J. Appl. Numer. Optim. 7 (2025), No. 1, pp. 1-28
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.7.2025.1.01

SOLVING BASIS PURSUIT REVISITED: WHEN ADMM MEETS THE INHERENT SEPARABLE STRUCTURE

YIFAN WU[1], KAI WANG[2], HONGJIN HE[1,*]

[1]*School of Mathematics and Statistics, Ningbo University, Ningbo 315211, China*
[2]*School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing 210094, China*

Abstract. The Basis Pursuit (BP) problem refers to the task of finding a minimum ℓ_1-norm solution to an underdetermined linear system, which is a fundamental problem in compressed sensing. It has been demonstrate that the BP problem can be efficiently solved by using some state-of-the-art first-order optimization algorithms based on the Augmented Lagrangian method, which can successfully circumvent the difficulty caused by the nondifferentiable objective. In this paper, we revisit the implementation of the Alternating Direction Method of Multipliers (ADMM) to solve the BP problem. Notably, by reformulating the BP problem as a naturally separable minimization problem, without relying on additional auxiliary variables, we can obtain two more efficient BP solvers, which can save storage space to speed up the process of solving the BP problem. Some numerical results demonstrate the reliability and efficiency of our approach. Furthermore, we conclude that ADMM would be more efficient when the inherent separable structure of the problem is effectively exploited.
Keywords. Basis pursuit; compressed sensing; ADMM; first-order algorithm; nonsmooth optimization.
2020 Mathematics Subject Classification. 90C25, 90C30.

1. INTRODUCTION

Compressed sensing (see seminal papers [3, 5]) is a breakthrough theory for information acquisition and has been widely applied in numerous areas of engineering, including magnetic resonance imaging, communication networks, signal/image processing[1]. It is known that a fundamental problem in compressed sensing is the following combinatorial optimization problem:

$$\min_{\boldsymbol{u}\in\mathbb{R}^n} \left\{ \|\boldsymbol{u}\|_0 \mid C\boldsymbol{u} = b \right\}, \tag{1.1}$$

where $C \in \mathbb{R}^{m\times n}$ is the so-called sensing (measurements) matrix, $b \in \mathbb{R}^m$ ($m \ll n$) is an observation vector, and $\|\boldsymbol{u}\|_0$ refers to the number of nonzero elements of $\boldsymbol{u}$ (note that $\|\cdot\|_0$ is often called ℓ_0-norm, which is indeed a pseudo norm). Unfortunately, it has been proven that directly solving (1.1) is an NP-hard problem. A celebrated result shows that the tightest convex surrogate ℓ_1-norm, i.e., $\|\boldsymbol{u}\|_1 = |\boldsymbol{u}_1| + |\boldsymbol{u}_2| + \cdots + |\boldsymbol{u}_n|$ is able to approximate $\|\boldsymbol{u}\|_0$ well, thereby

*Corresponding author.
E-mail address: hehongjin@nbu.edu.cn (H. He).
Received 25 October 2024; Accepted 19 December 2024; Published online 20 March 2025.

[1]See more articles and other materials on compressed sensing from https://dsp.rice.edu/cs/

leading to the well-known Basis Pursuit (BP, see [4]) problem:

$$\min_{\boldsymbol{u}\in\mathbb{R}^n} \{\|\boldsymbol{u}\|_1 \mid C\boldsymbol{u} = b\}, \tag{BP}$$

which, under certain conditions, guarantees that its solution coincides with the solution of (1.1). Consequently, researchers are motivated to design efficient algorithms for solving (BP), with the goal of finding the sparsest solution to an underdetermined linear system under reasonable circumstances. In this paper, we are concerned with the solution methods for (BP) and will not pay attention to (1.1).

Utilizing the definition of ℓ_1-norm, it was documented in [4] that (BP) can be recast as a linear program (LP):

$$\min_{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{y}\in\mathbb{R}^n} \left\{\mathbf{1}^\top \boldsymbol{x} + \mathbf{1}^\top \boldsymbol{y} \mid C\boldsymbol{x} - C\boldsymbol{y} = b,\ \boldsymbol{x} \geq 0,\ \boldsymbol{y} \geq 0\right\}. \tag{LP}$$

As a consequence, all the algorithms tailored for LPs are also applicable to solving (BP). However, it can be easily seen that (LP) is twice the size of (BP) on variables. More importantly, there is an augmented coefficient matrix in (LP) being the twice scale of the one in (BP). Consequently, the LP formulation (LP) is not an ideal way to solve (BP), which was shown numerically in [18], and also motivates the authors of [18] to design a more efficient Spectral Gradient-Projection (SGP) method to solve the following three variants of (BP):

$$\begin{cases} \min_{\boldsymbol{u}\in\mathbb{R}^n} \{\|\boldsymbol{u}\|_1 \mid \|C\boldsymbol{u} - b\| \leq \delta\}, \\ \min_{\boldsymbol{u}\in\mathbb{R}^n} \left\{\|\boldsymbol{u}\|_1 + \frac{\lambda}{2}\|C\boldsymbol{u} - b\|^2\right\}, \\ \min_{\boldsymbol{u}\in\mathbb{R}^n} \{\|C\boldsymbol{u} - b\| \mid \|\boldsymbol{u}\|_1 \leq \tau\}, \end{cases} \tag{1.2}$$

where $\|\cdot\|$ denotes the standard Euclidean ℓ_2-norm (i.e., $\|\boldsymbol{u}\| = \sqrt{\boldsymbol{u}^\top\boldsymbol{u}}$), $\delta > 0$, $\lambda > 0$, and $\tau > 0$ are appropriate tuning parameters. In the last decades, efficient solvers were developed for (1.2), e.g., ℓ_1-Homotopy [16, 17], Lars [7], ℓ_1-Magic[2], GPSR [8], FISTA [1], FPC [19], NESTA [2], YALL1 [21], to mention just a few. Here, we refer the reader to [14] for an extensive numerical comparison among some of the above state-of-the-art ℓ_1-solvers.

Although the three variants in (1.2) have the extra ability to deal with more general cases with Gaussian noise than (BP), choosing perfect tuning parameters δ, λ, and τ, which are important for the efficiency of ℓ_1-solvers to seek the sparsest solution to an underdetermined linear system, is usually not an easy task. Comparatively, (BP) is a parameter-free model so that we do not worry about how to choose model parameters. In this paper, we aim to solve (BP) directly by efficiently exploring its inherent favorable structure. Specifically, we first divide the original decision variable and coefficient matrix into two blocks. Then, we reformulate (BP) as a two-block optimization problem with a naturally separable structure. Unlike (LP), it is remarkable that our reformulation does not increase the scale of (BP), which is of importance for saving storage space when implementing algorithms to solve it. Due to the resulting separable structure, we can gainfully employ the state-of-the-art ADMM to solve the resulting formulation directly. When comparing with the existing application of ADMM to (BP) and its variants (see [21] and [22]), our numerical results show that applying ADMM to the new formulation of (BP) greatly reduces the number of iterations. We notice that, although our new application of ADMM requires an inner loop for its subproblems, it usually runs faster than the earlier

[2] http://www.l1-magic.org/

versions of ADMM in terms of computing time when their subproblems are inexactly solved. To make the ADMM more implementable, we further introduce a new fully linearized ADMM, which allows a larger stepsize for the second subproblem than the first one, to solve the resulting formulation of (BP). It is promising from the reported results that the fully linearized ADMM can further save much computing time to solve (BP). The computational results tell us how to explore the favorable inherent structure of a problem under consideration. In other words, we must prioritize exploring the inherent separable structure, rather than introducing additional auxiliary variables to artificially construct a separable form, when implementing ADMM to find a numerical solution for the problem under consideration.

The rest of this paper is organized as follows. In Section 2, we give a brief review on several Augmented Lagrangian Method (ALM) based optimization algorithms for (BP). In Section 3, we first give a two-block reformulation of (BP). Then, we show a direct application of ADMM to the resulting formulation and accordingly propose a fully linearized ADMM for the resulting two-block optimization model. In Section 4, we give an extensive numerical comparison of some ALM-based ℓ_1-solvers. Finally, some conclusions are given in Section 5 to complete this paper.

Notation. Throughout this paper, we let $\mathbb{R}^n$ be the n-dimensional Euclidean space endowed with the standard inner product $\langle\cdot,\cdot\rangle$. The superscript $^\top$ represents the transpose of vectors and matrices. For a given symmetric and positive (semi-) definite matrix H (denoted by $H(\succeq)\succ 0$), we let $\|\boldsymbol{x}\|_H=\sqrt{\langle \boldsymbol{x},H\boldsymbol{x}\rangle}\equiv\sqrt{\boldsymbol{x}^\top H\boldsymbol{x}}$ be the H-norm of $\boldsymbol{x}\in\mathbb{R}^n$. Moreover, we denote $\|\cdot\|_\infty$ as the ℓ_∞-norm of vectors, i.e., $\|\boldsymbol{x}\|_\infty=\max_{1\le i\le n}|\boldsymbol{x}_i|$ for any $\boldsymbol{x}\in\mathbb{R}^n$. Besides, we use $\rho(\cdot)$ to denote the spectral radius of a matrix.

2. Review on ALM-based Methods for (BP)

In this section, we first briefly review the standard Augmented Lagrangian Method (ALM) and its linearized version for (BP). Then, we review two versions of ADMM applied to (BP) that had been discussed in the literature.

2.1. The ALM and its linearized version for (BP).

It is well known that the ALM is a benchmark solver for equality constrained optimization problems [15, Ch. 17]. Therefore, we can directly employ the ALM to solve (BP). Letting $\boldsymbol{\lambda}$ be Lagrange multiplier associated to the linear constraint, the augmented Lagrangian function for (BP) reads as

$$\mathscr{L}(\boldsymbol{u},\boldsymbol{\lambda})=\|\boldsymbol{u}\|_1-\boldsymbol{\lambda}^\top(C\boldsymbol{u}-b)+\frac{\beta}{2}\|C\boldsymbol{u}-b\|^2,$$

where $\beta>0$ serves as a penalty parameter. Accordingly, for given $\boldsymbol{\lambda}^k$, the iterative scheme of ALM is

$$\text{(ALM)}\quad\begin{cases}\boldsymbol{u}^{k+1}=\arg\min\limits_{\boldsymbol{u}\in\mathbb{R}^n}\left\{\|\boldsymbol{u}\|_1+\dfrac{\beta}{2}\left\|C\boldsymbol{u}-b-\dfrac{\boldsymbol{\lambda}^k}{\beta}\right\|^2\right\}, & (2.1a)\\ \boldsymbol{\lambda}^{k+1}=\boldsymbol{\lambda}^k-\beta(C\boldsymbol{u}^{k+1}-b). & (2.1b)\end{cases}$$

We can easily observe that the $\boldsymbol{u}$-subproblem (2.1a) has no closed-form solution due to the appearance of matrix C. In this situation, we must employ another optimization solver to find $\boldsymbol{u}^{k+1}$ exactly. However, even if possible, such a procedure usually takes a high computational cost for large-scale problems. More importantly, it is sensitive to the quality of the solution of (2.1a) in accordance with our numerical experience on the datasets in Section 4. Therefore, we consider linearizing the quadratic term in (2.1a) to simplify the $\boldsymbol{u}$-subproblem. Specifically, we approximate the quadratic term in (2.1a) at $\boldsymbol{u}^k$ by

$$\frac{1}{2}\left\|C\boldsymbol{u}-b-\frac{\boldsymbol{\lambda}^k}{\beta}\right\|^2 \approx \frac{1}{2}\left\|C\boldsymbol{u}^k-b-\frac{\boldsymbol{\lambda}^k}{\beta}\right\|^2+\left(\boldsymbol{u}-\boldsymbol{u}^k\right)^\top C^\top\left(C\boldsymbol{u}^k-b-\frac{\boldsymbol{\lambda}^k}{\beta}\right)+\frac{1}{2\gamma}\|\boldsymbol{u}-\boldsymbol{u}^k\|^2,$$

where $\gamma > 0$ is a proximal (or linearization) parameter, which is required to be $\gamma \in \left(0, 1/\rho(C^\top C)\right]$. As a consequence, the $\boldsymbol{u}$-subproblem (2.1a) becomes

$$\boldsymbol{u}^{k+1} = \underset{\boldsymbol{u}\in\mathbb{R}^n}{\arg\min}\left\{\|\boldsymbol{u}\|_1+\frac{\beta}{2\gamma}\left\|\boldsymbol{u}-\left(\boldsymbol{u}^k-\gamma C^\top\left(C\boldsymbol{u}^k-b-\frac{\boldsymbol{\lambda}^k}{\beta}\right)\right)\right\|^2\right\}. \tag{2.2}$$

By invoking the proximal operator of $\|\cdot\|_1$, the new $\boldsymbol{u}$-subproblem (2.2) immediately enjoys a closed-form solution. Therefore, we obtain the linearized ALM (L-ALM) for (BP), which reads as

$$\text{(L-ALM)} \quad \begin{cases} \boldsymbol{u}^{k+1} = \text{shrink}\left(\boldsymbol{u}^k-\gamma C^\top\left(C\boldsymbol{u}^k-b-\frac{\boldsymbol{\lambda}^k}{\beta}\right), \frac{\gamma}{\beta}\right), \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k+\beta(C\boldsymbol{u}^{k+1}-b), \end{cases} \tag{2.3}$$

where shrink$(\cdot,\cdot)$ is the shrinkage operator defined by

$$\text{shrink}(\boldsymbol{x},t) = \text{sign}(\boldsymbol{x})\cdot\max\{|\boldsymbol{x}|-t, 0\},$$

with sign$(\cdot)$ being the sign function and $t > 0$. As shown in [20], the L-ALM (2.3) is globally convergent under mild conditions.

