}^k - (\mathbf{E}^{k+1} - \mathbf{E}^k)\mathbf{X}^T/\eta_L + \alpha(\tilde{\mathbf{L}}^k - \mathbf{L}^k)$.
Step 8: Update $\mathbf{Z}^{k+1} = \mathbf{Z}^k - \mathbf{X}^T(\mathbf{L}^{k+1} - \mathbf{L}^k)\mathbf{X}/\eta_Z - \mathbf{X}^T(\mathbf{E}^{k+1} - \mathbf{E}^k)/\eta_Z + \alpha(\tilde{\mathbf{Z}}^k - \mathbf{Z}^k)$.

Step 9: Check the convergence conditions:
$\frac{\|\mathbf{X}\mathbf{Z}^{k+1} + \mathbf{L}^{k+1}\mathbf{X} + \mathbf{E}^{k+1} - \mathbf{X}\|}{\|\mathbf{X}\|} \le \varepsilon_1$ and $\max\left(\frac{\|\mathbf{Z}^{k+1} - \mathbf{Z}^k\|}{\|\mathbf{X}\|}, \frac{\|\mathbf{L}^{k+1} - \mathbf{L}^k\|}{\|\mathbf{X}\|}, \frac{\|\mathbf{E}^{k+1} - \mathbf{E}^k\|}{\|\mathbf{X}\|}\right) \le \varepsilon_2$.
If they are satisfied, break.
Step 11: $k \leftarrow k + 1$.

end while

The naive LADM for latent LRR is just the LADM part of LADMGB. So we omit the details.

For LADM, we follow the suggestions in Yang and Yuan (2013) to fix its penalty parameter β at $2.5/\min(d, sp)$, where $d \times sp$ is the size of $\mathbf{X}$. For LADMGB, He et al. He and Yuan (2013) suggested $\alpha = 0.8$ but did not suggest how to choose a fixed β. So we follow the suggestions in Yang and Yuan (2013) to fix β at $2.5/\min(d, sp)$.

References

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009.

B. S. He and X. Yuan. Linearized alternating direction method with Gaussian back substitution for separable convex programming. *Numerical Algebra, Control and Optimization*, 3(2):247–260, 2013.

Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. *UIUC Technical Report UILU-ENG-09-2214*, 2009.

Z. Lin, R. Liu, and Z. Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *NIPS*, 2011.

G. Liu and S. Yan. Latent low-rank representation for subspace segmentation and feature extraction. In *ICCV*, 2011.

R. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

J. Yang and X. Yuan. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. *Mathematics of Computation*, 82(281):301–329, 2013.