Theorem 2.1. *Let $\boldsymbol{u}^*$ be a solution of (BP), and let $\boldsymbol{\lambda}^*$ be the optimal multiplier. Denote*

$$\boldsymbol{w} = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{\lambda} \end{pmatrix} \quad \text{and} \quad H_L = \begin{pmatrix} \frac{\beta}{\gamma}I-\beta C^\top C & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix}.$$

Suppose that $\gamma \in \left(0, 1/\rho(C^\top C)\right]$ and $\beta > 0$. Then, the sequence $\left\{\boldsymbol{w}^k = (\boldsymbol{u}^k, \boldsymbol{\lambda}^k)\right\}$ generated by the L-ALM (2.3) satisfies

(i). $\|\boldsymbol{w}^{k+1}-\boldsymbol{w}^*\|_{H_L}^2 \le \|\boldsymbol{w}^k-\boldsymbol{w}^*\|_{H_L}^2-\|\boldsymbol{w}^k-\boldsymbol{w}^{k+1}\|_{H_L}^2$;

(ii). $\|\boldsymbol{w}^k-\boldsymbol{w}^{k+1}\|_{H_L}^2 \le \frac{\|\boldsymbol{w}^0-\boldsymbol{w}^*\|_{H_L}^2}{k+1}$, *which means that the L-ALM enjoys an $\mathscr{O}(1/k)$ convergence rate.*

2.2. **ADMM for the primal form of (BP).** As shown in Section 2.1, the direct application of the ALM to (BP) yields a difficult subproblem (2.1a), which cannot efficiently exploit the explicit solutions of the proximal mapping of $\|\cdot\|_1$. Moreover, the L-ALM requires an extra proximal (or linearization) parameters γ, which depends on the spectral of $C^\top C$, resulting in a small step size for updating $\boldsymbol{u}^{k+1}$. Therefore, as suggested in [22], a natural way to address the above issues is to separate the shared variable $\boldsymbol{u}$ from the objective function $\|\boldsymbol{u}\|_1$ and linear constraints $C\boldsymbol{u} = b$ without identity coefficient matrix. Consequently, we can easily employ the well-known ADMM [9, 10] tailored for separable minimization problems to solve the resulting model. Specifically, we introduce an auxiliary variable $\boldsymbol{z}$ to separate $\boldsymbol{u}$ from the objective function of (BP) and let $\boldsymbol{u}=\boldsymbol{z}$. Then, we construct the following separable optimization problem:

$$\min_{\boldsymbol{u},\boldsymbol{z}} \{\|\boldsymbol{u}\|_1 \mid C\boldsymbol{z}=b, \boldsymbol{u}=\boldsymbol{z}\},$$

or in a compact form of

$$\min_{\boldsymbol{u},\boldsymbol{z}} \left\{\|\boldsymbol{u}\|_1 \mid \widehat{A}\boldsymbol{u}+\widehat{B}\boldsymbol{z}=\widehat{b}\right\} \tag{2.4}$$

with

$$\widehat{A}=\begin{pmatrix}\mathbf{0}\\ I\end{pmatrix},\ \widehat{B}=\begin{pmatrix}C\\ -I\end{pmatrix},\ \text{and}\ \widehat{b}=\begin{pmatrix}b\\ 0\end{pmatrix}.$$

Correspondingly, we let

$$\mathscr{L}(\boldsymbol{u},\boldsymbol{z},\boldsymbol{\lambda})=\|\boldsymbol{u}\|_1-\boldsymbol{\lambda}^\top\left(\widehat{A}\boldsymbol{u}+\widehat{B}\boldsymbol{z}-\widehat{b}\right)+\frac{\beta}{2}\left\|\widehat{A}\boldsymbol{u}+\widehat{B}\boldsymbol{z}-\widehat{b}\right\|^2,$$

be the augmented Lagrangian function with positive penalty parameter β and Lagrangian multiplier $\boldsymbol{\lambda}\in\mathbb{R}^{m+n}$. Then, the iterative scheme of ADMM for (2.4) reads as

$$\begin{cases}\boldsymbol{u}^{k+1}=\arg\min_{\boldsymbol{u}}\mathscr{L}(\boldsymbol{u},\boldsymbol{z}^k,\boldsymbol{\lambda}^k), & (2.5a)\\ \boldsymbol{z}^{k+1}=\arg\min_{\boldsymbol{z}}\mathscr{L}(\boldsymbol{u}^{k+1},\boldsymbol{z},\boldsymbol{\lambda}^k), & (2.5b)\\ \boldsymbol{\lambda}^{k+1}=\boldsymbol{\lambda}^k-\beta\left(\widehat{A}\boldsymbol{u}^{k+1}+\widehat{B}\boldsymbol{z}^{k+1}-\widehat{b}\right). & (2.5c)\end{cases}$$

By exploiting the block structure of the linear constraint in (2.4), the Lagrangian multiplier $\boldsymbol{\lambda}$ can also be decomposed into two blocks $\boldsymbol{\lambda}_1\in\mathbb{R}^m$ and $\boldsymbol{\lambda}_2\in\mathbb{R}^n$, where the former corresponds to $C\boldsymbol{z}=b$ and the latter is associated with $\boldsymbol{u}=\boldsymbol{z}$. Using the shrinkage operator and invoking the first-order optimality of (2.5b), the iterative scheme (2.5) can be specified as

$$\text{(P-ADMM)}\begin{cases}\boldsymbol{u}^{k+1}=\operatorname{shrink}\left(\boldsymbol{z}^k+\frac{1}{\beta}\boldsymbol{\lambda}_2^k,\frac{1}{\beta}\right),\\ \boldsymbol{z}^{k+1}=\left(\beta C^\top C+\beta I\right)^{-1}\left(C^\top\left(\beta b+\boldsymbol{\lambda}_1^k\right)+\beta\boldsymbol{u}^{k+1}-\boldsymbol{\lambda}_2^k\right),\\ \boldsymbol{\lambda}_1^{k+1}=\boldsymbol{\lambda}_1^k-\beta\left(C\boldsymbol{z}^{k+1}-b\right),\\ \boldsymbol{\lambda}_2^{k+1}=\boldsymbol{\lambda}_2^k-\beta\left(\boldsymbol{u}^{k+1}-\boldsymbol{z}^{k+1}\right).\end{cases} \tag{2.6}$$

Comparing with ALM (2.1), the application of ADMM to (2.4) yields a more implementable scheme in the sense that all subproblems have closed-form solutions. Note that we usually employ the Cholesky decomposition or Woodbury matrix identity for computing the inverse of matrix $(\beta C^\top C+\beta I)$ in practice. It follows from [13] that the P-ADMM) (2.6) is globally convergent.

Theorem 2.2. *Let $(\boldsymbol{u}^*,\boldsymbol{z}^*)$ be a solution of (2.4), and let $\boldsymbol{\lambda}^*$ be the optimal multiplier. Denote*

$$\boldsymbol{v}=\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{z}\\ \boldsymbol{\lambda}\end{pmatrix}\quad and\quad H_P=\begin{pmatrix}0&0&0\\ 0&\beta\widehat{B}^\top\widehat{B}&0\\ 0&0&\frac{1}{\beta}I\end{pmatrix}. \tag{2.7}$$

For given $\beta>0$, the sequence $\left\{\boldsymbol{v}^k=(\boldsymbol{u}^k,\boldsymbol{z}^k,\boldsymbol{\lambda}^k)\right\}$ generated by P-ADMM (2.6) has the following properties (i). *The sequence* $\{\boldsymbol{v}^k\}$ *is strictly contractive, i.e.,*

$$\|\boldsymbol{v}^{k+1}-\boldsymbol{v}^*\|_{H_P}^2 \le \|\boldsymbol{v}^k-\boldsymbol{v}^*\|_{H_P}^2 - \|\boldsymbol{v}^k-\boldsymbol{v}^{k+1}\|_{H_P}^2.$$

(ii). *The sequence* $\{\|\boldsymbol{v}^k-\boldsymbol{v}^{k+1}\|_{H_P}^2\}$ *is monotonically nonincreasing, i.e.,*

$$\|\boldsymbol{v}^k-\boldsymbol{v}^{k+1}\|_{H_P}^2 \le \|\boldsymbol{v}^{k-1}-\boldsymbol{v}^k\|_{H_P}^2 \quad \text{and} \quad \|\boldsymbol{v}^k-\boldsymbol{v}^{k+1}\|_{H_P}^2 \le \frac{\|\boldsymbol{v}^0-\boldsymbol{v}^*\|_{H_P}^2}{k+1},$$

which implies that P-ADMM (2.6) *also enjoys an* $\mathscr{O}(1/k)$ *convergence rate.*

2.3. **ADMM for the dual form of** (BP)**.** As shown in 2.2, applying the ADMM to primal form of the constructed separable model (2.4) leads to an implementable iterative scheme. Usually, applying some solvers to the dual form of some convex optimization problems possibly produces more efficient algorithms, e.g., the ADMM is essential a by-product of the Douglas-Rachford splitting method applying to the dual form of a convex composite optimization problem [6]. Therefore, a natural question is that can we apply ADMM to the dual form of (BP)? In this part, we review an efficient version developed in [21] that has been widely used in many fields.

First, it is trivial to obtain the dual problem of (BP) as follows:

$$\max_{\boldsymbol{\lambda}} \left\{ b^\top \boldsymbol{\lambda} \mid C^\top \boldsymbol{\lambda} \in \mathbb{B}_\infty := \{\boldsymbol{z} \in \mathbb{R}^n \mid \|\boldsymbol{z}\|_\infty \le 1\} \right\}, \tag{2.8}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^m$ is called the dual variable. Note that the appearance $C^\top$ in the constraint $C^\top\boldsymbol{\lambda} \in \mathbb{B}_\infty$ of (2.8) makes it a little complex. Therefore, we introduce an auxiliary variable $\boldsymbol{z}$ to extract $C^\top\boldsymbol{\lambda}$ from $\mathbb{B}_\infty$ and construct $\boldsymbol{z} = C^\top\boldsymbol{\lambda}$. Then, we immediately obtain a separable optimization problem, which reads as

$$\min_{\boldsymbol{\lambda},\boldsymbol{z}} \left\{ -b^\top\boldsymbol{\lambda} \mid \boldsymbol{z} - C^\top\boldsymbol{\lambda} = 0,\ \boldsymbol{z} \in \mathbb{B}_\infty \right\}. \tag{2.9}$$

For notational simplicity, we let $u \in \mathbb{R}^n$ be the Lagrangian multipliers (which actually corresponds to the primal variable of (BP)) of (2.9). Then, its augmented Lagrangian function reads as

$$\mathscr{L}(\boldsymbol{z},\boldsymbol{\lambda},\boldsymbol{u}) = -b^\top\boldsymbol{\lambda} - \boldsymbol{u}^\top(\boldsymbol{z}-C^\top\boldsymbol{\lambda}) + \frac{\beta}{2}\|\boldsymbol{z}-C^\top\boldsymbol{\lambda}\|^2.$$

As a consequence, the application of ADMM to (2.9) yields

$$\begin{cases} \boldsymbol{z}^{k+1} = \arg\min_{\boldsymbol{z}\in\mathbb{B}_\infty} \mathscr{L}(\boldsymbol{z},\boldsymbol{\lambda}^k,\boldsymbol{u}^k), & \text{(2.10a)} \\ \boldsymbol{\lambda}^{k+1} = \arg\min_{\boldsymbol{\lambda}\in\mathbb{R}^m} \mathscr{L}(\boldsymbol{z}^{k+1},\boldsymbol{\lambda},\boldsymbol{u}^k), & \text{(2.10b)} \\ \boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \beta(\boldsymbol{z}^{k+1} - C^\top\boldsymbol{\lambda}^{k+1}). & \text{(2.10c)} \end{cases}$$

Notice that the $\boldsymbol{\lambda}$-subproblem (2.10b) amounts to finding a solution of the following linear system:

$$\beta C(C^\top\boldsymbol{\lambda} - \boldsymbol{z}^{k+1}) + C\boldsymbol{u}^k - b = 0. \tag{2.11}$$

To reduce the computational cost of solving (2.11), Yang and Zhang [21] judiciously used one-step steepest descent method to find an approximate solution of (2.10b). By using the closed-form solution of (2.10a), a simplified ADMM for dual problem (2.9) reads as

$$\text{(D-ADMM)} \begin{cases} \boldsymbol{z}^{k+1} = \Pi_{\mathbb{B}_\infty}\left(C^\top \boldsymbol{\lambda}^k + \frac{1}{\beta}\boldsymbol{u}^k\right), \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \alpha_k g^k, \\ \boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \beta(\boldsymbol{z}^{k+1} - C^\top \boldsymbol{\lambda}^{k+1}), \end{cases} \tag{2.12}$$

where $\Pi_{\mathbb{B}_\infty}(\cdot)$ represents the projection onto $\mathbb{B}_\infty$, and α_k and g^k are respectively given by

$$\alpha_k = \frac{(g^k)^\top g^k}{(g^k)^\top(\beta CC^\top)g^k} \quad \text{and} \quad g^k = C\boldsymbol{u}^k - b + \beta C(C^\top \boldsymbol{\lambda}^k - \boldsymbol{z}^{k+1}).$$

Here we notice that the global convergence of D-ADMM (2.12) was not established in [21]. However, it performs better than P-ADMM (2.6) in practice, which will also be verified in Section 4.

3. Refined ADMM and Its Linearized Version for (BP)

We can see from the discussions in Sections 2.2 and 2.3 that applying ADMM to the constructed separable formulation of (BP) produces implementable iterative schemes, which is able to efficiently leverage the promising property of ℓ_1-norm minimization models. However, some extra variables must be introduced to construct the separable structure, which will increase the storage burden and possibly decrease the computational efficiency. In this section, we first introduce a natural separable reformulation for (BP) without increasing the problem scale. Then, we give a refined application of ADMM and its fully linearized version to the resulting separable model.

3.1. Refined ADMM for (3.2).

Revisiting the definition of ℓ_1-norm and the linear constraints, it is not difficult to observe that both terms have a natural separable structure. Specifically, we divide the variable $\boldsymbol{u}$ and coefficient matrix C into two blocks, i.e.,

$$\boldsymbol{u} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \quad \text{and} \quad C = (A, B), \tag{3.1}$$

where $\boldsymbol{x} \in \mathbb{R}^{n_1}$, $\boldsymbol{y} \in \mathbb{R}^{n_2}$, $A \in \mathbb{R}^{m\times n_1}$, $B \in \mathbb{R}^{m\times n_2}$, and $n_1 + n_2 = n$. Then, original problem (BP) immediately becomes

$$\min_{\boldsymbol{x}\in\mathbb{R}^{n_1}, \boldsymbol{y}\in\mathbb{R}^{n_2}} \{\|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1 \mid A\boldsymbol{x} + B\boldsymbol{y} = b\}, \tag{3.2}$$

which obviously appears a natural separable structure without any auxiliary variable. A promising result is that we do not require extra space to store auxiliary variables. Let

$$\mathscr{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) = \|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1 - \boldsymbol{\lambda}^\top(A\boldsymbol{x} + B\boldsymbol{y} - b) + \frac{\beta}{2}\|A\boldsymbol{x} + B\boldsymbol{y} - b\|^2$$

be the augmented Lagrangian function associated with (3.2). Consequently, applying ADMM to (3.2) immediately leads to the Refined ADMM (R-ADMM) for (BP):

$$
\text{(R-ADMM)}\quad
\begin{cases}
\boldsymbol{x}^{k+1} = \arg\min_{\boldsymbol{x}} \left\{ \|\boldsymbol{x}\|_1 + \frac{\beta}{2}\left\| A\boldsymbol{x} + B\boldsymbol{y}^k - b - \frac{\boldsymbol{\lambda}^k}{\beta}\right\|^2 \right\}, & \text{(3.3a)}\\
\boldsymbol{y}^{k+1} = \arg\min_{\boldsymbol{y}} \left\{ \|\boldsymbol{y}\|_1 + \frac{\beta}{2}\left\| A\boldsymbol{x}^{k+1} + B\boldsymbol{y} - b - \frac{\boldsymbol{\lambda}^k}{\beta}\right\|^2 \right\}, & \text{(3.3b)}\\
\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta(A\boldsymbol{x}^{k+1} + B\boldsymbol{y}^{k+1} - b). & \text{(3.3c)}
\end{cases}
$$

Theoretically, it follows from [13] that the R-ADMM (3.3) has the similar convergence property as shown in Theorem 2.2. For a better understanding, we state it completely by the following theorem.

Theorem 3.1. *Let $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ be a solution of (3.2), and let $\boldsymbol{\lambda}^*$ be the optimal multiplier. Denote*

$$
\boldsymbol{w} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{\lambda} \end{pmatrix} \quad \text{and} \quad H_R = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta B^\top B & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}. \tag{3.4}
$$

For given $\beta > 0$, the sequence $\left\{\boldsymbol{w}^k = (\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{\lambda}^k)\right\}$ generated by R-ADMM (3.3) has the following properties

(i). *The sequence $\left\{\boldsymbol{w}^k\right\}$ is strictly contractive, i.e.,*

$$
\|\boldsymbol{w}^{k+1} - \boldsymbol{w}^*\|_{H_R}^2 \le \|\boldsymbol{w}^k - \boldsymbol{w}^*\|_{H_R}^2 - \|\boldsymbol{w}^k - \boldsymbol{w}^{k+1}\|_{H_R}^2.
$$

(ii). *The sequence $\left\{\|\boldsymbol{w}^k - \boldsymbol{w}^{k+1}\|_{H_R}^2\right\}$ is monotonically nonincreasing, i.e.,*

$$
\|\boldsymbol{w}^k - \boldsymbol{w}^{k+1}\|_{H_R}^2 \le \|\boldsymbol{w}^{k-1} - \boldsymbol{w}^k\|_{H_R}^2 \quad \text{and} \quad \|\boldsymbol{w}^k - \boldsymbol{w}^{k+1}\|_{H_R}^2 \le \frac{\|\boldsymbol{w}^0 - \boldsymbol{w}^*\|_{H_R}^2}{k+1},
$$

which implies that R-ADMM (3.3) also enjoys an $\mathscr{O}(1/k)$ convergence rate.

Remark 3.1. It is easy to observe that P-ADMM (2.5), and R-ADMM (3.3) share the same $\mathscr{O}(1/k)$ convergence rate, but with different upper bounds due to their matrices H_P and H_R given in (2.7) and (3.4), respectively. We can verify that the spectral radius of H_R is smaller than the one of H_P, which implies that R-ADMM (3.3) has a tighter upper bound than P-ADMM (2.5). To some extent, it possibly explains why R-ADMM (3.3) takes much fewer iterations than P-ADMM (2.5) (see results in Section 4).

3.2. **Fully linearized R-ADMM for (3.2).** It is clear that R-ADMM (3.3) has no closed-form solutions for its $\boldsymbol{x}$- and $\boldsymbol{y}$-subproblems. Therefore, we employ the approximation strategy used in L-ALM (2.3) to simplify R-ADMM (3.3). More concretely, we first linearize the quadratic term of (3.3a) and arrive at

$$
\boldsymbol{x}^{k+1} = \arg\min_{\boldsymbol{x}} \left\{ \|\boldsymbol{x}\|_1 + \frac{\beta}{2\gamma_1}\left\| \boldsymbol{x} - \left(\boldsymbol{x}^k - \gamma_1 A^\top \left(A\boldsymbol{x}^k + B\boldsymbol{y}^k - b - \frac{\boldsymbol{\lambda}^k}{\beta} \right)\right)\right\|^2 \right\}
$$

$$= \text{shrink}\left(\boldsymbol{x}^k - \gamma_1 A^\top \left(A\boldsymbol{x}^k + B\boldsymbol{y}^k - b - \frac{\boldsymbol{\lambda}^k}{\beta}\right), \frac{\gamma_1}{\beta}\right), \tag{3.5}$$

where $\gamma_1 \in \left(0, 1/\rho(A^\top A)\right]$ is a proximal (or linearization) parameter to ensure convergence. Unlike the approximation on the $\boldsymbol{x}$-subproblem, we follow the idea introduced in [11] to linearize $\boldsymbol{y}$-subproblem (3.3b) and obtain

$$\boldsymbol{y}^{k+1} = \arg\min_{\boldsymbol{y}} \left\{ \|\boldsymbol{y}\|_1 + \frac{\beta}{2\tau\gamma_2} \left\| \boldsymbol{y} - \left(\boldsymbol{y}^k - \tau\gamma_2 B^\top \left(A\boldsymbol{x}^{k+1} + B\boldsymbol{y}^k - b - \frac{\boldsymbol{\lambda}^k}{\beta}\right)\right) \right\|^2 \right\}$$
$$= \text{shrink}\left(\boldsymbol{y}^k - \tau\gamma_2 B^\top \left(A\boldsymbol{x}^{k+1} + B\boldsymbol{y}^k - b - \frac{\boldsymbol{\lambda}^k}{\beta}\right), \frac{\tau\gamma_2}{\beta}\right), \tag{3.6}$$

with convergence-guaranteeing parameters $\gamma_2 \in \left(0, 1/\rho(B^\top B)\right]$ and $\tau \in (1, 4/3)$, where τ serves the role of enlarging the step size for updating $\boldsymbol{y}^{k+1}$ to compensate for the approximation. As a consequence, we obtain a fully linearized R-ADMM (LR-ADMM in short) for (3.2) by Algorithm 1.

Algorithm 1 Fully Linearized R-ADMM for (3.2).

1: Select $\beta > 0$, $\gamma_1 \in \left(0, 1/\rho(A^\top A)\right]$, $\gamma_2 \in \left(0, 1/\rho(B^\top B)\right]$ and $\tau \in (1, 4/3)$.
2: **for** $k = 0, 1, 2, \cdots$ **do**
3: Update $\boldsymbol{x}^{k+1}$ via (3.5).
4: Update $\boldsymbol{y}^{k+1}$ via (3.6).
5: Update $\boldsymbol{\lambda}^{k+1}$ via (3.3c).
6: **end for**

Below, with the notation $\boldsymbol{u}$ and $\boldsymbol{w}$ given in (3.1) and (3.4), respectively, we further introduce the following notations to state the convergence results of Algorithm 1.

$$F(\boldsymbol{w}) = \begin{pmatrix} -A^\top \boldsymbol{\lambda} \\ -B^\top \boldsymbol{\lambda} \\ A\boldsymbol{x} + B\boldsymbol{y} - b \end{pmatrix} \quad \text{and} \quad H_{LR} = \begin{pmatrix} \frac{\beta}{\gamma_1} I - \beta A^\top A & 0 & 0 \\ 0 & \frac{\beta}{\tau\gamma_2} I & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}. \tag{3.7}$$

Theorem 3.2. *Let $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ be a solution of (3.2), and let $\boldsymbol{\lambda}^*$ be the optimal multiplier. Suppose that $\gamma_1 \in \left(0, 1/\rho(A^\top A)\right]$, $\gamma_2 \in \left(0, 1/\rho(B^\top B)\right]$, and $\tau \in (1, 4/3)$. For any $\beta > 0$, the sequence $\{\boldsymbol{w}^k = (\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{\lambda}^k)\}$ generated by Algorithm 1 satisfies*

$$\theta(\widehat{\boldsymbol{u}}_t) - \theta(\boldsymbol{u}) + (\widehat{\boldsymbol{w}}_t - \boldsymbol{w})^\top F(\boldsymbol{w})$$
$$\leq \frac{1}{2t}\left(\|\boldsymbol{w} - \boldsymbol{w}^1\|_{H_{LR}}^2 + \frac{1}{2}\left(\tau \|\boldsymbol{y}^0 - \boldsymbol{y}^1\|^2 + (1-\tau)\beta \|B(\boldsymbol{y}^0 - \boldsymbol{y}^1)\|^2\right)\right),$$

where $\theta(\boldsymbol{u}) \equiv \theta(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1$, $\widehat{\boldsymbol{u}}_t = \frac{1}{t}\sum_{k=1}^{t} \widehat{\boldsymbol{u}}^k$, $\widehat{\boldsymbol{w}}_t = \frac{1}{t}\sum_{k=1}^{t} \widehat{\boldsymbol{w}}^k$ *with*

$$\widehat{\boldsymbol{u}}^k = \begin{pmatrix} \boldsymbol{x}^{k+1} \\ \boldsymbol{y}^{k+1} \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{w}}^k = \begin{pmatrix} \boldsymbol{x}^{k+1} \\ \boldsymbol{y}^{k+1} \\ \boldsymbol{\lambda}^k - \beta(A\boldsymbol{x}^{k+1} + B\boldsymbol{y}^k - b) \end{pmatrix}.$$

The result implies that Algorithm 1 has an $\mathscr{O}(1/t)$ *convergence rate.*

Proof. See the proof of Theorem A.3. □

4. Numerical Experiments

In this section, we conduct the numerical performance of the aforementioned five ALM-based first-order algorithms on solving (BP) with synthetic datasets. Here, we aim to show that some structured optimization problems would be solved efficiently if we could fully exploit its inherent favorable structure.

Note that ALM (2.1) and R-ADMM (3.3) have subproblems that should be solved by the employment of some optimization solvers. Here, we employ the famous FISTA [1] to solve them and take the $\boldsymbol{u}$-subproblem (2.1a) as an example to elaborate the implementation details. Specifically, applying FISTA to (2.1a) yields the following iterative scheme:

$$\begin{cases} \bar{\boldsymbol{z}}^j = \boldsymbol{u}^{k,j} + \dfrac{t_j - 1}{t_{j+1}}\left(\boldsymbol{u}^{k,j} - \boldsymbol{u}^{k,j-1}\right), \\ \boldsymbol{u}^{k,j+1} = \operatorname{shrink}\left(\bar{\boldsymbol{z}}^j - \dfrac{\beta}{\zeta} C^\top \left(C\bar{\boldsymbol{z}}^j - b - \dfrac{\boldsymbol{\lambda}^k}{\beta}\right), \dfrac{1}{\zeta}\right), \end{cases}$$

where ζ is a constant being larger than the Lipschitz constant of the quadratic term in (2.1a) and $t_{j+1} = \left(1 + \sqrt{1 + 4t_j^2}\right)/2$ starts from $t_1 = 1$, and k, and j are the outer and inner iteration counters, respectively. Moreover, the FISTA will be terminated when $\|\boldsymbol{u}^{k,j+1} - \boldsymbol{u}^{k,j}\| \le 10^{-3}/k^2$ or the number of inner iterations exceed a preset number 10. Moreover, it is noticed that the penalty parameter β is playing a great role in these five ALM-based algorithms. As shown in [21], a dynamical β is of benefit for better numerical behaviors. Therefore, we show their convergence results with a fixed constant for simplicity, while an increasing strategy (i.e., $\beta_k = \min\{10\beta_{k-1}, \beta_{\max}\}$ such that $0 < \beta_0 \le \beta_1 \le \cdots \le \beta_k \le \cdots \le \beta_{\max}$) will be employed in algorithmic implementation for all of them. Certainly, we also show that such a dynamical strategy still keeps the global convergence for the proposed LR-ADMM (3.3) (see our proof in Appendix A).

All algorithms are implemented by MATLAB R2021a, and all experiments are conducted on a Lenovo laptop with Windows 11 system and Intel(R) Core(TM) i5-12500 CPU processor with 16GB memory.

4.1. Sensitivity simulation on the block size of (3.2).

As shown in Section 3, the variable $\boldsymbol{u} \in \mathbb{R}^n$ of (BP) is decomposed into two blocks $\boldsymbol{x} \in \mathbb{R}^{n_1}$ and $\boldsymbol{y} \in \mathbb{R}^{n_2}$. Hence, we are here concerned with the numerical sensitivity on the size of the two decomposed vectors. In other words, whether the ratio of n_1 and n_2 (i.e., $n_1 : n_2$) will affect the numerical performance of R-ADMM (3.3) and LR-ADMM (i.e., Algorithm 1)?

In the experiment, we first state the data construction for (BP). Concretely, we generate C in a random way, which is a standard Gaussian matrix. Then, we form a random sparse solution $\boldsymbol{u}^\star \in \mathbb{R}^n$, which has s nonzero components. Finally, we let $b = C\boldsymbol{u}^\star$ be the observed vector. The corresponding MATLAB scripts are:

```
        C = randn(m,n);   C = C/norm(C);            p = randi(n,s,1);
   u_star = zeros(n,1);  us = randn(s,1);  u_star(p) = us;  b = C * u_star;
```

To implement our R-ADMM and LR-ADMM, we set $\beta_0 = \|b\|_1/n$ and $\beta_{\max} = 10^8$ for both algorithms. Then, we set $\gamma_1 = 1/\rho(A^\top A)$, $\gamma_2 = 1/\rho(B^\top B)$ and $\tau = 4/3$ for LR-ADMM. Moreover, both algorithms start from the same random initial points $\boldsymbol{u}^0 = (\boldsymbol{x}^0, \boldsymbol{y}^0)$ and $\boldsymbol{\lambda}^0$ that all of them are normally distributed random numbers (by MATLAB script '`randn(n,1)`'), and terminate at

$$\frac{\|\boldsymbol{u}^k - \boldsymbol{u}^\star\|}{\|\boldsymbol{u}^\star\|} \le 10^{-8}.$$

We consider three cases on the size of the problem, i.e., $(m,n,s) = (512i, 1024i, 80i)$ with $i = 4,5,6$. For the decomposition of $\boldsymbol{u}$ and C, we consider $\boldsymbol{u} = (\boldsymbol{x}^\top, \boldsymbol{y}^\top)^\top$ with $\boldsymbol{x} \in \mathbb{R}^{n_1}$ and $\boldsymbol{y} \in \mathbb{R}^{n_2}$, and $C = (A, B)$ with $A \in \mathbb{R}^{m\times n_1}$ and $B \in \mathbb{R}^{m\times n_2}$, where $n_1 : n_2$ will be conducted through three scenarios, i.e., $n_1 : n_2 = \{2:1,\ 1:1,\ 1:2\}$. Here, we should notice that R-ADMM (3.3) is quite sensitive to the block sizes of $\boldsymbol{x}$ and $\boldsymbol{y}$ when both of them are very unbalanced. Therefore, we do not report these cases where $n_1 : n_2 = 1:3$ and $3:1$. Considering the randomness of the data, we conduct the average performance of R-ADMM and LR-ADMM on 10 groups of randomly generated datasets for each case.

In Figure 1, we report the average iterations and computing time in seconds by bar charts, where the line segments particularly represents the standard deviation of 10 trials. Clearly, a longer line segment means that the algorithm performs less stably.

It is easy to see that R-ADMM runs faster for the case where the variable $\boldsymbol{u}$ is equally divided into two blocks (i.e., $n_1 : n_2 = 1:1$) than the other unbalanced cases. Comparatively, LR-ADMM works much more stably than R-ADMM, since the former is insensitive to the block sizes of $\boldsymbol{x}$ and $\boldsymbol{y}$. The results in Figure 1 tell us that the variable $\boldsymbol{u}$ should be decomposed into two blocks with approximately equal size (i.e., $n_1 : n_2 \approx 1:1$).

4.2. **Numerical comparisons.** The above experiments tell us how to decompose the original variable $\boldsymbol{u}$ in (BP). In this section, we will aim to show that the BP can be solved in a faster way than some existing ADMM-type solvers when its inherent separable structure could be exploited efficiently. We compare R-ADMM (3.3) and LR-ADMM (Algorithm 1) with other methods including ALM (2.1), L-ALM (2.3), P-ADMM (2.5), and D-ADMM (2.10).

In the following experiments, we generate the datasets by the same way described in Section 4.1, where we set $n_1 = n/2$ and $n_2 = n - n_1$ for the lengths of $\boldsymbol{x} \in \mathbb{R}^{n_1}$ and $\boldsymbol{y} \in \mathbb{R}^{n_2}$, respectively. Note that the starting points often affect the numerical behaviors of numerical algorithms. Therefore, we first conduct the numerical sensitivity of starting points to all compared algorithms. Here, we consider two ways to start the algorithms: (i) the first way is to set $\boldsymbol{x}^0$, $\boldsymbol{y}^0$, and $\boldsymbol{\lambda}^0$ as zeros; (ii) the second way is starting iteration from random $\boldsymbol{x}^0$, $\boldsymbol{y}^0$, and $\boldsymbol{\lambda}^0$, which are normally distributed. All algorithms will terminate when

$$\max\left\{\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^k\|,\ \|C\boldsymbol{u}^{k+1} - b\|\right\} \le \varepsilon,$$

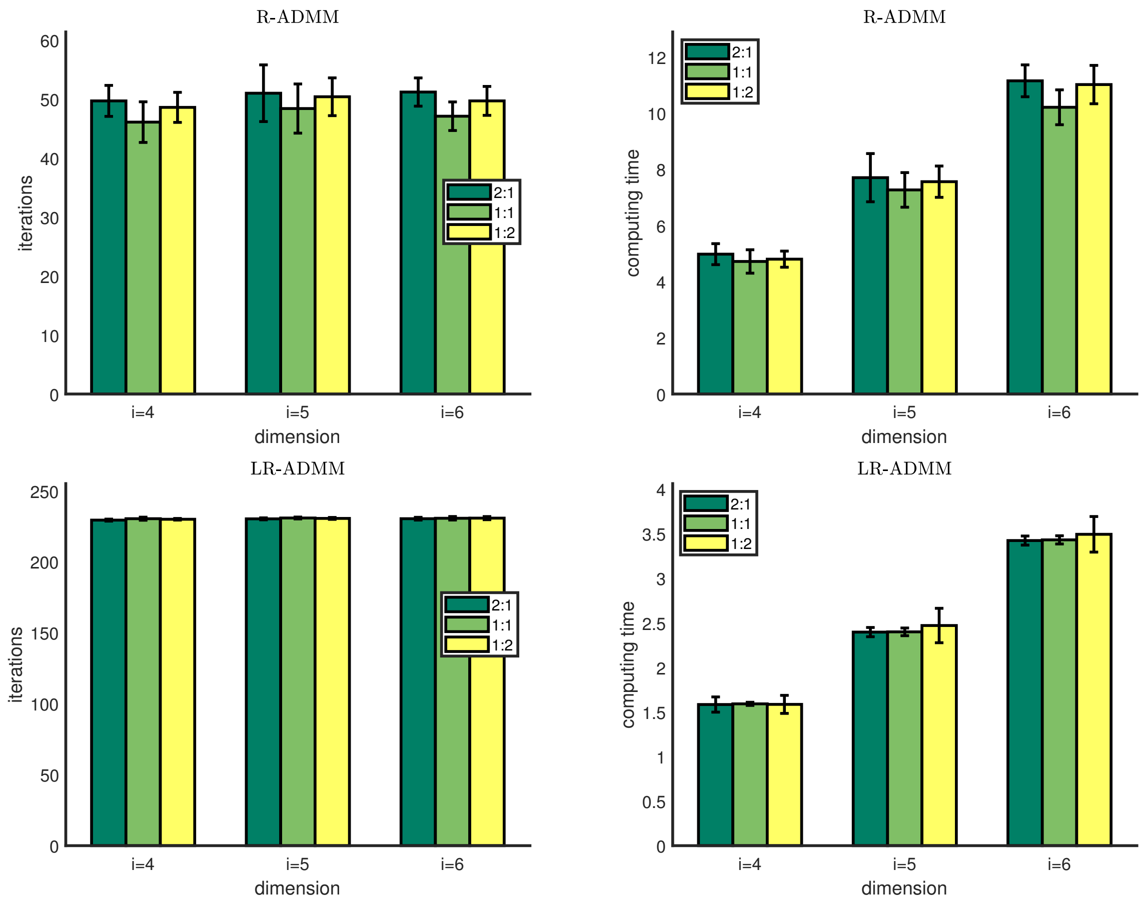

FIGURE 1. Numerical sensitivity of the block sizes of $\boldsymbol{x}$ and $\boldsymbol{y}$ in (3.2) to R-ADMM and LR-ADMM.

where ε is a preset precision, which will be set as $\varepsilon = \{10^{-8}, 10^{-12}\}$. In our experiments, we set $\beta_0 = \|b\|_1/n$ for all algorithms. Then, we take $\gamma = 1/\rho(C^\top C)$ for L-ALM, and set LR-ADMM with the same settings used in Section 4.1. The code of D-ADMM (2.10) comes from the original Matlab package YALL1[3] introduced in [21].

Now, we first set the stopping precision as $\varepsilon = 10^{-8}$ and consider the case $(m,n,s) = (2048, 8192, 320)$. In Figure 2, we show the convergence curves with respect to iterations and computing time for all algorithms starting from two different initial points. Clearly, we can see that ALM and R-ADMM requires much less iterations to achieve a little better solution than the other solvers that have closed-form solutions. Comparatively, R-ADMM runs a little faster than ALM in terms of iterations and computing time. Moreover, LR-ADMM takes the least computing time, which supports that the new reformulation (3.2) equipped with the LR-ADMM is reliable way to improve the efficiency of solving (BP).

Considering the randomness of the datasets, we finally consider the problem (BP) with 10 different sizes, i.e., $(m,n,s) = (512i, 1024i, 80i)$ with $i = 1, 2, \cdots, 10$. Moreover, we report the results for the cases where the algorithms start from two different types of initial points and

[3]https://yall1.blogs.rice.edu/

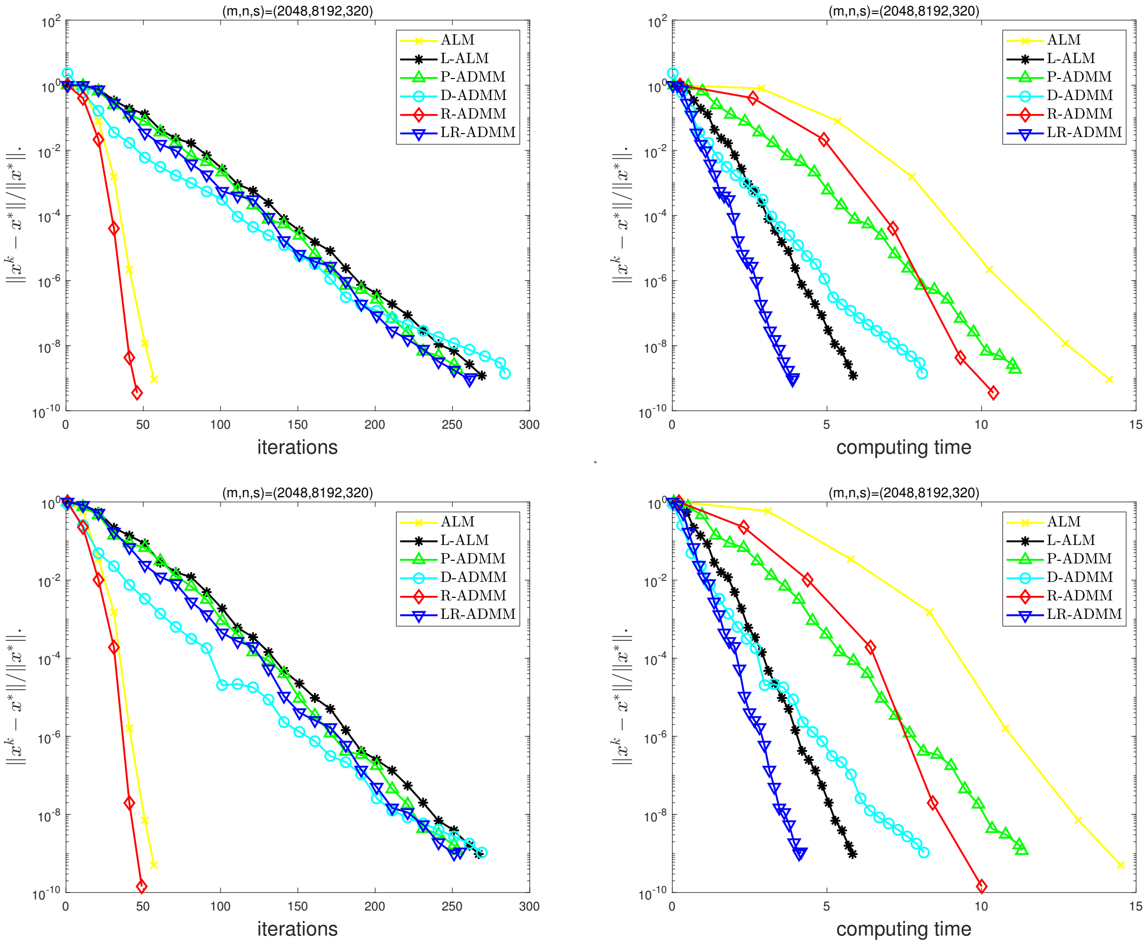

FIGURE 2. Convergence curves with respect to iterations and computing time for numerical sensitivity of the algorithms starting from different initial points. The top two plots correspond to the case where initial points are normally distributed. The bottom two plots correspond to the case where initial points are zeros.

two different stopping precision $\varepsilon = 10^{-8}$ and 10^{-12}, respectively. In Tables 1-4, we report the average iterations (**iter.**) and computing time in seconds (**time**) of 10 random trials.

We can see from the results in Tables 1-4 that R-ADMM always takes the least iterations. When comparing with P-ADMM and D-ADMM, R-ADMM also takes less computing time than the convergent P-ADMM, and runs a little faster than the D-ADMM without convergence guarantee to achieve higher quality solutions. In particular, we can see that our LR-ADMM always takes the least computing time to reach the same stopping tolerance, even though it sometimes takes more iterations than P-ADMM and D-ADMM. Those results in Tables 1-4 only show the average performance of the six algorithms. So, we would like to see their best and worst behaviors on these random datasets, which to some extent illustrate their stability. In Figure 3, we use radar plots to report the best results (i.e., **iter-min** and **time-min**), the average performance (i.e., **iter-aver** and **time-aver**), and the worst values (i.e., **iter-max** and **time-max**)

TABLE 1. Numerical results of the six algorithms starting random points under precision $\varepsilon = 10^{-8}$.

i	ALM	L-ALM	P-ADMM	D-ADMM	R-ADMM	LR-ADMM
	iter. / time	iter. / time	iter. / time	iter. / time	iter. / time	iter. / time
1	38.3 / 0.13	255.3 / 0.08	252.4 / 0.94	258.8 / 0.13	32.8 / 0.12	258.7 / 0.06
2	39.5 / 3.65	265.3 / 2.44	258.2 / 5.16	269.7 / 3.14	33.6 / 1.44	267.3 / 1.24
3	39.6 / 7.05	267.2 / 4.44	261.9 / 9.84	281.0 / 5.11	33.0 / 5.08	272.3 / 2.88
4	40.2 / 11.08	269.6 / 6.63	263.0 / 17.43	283.5 / 8.10	34.3 / 8.44	275.1 / 4.61
5	40.2 / 15.55	271.9 / 9.71	262.9 / 24.53	278.0 / 10.52	34.2 / 11.96	281.0 / 7.03
6	40.5 / 19.76	273.0 / 13.35	263.0 / 35.06	288.5 / 14.63	34.5 / 16.84	282.3 / 10.07
7	40.9 / 25.65	280.2 / 16.11	263.5 / 46.53	279.6 / 18.60	34.2 / 21.54	284.7 / 12.31
8	43.7 / 32.98	280.4 / 20.61	263.6 / 62.02	295.1 / 25.21	35.2 / 27.85	286.1 / 15.87
9	41.0 / 38.16	281.7 / 24.67	263.8 / 75.67	290.3 / 31.19	34.7 / 32.23	289.4 / 18.21
10	41.0 / 44.67	285.0 / 29.20	264.0 / 93.79	296.1 / 38.28	35.2 / 38.52	290.4 / 21.12

TABLE 2. Numerical results of the six algorithms starting random points under precision $\varepsilon = 10^{-12}$.

i	ALM	L-ALM	P-ADMM	D-ADMM	R-ADMM	LR-ADMM
	iter. / time	iter. / time	iter. / time	iter. / time	iter. / time	iter. / time
1	249.0 / 0.25	574.7 / 0.19	356.7 / 1.32	444.5 / 0.22	82.8 / 0.16	482.2 / 0.11
2	240.5 / 6.31	600.2 / 5.01	359.5 / 6.61	442.5 / 4.53	97.7 / 2.07	493.4 / 2.09
3	222.6 / 11.69	609.0 / 9.25	364.7 / 13.03	459.7 / 7.28	49.6 / 5.99	505.6 / 5.04
4	205.4 / 16.82	625.9 / 13.63	366.8 / 23.12	459.8 / 11.96	57.5 / 9.29	511.6 / 8.01
5	190.4 / 25.03	629.6 / 19.48	367.0 / 32.98	462.6 / 16.27	57.3 / 14.62	518.6 / 12.06
6	186.5 / 31.55	632.6 / 26.46	367.1 / 47.73	471.2 / 22.95	57.3 / 19.52	519.2 / 16.54
7	176.4 / 41.00	641.1 / 34.92	367.3 / 64.13	480.3 / 31.25	59.6 / 25.69	522.5 / 21.65
8	171.3 / 50.30	641.4 / 43.74	367.7 / 85.30	476.4 / 39.73	59.0 / 33.08	523.8 / 26.88
9	166.8 / 61.24	644.0 / 52.96	368.0 / 103.67	466.7 / 49.30	63.5 / 39.63	527.9 / 31.72
10	159.5 / 72.78	646.4 / 63.74	368.0 / 128.00	484.8 / 61.57	66.4 / 48.20	526.2 / 37.94

for the case $(m,n,s) = (2048, 8192, 320)$, where all algorithms start from two different initial points.

We can see from Figure 3 that LR-ADMM is relatively more stable than the others. Those results in Tables 1-4 and Figure 3 further validate that the new reformulation (3.2) exploiting the inherent separable structure is beneficial for improving the efficiency of solving (BP). More

TABLE 3. Numerical results of the six algorithms starting zero points under precision $\varepsilon = 10^{-8}$.

i	ALM	L-ALM	P-ADMM	D-ADMM	R-ADMM	LR-ADMM
	iter. / time	iter. / time	iter. / time	iter. / time	iter. / time	iter. / time
1	37.9 / 0.13	255.3 / 0.09	252.4 / 0.90	248.6 / 0.12	32.4 / 0.11	257.7 / 0.06
2	38.7 / 3.61	264.5 / 2.29	256.6 / 5.10	264.3 / 3.20	34.2 / 1.49	267.6 / 1.23
3	39.2 / 7.13	268.1 / 4.94	261.0 / 9.88	271.4 / 5.07	33.0 / 4.89	270.4 / 2.95
4	40.2 / 10.95	268.1 / 7.15	262.9 / 17.34	276.3 / 8.20	33.3 / 7.95	277.9 / 5.14
5	40.6 / 14.89	273.4 / 9.36	263.0 / 24.12	278.9 / 10.47	34.1 / 11.70	280.4 / 7.05
6	40.7 / 20.27	274.2 / 12.53	263.1 / 34.36	282.6 / 14.37	33.8 / 15.41	281.4 / 10.60
7	40.8 / 26.75	279.4 / 16.66	263.6 / 46.08	283.3 / 19.13	34.5 / 21.10	284.6 / 12.94
8	41.0 / 33.38	281.1 / 21.15	263.2 / 60.98	280.2 / 23.91	34.7 / 26.45	285.4 / 16.02
9	41.0 / 37.91	283.3 / 24.65	263.8 / 74.68	291.0 / 31.27	35.0 / 32.35	287.7 / 17.89
10	41.0 / 44.10	285.6 / 29.30	264.0 / 92.34	288.1 / 36.81	36.5 / 40.13	291.6 / 21.34

TABLE 4. Numerical results of the six algorithms starting zero points under precision $\varepsilon = 10^{-12}$.

i	ALM	L-ALM	P-ADMM	D-ADMM	R-ADMM	LR-ADMM
	iter. / time	iter. / time	iter. / time	iter. / time	iter. / time	iter. / time
1	237.6 / 0.23	575.6 / 0.18	356.2 / 1.27	429.3 / 0.19	75.7 / 0.14	472.5 / 0.11
2	243.9 / 6.39	609.1 / 5.04	360.0 / 6.63	449.0 / 4.44	79.8 / 2.04	495.8 / 2.13
3	216.4 / 11.81	609.6 / 9.22	366.4 / 12.90	445.0 / 7.21	56.1 / 5.65	512.4 / 5.41
4	205.1 / 19.09	621.7 / 14.58	366.8 / 22.93	458.9 / 12.16	75.3 / 10.23	512.7 / 8.71
5	195.1 / 23.91	625.1 / 19.87	367.0 / 32.89	467.2 / 16.70	47.1 / 14.15	517.6 / 13.50
6	187.4 / 31.92	630.1 / 26.33	367.0 / 47.13	465.3 / 22.58	62.5 / 20.10	518.7 / 16.85
7	177.4 / 40.85	635.3 / 34.46	367.1 / 63.23	473.6 / 30.93	59.9 / 26.21	521.6 / 21.62
8	174.9 / 50.53	636.5 / 44.04	367.7 / 84.39	450.7 / 37.96	62.8 / 32.61	524.3 / 26.45
9	166.6 / 61.92	643.3 / 53.27	368.0 / 103.63	466.8 / 49.60	82.9 / 42.75	525.6 / 31.47
10	162.4 / 73.15	647.2 / 64.29	368.0 / 127.93	469.2 / 60.46	79.1 / 50.94	527.9 / 37.51

importantly, it tells us that, when implementing ADMM to solve a real-world problem, exploiting the inherent separable structure is more efficient than the way of introducing extra auxiliary variable to construct a separable structure.

5. CONCLUSION

In this paper, we revisited the solution methods for Basis Pursuit (BP) problems, which can be solved by many efficient optimization solvers. Observing that some state-of-the-art first-order

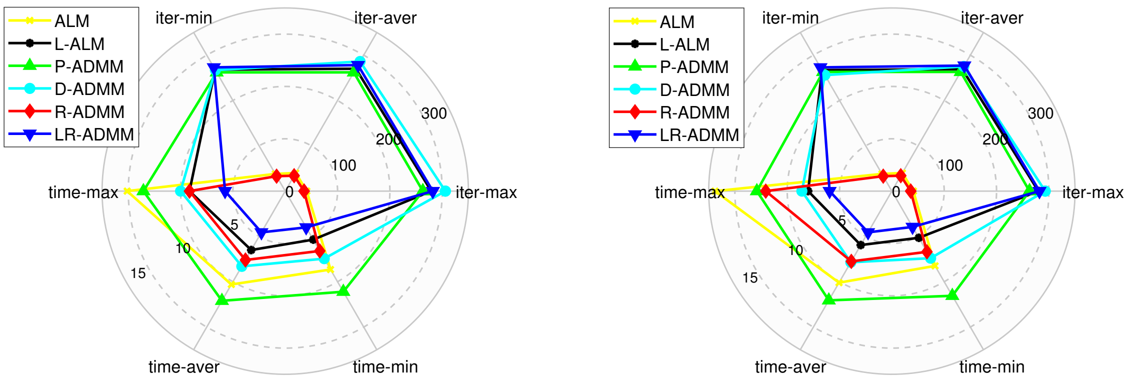

FIGURE 3. Radar plots with respect to iterations and computing time of the algorithms starting from different initial points. The left plot corresponds to the case where initial points are normally distributed random numbers. The right plot corresponds to the case where initial points are zeros.

algorithms, especially ALM-based methods for BP did not efficiently explore the inherent separable structure, we accordingly reformulated BP as a natural separable minimization problem without increasing the size of the problem under consideration. Then, we showed two refined application of ADMM to the resulting problem. One of the algorithms is newly introduced here, which has two simple subproblems with closed-form solutions and enjoys an $\mathscr{O}(1/t)$ convergence rate. A series of computational results shows that BP can be solved faster than some existing first-order algorithms when its inherent separable structure has been efficiently exploited.

APPENDIX A. CONVERGENCE OF ALGORITHM 1

In this appendix, we establish the convergence of the newly introduced fully linearized ADMM (i.e., Algorithm 1) for (BP). Actually, we can extend such an algorithm to handle a more general problem of (3.2), i.e.,

$$\min_{\boldsymbol{x}\in\mathbb{R}^{n_1},\boldsymbol{y}\in\mathbb{R}^{n_2}} \left\{\theta(\boldsymbol{x},\boldsymbol{y}) := \theta_1(\boldsymbol{x})+\theta_2(\boldsymbol{y}) \mid A\boldsymbol{x}+B\boldsymbol{y}=b,\ \boldsymbol{x}\in\mathscr{X},\ \boldsymbol{y}\in\mathscr{Y}\right\}, \tag{A.1}$$

where θ_1 and θ_2 are assumed to be convex functions, and $\mathscr{X}\subseteq\mathbb{R}^{n_1}$ and $\mathscr{Y}\subseteq\mathbb{R}^{n_2}$ are two nonempty convex sets. Here, the solution set of (A.1) is also supposed to be nonempty. Apparently, (A.1) immediately falls into (3.2) when $\theta_1(\boldsymbol{x}) := \|\boldsymbol{x}\|_1$, $\theta_2(\boldsymbol{y}) := \|\boldsymbol{y}\|_1$, $\mathscr{X}=\mathbb{R}^{n_1}$ and $\mathscr{Y}=\mathbb{R}^{n_2}$. Accordingly, the fully linearized ADMM (LR-ADMM in short) for (A.1) reads as

$$\text{(LR-ADMM)}\quad \begin{cases} \boldsymbol{x}^{k+1} = \arg\min\limits_{\boldsymbol{x}\in\mathscr{X}} \left\{\theta_1(\boldsymbol{x}) + \dfrac{\beta_k}{2\gamma_1}\left\|\boldsymbol{x}-\boldsymbol{\xi}_1^k\right\|^2\right\}, & \text{(A.2a)}\\ \boldsymbol{y}^{k+1} = \arg\min\limits_{\boldsymbol{y}\in\mathscr{Y}} \left\{\theta_2(\boldsymbol{y}) + \dfrac{\beta_k}{2\tau\gamma_2}\left\|\boldsymbol{y}-\boldsymbol{\xi}_2^k\right\|^2\right\}, & \text{(A.2b)}\\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta_k(A\boldsymbol{x}^{k+1}+B\boldsymbol{y}^{k+1}-b), & \text{(A.2c)} \end{cases}$$

where

$$\begin{cases} \boldsymbol{\xi}_1^k = \boldsymbol{x}^k - \gamma_1 A^\top \left(A\boldsymbol{x}^k + B\boldsymbol{y}^k - b - \dfrac{\boldsymbol{\lambda}^k}{\beta_k} \right), \\ \boldsymbol{\xi}_2^k = \boldsymbol{y}^k - \tau\gamma_2 B^\top \left(A\boldsymbol{x}^{k+1} + B\boldsymbol{y}^k - b - \dfrac{\boldsymbol{\lambda}^k}{\beta_k} \right). \end{cases}$$

Note that the involved parameters satisfy $0 < \beta_0 \le \beta_1 \le \cdots \le \beta_k \le \cdots \le \beta_{\max}$, $\gamma_1 \in \left(0, 1/\rho(A^\top A)\right]$, $\gamma_2 \in \left(0, 1/\rho(B^\top B)\right]$ and $\tau \in (1, 4/3)$. Below, we consider the convergence of (A.2) for (A.1).

As a preparation for proving the convergence of (A.2), we follow the way introduced in [11] to recall the Variational Inequalities (VI) characterization of (A.1). Specifically, solving (A.1) amounts to finding a triple $\boldsymbol{w}^* = (\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\lambda}^*) \in \mathscr{W}$ such that

$$\begin{cases} \theta_1(\boldsymbol{x}) - \theta_1(\boldsymbol{x}^*) + (\boldsymbol{x} - \boldsymbol{x}^*)^\top (-A^\top \boldsymbol{\lambda}^*) \ge 0, \ \ \forall \boldsymbol{x} \in \mathscr{X}, \\ \theta_2(\boldsymbol{y}) - \theta_2(\boldsymbol{y}^*) + (\boldsymbol{y} - \boldsymbol{y}^*)^\top (-B^\top \boldsymbol{\lambda}^*) \ge 0, \ \ \forall \boldsymbol{y} \in \mathscr{Y}, \\ (\boldsymbol{\lambda} - \boldsymbol{\lambda}^*)^\top (A\boldsymbol{x}^* + B\boldsymbol{y}^* - b) \ge 0, \ \ \forall \boldsymbol{\lambda} \in \mathbb{R}^m, \end{cases}$$

which can be rewritten as a compact form

$$\theta(\boldsymbol{u}) - \theta(\boldsymbol{u}^*) + (\boldsymbol{w} - \boldsymbol{w}^*)^\top F(\boldsymbol{w}^*) \ge 0, \ \ \forall \boldsymbol{w} \in \mathscr{W}, \tag{A.5}$$

where $\boldsymbol{u}$, $\boldsymbol{w}$ and $F(\boldsymbol{w})$ are given in (3.1), (3.4), and (3.7), respectively, and $\mathscr{W} := \mathscr{X} \times \mathscr{Y} \times \mathbb{R}^m$. In what follows, we denote (A.5) by VI($\mathscr{W}, F, \theta$) for simplicity. It is easy to verify that

$$(\boldsymbol{w}' - \boldsymbol{w})^\top (F(\boldsymbol{w}') - F(\boldsymbol{w})) = 0, \ \ \forall \boldsymbol{w}', \boldsymbol{w} \in \mathscr{W}. \tag{A.6}$$

Hence, VI($\mathscr{W}, F, \theta$) is a monotone VI problem. Moreover, under the problem setting, the solution set of VI($\mathscr{W}, F, \theta$), denoted by $\mathscr{W}^*$, is also nonempty and convex.

To facilitate the convergence proof, we can follow the way used in [11] to reformulate the fully linearized ADMM (A.2) as a two-step (prediction-correction) method, which can be stated as the following proposition.

Proposition A.1. *Under the notations defined in* (A.5)*, the iterative scheme LR-ADMM* (A.2) *can be regarded as a prediction-correction method, i.e.,*

(**prediction step**) Find a prediction $\widehat{\boldsymbol{w}}^k \in \mathscr{W}$ such that

$$\theta(\boldsymbol{u}) - \theta(\widehat{\boldsymbol{u}}^k) + (\boldsymbol{w} - \widehat{\boldsymbol{w}}^k)^\top F(\widehat{\boldsymbol{w}}^k) \ge (\boldsymbol{w} - \widehat{\boldsymbol{w}}^k)^\top Q_k (\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k), \ \ \forall \boldsymbol{w} \in \mathscr{W}, \tag{A.7}$$

(**correction step**) Compute a correction via

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k - M_k(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k), \tag{A.8}$$

where

$$\widehat{\boldsymbol{u}}^k = \begin{pmatrix} \widehat{\boldsymbol{x}}^k \\ \widehat{\boldsymbol{y}}^k \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}^{k+1} \\ \boldsymbol{y}^{k+1} \end{pmatrix}, \quad \widehat{\boldsymbol{w}}^k = \begin{pmatrix} \widehat{\boldsymbol{x}}^k \\ \widehat{\boldsymbol{y}}^k \\ \widehat{\boldsymbol{\lambda}}^k \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}^{k+1} \\ \boldsymbol{y}^{k+1} \\ \boldsymbol{\lambda}^k - \beta_k \left(A\boldsymbol{x}^{k+1} + B\boldsymbol{y}^k - b \right) \end{pmatrix}, \tag{A.9}$$

$$Q_k = \begin{pmatrix} \frac{\beta_k}{\gamma_1} I - \beta_k A^\top A & 0 & 0 \\ 0 & \frac{\beta_k}{\tau\gamma_2} I & 0 \\ 0 & -B & \frac{1}{\beta_k} I \end{pmatrix} \ \ \textit{and} \ \ M_k = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -\beta_k B & I \end{pmatrix}. \tag{A.10}$$

Proof. By invoking the first-order optimality conditions of (A.2a) and (A.2b), it follows from the notation $\boldsymbol{\xi}_1^k$ and $\boldsymbol{\xi}_2^k$ that

$$\theta_1(\boldsymbol{x})-\theta_1(\boldsymbol{x}^{k+1})+(\boldsymbol{x}-\boldsymbol{x}^{k+1})^\top\left[-A^\top\boldsymbol{\lambda}^k+\beta_kA^\top(A\boldsymbol{x}^k+B\boldsymbol{y}^k-b)+\frac{\beta_k}{\gamma_1}(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k)\right]\geq 0 \tag{A.11}$$

and

$$\theta_2(\boldsymbol{y})-\theta_2(\boldsymbol{y}^{k+1})+(\boldsymbol{y}-\boldsymbol{y}^{k+1})^\top\left[-B^\top\boldsymbol{\lambda}^k+\beta_kB^\top(A\boldsymbol{x}^{k+1}+B\boldsymbol{y}^k-b)+\frac{\beta_k}{\tau\gamma_2}(\boldsymbol{y}^{k+1}-\boldsymbol{y}^k)\right]\geq 0 \tag{A.12}$$

holds for all $\boldsymbol{x}\in\mathscr{X}$ and $\boldsymbol{y}\in\mathscr{Y}$. According to the definition of $\widehat{\boldsymbol{w}}^k$ in (A.9), we can further rewrite (A.11) and (A.12) as

$$\begin{cases}\theta_1(\boldsymbol{x})-\theta_1(\widehat{\boldsymbol{x}}^k)+(\boldsymbol{x}-\widehat{\boldsymbol{x}}^k)^\top\left[-A^\top\widehat{\boldsymbol{\lambda}}^k+\left(\frac{\beta_k}{\gamma_1}I-\beta_kA^\top A\right)(\widehat{\boldsymbol{x}}^k-\boldsymbol{x}^k)\right]\geq 0, & \text{(A.13)}\\[2mm] \theta_2(\boldsymbol{y})-\theta_2(\widehat{\boldsymbol{y}}^k)+(\boldsymbol{y}-\widehat{\boldsymbol{y}}^k)^\top\left[-B^\top\widehat{\boldsymbol{\lambda}}^k+\frac{\beta_k}{\tau\gamma_2}(\widehat{\boldsymbol{y}}^k-\boldsymbol{y}^k)\right]\geq 0, & \text{(A.14)}\end{cases}$$

which holds for all $\boldsymbol{x}\in\mathscr{X}$ and $\boldsymbol{y}\in\mathscr{Y}$. Similarly, (A.2c) can be recast as

$$(\boldsymbol{\lambda}-\widehat{\boldsymbol{\lambda}}^k)^\top\left[(A\widehat{\boldsymbol{x}}^k+B\widehat{\boldsymbol{y}}^k-b)-B(\widehat{\boldsymbol{y}}^k-y^k)+\frac{1}{\beta_k}(\widehat{\boldsymbol{\lambda}}^k-\boldsymbol{\lambda}^k)\right]\geq 0,\ \ \forall\boldsymbol{\lambda}\in\mathbb{R}^m. \tag{A.15}$$

With the help of notation Q_k given in (A.10), we can rewrite (A.13), (A.14), and (A.15) into a compact form as follows:

$$\theta(\boldsymbol{u})-\theta(\widehat{\boldsymbol{u}}^k)+(\boldsymbol{w}-\widehat{\boldsymbol{w}}^k)^\top F(\widehat{\boldsymbol{w}}^k)\geq(\boldsymbol{w}-\widehat{\boldsymbol{w}}^k)^\top Q_k(\boldsymbol{w}^k-\widehat{\boldsymbol{w}}^k),\ \ \forall\boldsymbol{w}\in\mathscr{W},$$

which corresponds to the prediction step (A.7).

Recalling the definition $\widehat{\boldsymbol{\lambda}}^k$ in (A.9), we have

$$\boldsymbol{\lambda}^{k+1}=\boldsymbol{\lambda}^k-\left[-\beta_kB(\boldsymbol{y}^k-\widehat{\boldsymbol{y}}^k)+(\boldsymbol{\lambda}^k-\widehat{\boldsymbol{\lambda}}^k)\right],$$

which, together with $\widehat{\boldsymbol{w}}^k$ given in (A.9) and M_k defined by (A.10), implies the correction step:

$$\boldsymbol{w}^{k+1}=\boldsymbol{w}^k-M_k(\boldsymbol{w}^k-\widehat{\boldsymbol{w}}^k).$$

The proof is complete. □

Before starting the convergence analysis of (A.2), we first introduce some more notations:

$$\widetilde{A}_k=\frac{\beta_k}{\gamma_1}I-\beta_kA^\top A,\quad D_k=\frac{\beta_k}{\gamma_2}I-\beta_kB^\top B,\quad E_k=\frac{\beta_k}{\tau\gamma_2}I-\beta_kB^\top B, \tag{A.16}$$

$$H_k=\begin{pmatrix}\widetilde{A}_k & 0 & 0\\ 0 & \frac{\beta_k}{\tau\gamma_2}I & 0\\ 0 & 0 & \frac{1}{\beta_k}I\end{pmatrix}\quad\text{and}\quad G_k=\begin{pmatrix}\widetilde{A}_k & 0 & 0\\ 0 & E_k & 0\\ 0 & 0 & \frac{1}{\beta_k}I\end{pmatrix}. \tag{A.17}$$

Clearly, $\gamma_1\in\left(0,1/\rho(A^\top A)\right]$ can ensure that $\widetilde{A}_k$ is a positive definite matrix, so H_k is also positive definite. Similarly, $\gamma_2\in\left(0,1/\rho(B^\top B)\right]$ can ensure that D_k is a positive definite matrix, while it cannot guarantee the positive definiteness of E_k and G_k due to the appearance of τ in E_k. Recalling the definitions of Q_k and M_k in (A.10), we have

$$H_k = Q_k {M_k}^{-1} \quad and \quad G_k = {Q_k}^\top + Q_k - {M_k}^\top H_k M_k. \tag{A.18}$$

Below, we establish an inequality for the prediction point.

Lemma A.1. *Let $\{\boldsymbol{w}^k\}$ be the sequence generated by LR-ADMM (A.2) and $\widehat{\boldsymbol{w}}^k$ be a prediction point defined in (A.9). Then, for all $\boldsymbol{w} \in \mathscr{W}$,*

$$\begin{aligned} &\theta(\boldsymbol{u}) - \theta(\widehat{\boldsymbol{u}}^k) + (\boldsymbol{w} - \widehat{\boldsymbol{w}}^k)^\top F(\boldsymbol{w}) \\ &\geq \frac{1}{2}\left(\left\|\boldsymbol{w} - \boldsymbol{w}^{k+1}\right\|_{H_k}^2 - \left\|\boldsymbol{w} - \boldsymbol{w}^k\right\|_{H_k}^2\right) + \frac{1}{2}(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k)^\top G_k(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k), \end{aligned} \tag{A.19}$$

where H_k and G_k are defined in (A.17).

Proof. Combining $H_k = Q_k {M_k}^{-1}$ with (A.8), we have

$$(\boldsymbol{w} - \widehat{\boldsymbol{w}}^k)^\top Q_k(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k) = (\boldsymbol{w} - \widehat{\boldsymbol{w}}^k)^\top H_k(\boldsymbol{w}^k - \boldsymbol{w}^{k+1}),$$

which, together with (A.7), leads to

$$\theta(\boldsymbol{u}) - \theta(\widehat{\boldsymbol{u}}^k) + (\boldsymbol{w} - \widehat{\boldsymbol{w}}^k)^\top F(\widehat{\boldsymbol{w}}^k) \geq (\boldsymbol{w} - \widehat{\boldsymbol{w}}^k)^\top H_k(\boldsymbol{w}^k - \boldsymbol{w}^{k+1}), \ \forall \boldsymbol{w} \in \mathscr{W}. \tag{A.20}$$

We now recall the identity that holds for all $a, b, c, d \in \mathbb{R}^n$ and a given positive semi-definite matrix H that

$$(a-b)^\top H(c-d) = \frac{1}{2}\left(\|a-d\|_H^2 - \|a-c\|_H^2\right) + \frac{1}{2}\left(\|c-b\|_H^2 - \|d-b\|_H^2\right).$$

As a consequence, applying the above identity to the right-hand side of (A.20) with settings $a = \boldsymbol{w}$, $b = \widehat{\boldsymbol{w}}^k$, $c = \boldsymbol{w}^k$, $d = \boldsymbol{w}^{k+1}$, and $H = H_k$ produces

$$\begin{aligned} &(\boldsymbol{w} - \widehat{\boldsymbol{w}}^k)^\top H_k(\boldsymbol{w}^k - \boldsymbol{w}^{k+1}) - \frac{1}{2}\left(\left\|\boldsymbol{w} - \boldsymbol{w}^{k+1}\right\|_{H_k}^2 - \left\|\boldsymbol{w} - \boldsymbol{w}^k\right\|_{H_k}^2\right) \\ &= \frac{1}{2}\left(\left\|\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k\right\|_{H_k}^2 - \left\|\boldsymbol{w}^{k+1} - \widehat{\boldsymbol{w}}^k\right\|_{H_k}^2\right) \\ &= \frac{1}{2}\left(\left\|\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k\right\|_{H_k}^2 - \left\|(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k) - M_k(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k)\right\|_{H_k}^2\right) \\ &= (\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k)^\top H_k M_k(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k) - \frac{1}{2}(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k)^\top {M_k}^\top H_k M_k(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k) \\ &= \frac{1}{2}(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k)^\top ({Q_k}^\top + Q_k - {M_k}^\top H_k M_k)(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k) \\ &= \frac{1}{2}(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k)^\top G_k(\boldsymbol{w}^k - \widehat{\boldsymbol{w}}^k), \end{aligned} \tag{A.21}$$

where the second equality follows from (A.8), and the last two equalities come from (A.18).

On the other hand, it follows from (A.6) that

$$(\boldsymbol{w} - \widehat{\boldsymbol{w}}^k)^\top F(\widehat{\boldsymbol{w}}^k) = (\boldsymbol{w} - \widehat{\boldsymbol{w}}^k)^\top F(\boldsymbol{w}). \tag{A.22}$$

Consequently, substituting (A.21) and (A.22) into (A.20) arrives at the assertion of this lemma. We complete its proof. □

Lemma A.2. *Let $\{\boldsymbol{w}^k\}$ be the sequence generated by LR-ADMM* (A.2) *and $\widehat{\boldsymbol{w}}^k$ be a prediction point defined in* (A.9)*. Then,*

$$(\boldsymbol{w}^k-\widehat{\boldsymbol{w}}^k)^\top G_k(\boldsymbol{w}^k-\widehat{\boldsymbol{w}}^k)=\left\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\right\|_{\widetilde{A}_k}^2+\frac{1}{\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2+\frac{\beta_k}{\tau}\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2$$
$$+\frac{1}{\beta_k}\left\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\right\|^2+2(\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1})^\top B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1}). \qquad \text{(A.23)}$$

Proof. It follows from the definitions of G_k in (A.17) and $\widehat{\boldsymbol{w}}^k$ in (A.9) that

$$\begin{aligned}
&(\boldsymbol{w}^k-\widehat{\boldsymbol{w}}^k)^\top G_k(\boldsymbol{w}^k-\widehat{\boldsymbol{w}}^k)\\
&=\left\|\boldsymbol{x}^k-\widehat{\boldsymbol{x}}^k\right\|_{\widetilde{A}_k}^2+\frac{\beta_k}{\tau\gamma_2}\left\|\boldsymbol{y}^k-\widehat{\boldsymbol{y}}^k\right\|^2-\beta_k\left\|B(\boldsymbol{y}^k-\widehat{\boldsymbol{y}}^k)\right\|^2+\frac{1}{\beta_k}\left\|\boldsymbol{\lambda}^k-\widehat{\boldsymbol{\lambda}}^k\right\|^2\\
&=\left\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\right\|_{\widetilde{A}_k}^2+\frac{\beta_k}{\tau\gamma_2}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|^2-\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2+\frac{1}{\beta_k}\left\|\boldsymbol{\lambda}^k-\widehat{\boldsymbol{\lambda}}^k\right\|^2,
\end{aligned}$$

where the last term, based on the definition of $\widehat{\boldsymbol{\lambda}}^k$ in (A.9), can be further rewritten as

$$\begin{aligned}
\frac{1}{\beta_k}\left\|\boldsymbol{\lambda}^k-\widehat{\boldsymbol{\lambda}}^k\right\|^2&=\beta_k\left\|(A\boldsymbol{x}^{k+1}+B\boldsymbol{y}^{k+1}-b)+B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2\\
&=\beta_k\left\|\frac{1}{\beta_k}(\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1})+B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2\\
&=\frac{1}{\beta_k}\left\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\right\|^2+2(\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1})^\top B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})+\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2.
\end{aligned}$$

Hence, combining the above two equalities, we arrive at

$$(\boldsymbol{w}^k-\widehat{\boldsymbol{w}}^k)^\top G_k(\boldsymbol{w}^k-\widehat{\boldsymbol{w}}^k)=\left\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\right\|_{\widetilde{A}_k}^2+\frac{\beta_k}{\tau\gamma_2}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|^2+\frac{1}{\beta_k}\left\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\right\|^2$$
$$+2(\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1})^\top B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1}). \qquad \text{(A.24)}$$

On the other hand, from the definition of D_k in (A.16) and the positive definiteness of D_k under the condition $\gamma_2\in\big(0,1/\rho(B^\top B)\big]$, we have

$$\frac{\beta_k}{\tau\gamma_2}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|^2=\frac{1}{\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2+\frac{\beta_k}{\tau}\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2,$$

which immediately leads to the assertion of this lemma by plugging it into (A.24). □

Lemma A.3. *Let $\{\boldsymbol{w}^k\}$ be the sequence generated by LR-ADMM* (A.2) *and $\widehat{\boldsymbol{w}}^k$ be a prediction point defined in* (A.9)*. Then,*

$$\begin{aligned}
&(\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1})^\top B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\\
&\geq\frac{1}{2\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2-\frac{3(\tau-1)}{2\tau}\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2-\frac{1}{2\tau}\left\|\boldsymbol{y}^{k-1}-\boldsymbol{y}^k\right\|_{D_{k-1}}^2\\
&\quad-\frac{\tau-1}{2\tau}\beta_{k-1}\left\|B(\boldsymbol{y}^{k-1}-\boldsymbol{y}^k)\right\|^2.
\end{aligned}\qquad \text{(A.25)}$$

Proof. It follows from (A.2c), (A.12), and the definition of E_k in (A.16) that

$$\theta_2(\boldsymbol{y}) - \theta_2(\boldsymbol{y}^{k+1}) + (\boldsymbol{y} - \boldsymbol{y}^{k+1})^\top \left[-B^\top \boldsymbol{\lambda}^{k+1} + E_k(\boldsymbol{y}^{k+1} - \boldsymbol{y}^k) \right] \geq 0 \tag{A.26}$$

holds for any $\boldsymbol{y} \in \mathscr{Y}$. Similarly, we also have

$$\theta_2(\boldsymbol{y}) - \theta_2(\boldsymbol{y}^k) + (\boldsymbol{y} - \boldsymbol{y}^k)^\top \left[-B^\top \boldsymbol{\lambda}^k + E_{k-1}(\boldsymbol{y}^k - \boldsymbol{y}^{k-1}) \right] \geq 0, \ \ \forall \boldsymbol{y} \in \mathscr{Y}. \tag{A.27}$$

Consequently, setting $\boldsymbol{y} = \boldsymbol{y}^k$ and $\boldsymbol{y} = \boldsymbol{y}^{k+1}$ in (A.26) and (A.27), respectively, and adding both of them immediately yield

$$(\boldsymbol{y}^k - \boldsymbol{y}^{k+1})^\top \left[B^\top (\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k+1}) + E_k(\boldsymbol{y}^{k+1} - \boldsymbol{y}^k) - E_{k-1}(\boldsymbol{y}^k - \boldsymbol{y}^{k-1}) \right] \geq 0,$$

which implies

$$\begin{aligned}
&(\boldsymbol{y}^k - \boldsymbol{y}^{k+1})^\top B^\top (\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k+1}) \\
&\geq (\boldsymbol{y}^k - \boldsymbol{y}^{k+1})^\top E_k (\boldsymbol{y}^k - \boldsymbol{y}^{k+1}) - (\boldsymbol{y}^k - \boldsymbol{y}^{k+1})^\top E_{k-1} (\boldsymbol{y}^{k-1} - \boldsymbol{y}^k) \\
&\overset{(A.16)}{=} (\boldsymbol{y}^k - \boldsymbol{y}^{k+1})^\top \left(\frac{1}{\tau} D_k - \frac{\tau - 1}{\tau} \beta_k B^\top B \right) (\boldsymbol{y}^k - \boldsymbol{y}^{k+1}) \\
&\quad - (\boldsymbol{y}^k - \boldsymbol{y}^{k+1})^\top \left(\frac{1}{\tau} D_{k-1} - \frac{\tau - 1}{\tau} \beta_{k-1} B^\top B \right) (\boldsymbol{y}^{k-1} - \boldsymbol{y}^k) \\
&= \frac{1}{\tau} \left\| \boldsymbol{y}^k - \boldsymbol{y}^{k+1} \right\|_{D_k}^2 - \frac{1}{\tau} (\boldsymbol{y}^k - \boldsymbol{y}^{k+1})^\top D_{k-1} (\boldsymbol{y}^{k-1} - \boldsymbol{y}^k) \\
&\quad - \frac{\tau - 1}{\tau} \beta_k \left\| B(\boldsymbol{y}^k - \boldsymbol{y}^{k+1}) \right\|^2 + \frac{\tau - 1}{\tau} \beta_{k-1} (\boldsymbol{y}^k - \boldsymbol{y}^{k+1})^\top B^\top B (\boldsymbol{y}^{k-1} - \boldsymbol{y}^k) \\
&\geq \frac{1}{\tau} \left\| \boldsymbol{y}^k - \boldsymbol{y}^{k+1} \right\|_{D_k}^2 - \frac{1}{2\tau} \left\| \boldsymbol{y}^k - \boldsymbol{y}^{k+1} \right\|_{D_{k-1}}^2 - \frac{1}{2\tau} \left\| \boldsymbol{y}^{k-1} - \boldsymbol{y}^k \right\|_{D_{k-1}}^2 \\
&\quad - \left(\frac{\tau - 1}{\tau} \beta_k + \frac{\tau - 1}{2\tau} \beta_{k-1} \right) \left\| B(\boldsymbol{y}^k - \boldsymbol{y}^{k+1}) \right\|^2 - \frac{\tau - 1}{2\tau} \beta_{k-1} \left\| B(\boldsymbol{y}^{k-1} - \boldsymbol{y}^k) \right\|^2,
\end{aligned}$$

where the last inequality follows from the application of $2a^\top b \geq -\|a\|^2 - \|b\|^2$ for any $a, b \in \mathbb{R}^n$. Thanks to the assumption $\beta_{k-1} \leq \beta_k$, we know from the definition of D_k in (A.16) that $D_k \succeq D_{k-1}$, and then the above inequality turns out

$$\begin{aligned}
(\boldsymbol{y}^k - \boldsymbol{y}^{k+1})^\top B^\top (\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k+1}) \geq &\frac{1}{2\tau} \left\| \boldsymbol{y}^k - \boldsymbol{y}^{k+1} \right\|_{D_k}^2 - \frac{1}{2\tau} \left\| \boldsymbol{y}^{k-1} - \boldsymbol{y}^k \right\|_{D_{k-1}}^2 \\
&- \frac{3(\tau - 1)}{2\tau} \beta_k \left\| B(\boldsymbol{y}^k - \boldsymbol{y}^{k+1}) \right\|^2 - \frac{\tau - 1}{2\tau} \beta_{k-1} \left\| B(\boldsymbol{y}^{k-1} - \boldsymbol{y}^k) \right\|^2,
\end{aligned}$$

which arrives at the assertion of this lemma. □

Lemma A.4. *Let* $\{\boldsymbol{w}^k\}$ *be the sequence generated by LR-ADMM* (A.2) *and* $\widehat{\boldsymbol{w}}^k$ *be a prediction point defined in* (A.9)*. Then, there exists a constant* $\delta \in (0, 1/2)$ *such that*

$$(\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k+1})^\top B(\boldsymbol{y}^k - \boldsymbol{y}^{k+1}) \geq - \left(\frac{1}{4} + \frac{1}{2}\delta \right) \beta_k \left\| B(\boldsymbol{y}^k - \boldsymbol{y}^{k+1}) \right\|^2 - \frac{1 - \delta}{\beta_k} \left\| \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k+1} \right\|^2. \tag{A.28}$$

Proof. It follows from the Young's inequality that

$$(\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1})^\top B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1}) \geq -\frac{\beta_k}{4(1-\delta)}\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2-\frac{1-\delta}{\beta_k}\left\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\right\|^2 \tag{A.29}$$

always holds for any $\delta \in (0,1)$. Consequently, it is clear that $\delta(1-2\delta)>0$ for any $\delta \in (0,1/2)$, and then

$$\frac{1}{4(1-\delta)}<\frac{1}{4}+\frac{\delta}{2},$$

which together with (A.29) immediately implies (A.28). □

With the preparations of Lemmas A.1-A.4, we can obtain the following inequality.

Lemma A.5. *Suppose that $\gamma_1 \in \left(0,1/\rho(A^\top A)\right]$, $\gamma_2 \in \left(0,1/\rho(B^\top B)\right]$, $\tau \in (1,4/3)$, and $\delta = 2\left(\frac{1}{\tau}-\frac{3}{4}\right)$. Let $\{\boldsymbol{w}^k\}$ be the sequence generated by LR-ADMM (A.2) and $\widehat{\boldsymbol{w}}^k$ be a prediction point defined in (A.9). Then,*

$$\begin{aligned}
&\theta(\boldsymbol{u})-\theta(\widehat{\boldsymbol{u}}^k)+(\boldsymbol{w}-\widehat{\boldsymbol{w}}^k)^\top F(\boldsymbol{w})\\
\geq&\left(\frac{1}{2}\left\|\boldsymbol{w}-\boldsymbol{w}^{k+1}\right\|_{H_k}^2+\frac{1}{4}\left(\frac{1}{\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2+\frac{\tau-1}{\tau}\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2\right)\right)\\
&-\left(\frac{1}{2}\left\|\boldsymbol{w}-\boldsymbol{w}^{k}\right\|_{H_k}^2+\frac{1}{4}\left(\frac{1}{\tau}\left\|\boldsymbol{y}^{k-1}-\boldsymbol{y}^{k}\right\|_{D_{k-1}}^2+\frac{\tau-1}{\tau}\beta_{k-1}\left\|B(\boldsymbol{y}^{k-1}-\boldsymbol{y}^{k})\right\|^2\right)\right)\\
&+\frac{1}{2}\left\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\right\|_{\widetilde{A}_k}^2+\frac{1}{2\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2\\
&+\left(\frac{1}{\tau}-\frac{3}{4}\right)\left(\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2+\frac{1}{\beta_k}\left\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\right\|^2\right).
\end{aligned} \tag{A.30}$$

Proof. It is clear from $\tau \in (1,4/3)$ and $\delta = 2\left(\frac{1}{\tau}-\frac{3}{4}\right)$ that $\delta \in (0,1/2)$ holds. Thus, Lemma A.4 holds by setting $\delta = 2\left(\frac{1}{\tau}-\frac{3}{4}\right)$. Consequently, adding (A.25) and (A.28) leads to

$$\begin{aligned}
&2(\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1})^\top B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\\
\geq&\left(\frac{1}{2\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2+\frac{\tau-1}{2\tau}\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2\right)\\
&-\left(\frac{1}{2\tau}\left\|\boldsymbol{y}^{k-1}-\boldsymbol{y}^{k}\right\|_{D_{k-1}}^2+\frac{\tau-1}{2\tau}\beta_{k-1}\left\|B(\boldsymbol{y}^{k-1}-\boldsymbol{y}^{k})\right\|^2\right)\\
&-\frac{2\beta_k(\tau-1)}{\tau}\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2-\left(\frac{1}{4}+\frac{1}{2}\delta\right)\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2-\frac{1-\delta}{\beta_k}\left\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\right\|^2\\
=&\left(\frac{1}{2\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2+\frac{\tau-1}{2\tau}\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2\right)\\
&-\left(\frac{1}{2\tau}\left\|\boldsymbol{y}^{k-1}-\boldsymbol{y}^{k}\right\|_{D_{k-1}}^2+\frac{\tau-1}{2\tau}\beta_{k-1}\left\|B(\boldsymbol{y}^{k-1}-\boldsymbol{y}^{k})\right\|^2\right)\\
&+\left(\frac{1}{\tau}-\frac{3}{2}\right)\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2+\left(\frac{2}{\tau}-\frac{5}{2}\right)\frac{1}{\beta_k}\left\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\right\|^2.
\end{aligned}$$

Substituting it into (A.23) produces

$$(\boldsymbol{w}^k-\widehat{\boldsymbol{w}}^k)^\top G_k(\boldsymbol{w}^k-\widehat{\boldsymbol{w}}^k)$$

$$\geq \left(\frac{1}{2\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2+\frac{\tau-1}{2\tau}\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2\right)$$
$$-\left(\frac{1}{2\tau}\left\|\boldsymbol{y}^{k-1}-\boldsymbol{y}^k\right\|_{D_{k-1}}^2+\frac{\tau-1}{2\tau}\beta_{k-1}\left\|B(\boldsymbol{y}^{k-1}-\boldsymbol{y}^k)\right\|^2\right)+\left\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\right\|_{\widetilde{A}_k}^2$$
$$+\frac{1}{\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2+2\left(\frac{1}{\tau}-\frac{3}{4}\right)\left(\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2+\frac{1}{\beta_k}\left\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\right\|^2\right),$$

which together with (A.19) implies the assertion of this lemma. □

Lemma A.6. *Suppose that* $\gamma_1\in\left(0,1/\rho(A^\top A)\right]$, $\gamma_2\in\left(0,1/\rho(B^\top B)\right]$, $\tau\in(1,4/3)$, *and* $\delta=2\left(\frac{1}{\tau}-\frac{3}{4}\right)$. *Let* $\{\boldsymbol{w}^k\}$ *be the sequence generated by LR-ADMM* (A.2). *Then,*

$$\left\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\right\|_{\widetilde{A}_k}^2+\frac{1}{\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2$$
$$+2\left(\frac{1}{\tau}-\frac{3}{4}\right)\left(\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2+\frac{1}{\beta_k}\left\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\right\|^2\right)$$
$$\leq\left(\left\|\boldsymbol{w}^k-\boldsymbol{w}^*\right\|_{H_k}^2+\frac{1}{2}\left(\frac{1}{\tau}\left\|\boldsymbol{y}^{k-1}-\boldsymbol{y}^k\right\|_{D_{k-1}}^2+\frac{\tau-1}{\tau}\beta_{k-1}\left\|B(\boldsymbol{y}^{k-1}-\boldsymbol{y}^k)\right\|^2\right)\right)$$
$$-\left(\left\|\boldsymbol{w}^{k+1}-\boldsymbol{w}^*\right\|_{H_k}^2+\frac{1}{2}\left(\frac{1}{\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2+\frac{\tau-1}{\tau}\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2\right)\right). \tag{A.31}$$

Proof. Due to arbitrariness of $\boldsymbol{w}$ in (A.30), we set $\boldsymbol{w}=\boldsymbol{w}^*$ and obtain

$$\left(\frac{1}{2}\left\|\boldsymbol{w}^k-\boldsymbol{w}^*\right\|_{H_k}^2+\frac{1}{4}\left(\frac{1}{\tau}\left\|\boldsymbol{y}^{k-1}-\boldsymbol{y}^k\right\|_{D_{k-1}}^2+\frac{\tau-1}{\tau}\beta_{k-1}\left\|B(\boldsymbol{y}^{k-1}-\boldsymbol{y}^k)\right\|^2\right)\right)$$
$$-\left(\frac{1}{2}\left\|\boldsymbol{w}^{k+1}-\boldsymbol{w}^*\right\|_{H_k}^2+\frac{1}{4}\left(\frac{1}{\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2+\frac{\tau-1}{\tau}\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2\right)\right)$$
$$\geq\frac{1}{2}\left\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\right\|_{\widetilde{A}_k}^2+\frac{1}{2\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_{D_k}^2$$
$$+\left(\frac{1}{\tau}-\frac{3}{4}\right)\left(\beta_k\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2+\frac{1}{\beta_k}\left\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\right\|^2\right)$$
$$+(\theta(\widehat{\boldsymbol{u}}^k)-\theta(\boldsymbol{u}^*))+(\widehat{\boldsymbol{w}}^k-\boldsymbol{w}^*)^\top F(\boldsymbol{w}^*). \tag{A.32}$$

Similarly, thanks to the arbitrariness of $\boldsymbol{u}$ and $\boldsymbol{w}$, we take $\boldsymbol{u}=\widehat{\boldsymbol{u}}^k$ and $\boldsymbol{w}=\widehat{\boldsymbol{w}}^k$ in (A.5), then obtain

$$\theta(\widehat{\boldsymbol{u}}^k)-\theta(\boldsymbol{u}^*)+(\widehat{\boldsymbol{w}}^k-\boldsymbol{w}^*)^\top F(\boldsymbol{w}^*)\geq 0,$$

which, together with (A.32), yields the conclusion of this lemma. □

Theorem A.1. *Suppose that* $\gamma_1\in\left(0,1/\rho(A^\top A)\right]$, $\gamma_2\in\left(0,1/\rho(B^\top B)\right]$, $\tau\in(1,4/3)$, *and* $\delta=2\left(\frac{1}{\tau}-\frac{3}{4}\right)$. *Let* $\{\boldsymbol{w}^k\}$ *be the sequence generated by LR-ADMM* (A.2). *Then, LR-ADMM* (A.2) *is convergent.*

Proof. For notational simplicity, we first denote

$$\eta_k=\frac{\beta_k}{\beta_{k-1}}-1,\ \ \widetilde{C}_s=\sum_{k=1}^{\infty}\eta_k,\ \text{ and }\ \widetilde{C}_p=\prod_{k=1}^{\infty}(\eta_k+1).$$

Due to the monotone increasing property and upper boundedness of $\{\beta_k\}$, we can easily check that $\eta_k \geq 0$, $\widetilde{C}_s < +\infty$, and $\widetilde{C}_p < +\infty$. Additionally, we let

$$\begin{aligned}\phi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1}) = &\left\| \boldsymbol{x}^k - \boldsymbol{x}^{k+1} \right\|_{\widetilde{A}_k}^2 + \frac{1}{\tau} \left\| \boldsymbol{y}^k - \boldsymbol{y}^{k+1} \right\|_{D_k}^2 \\ &+ 2\left(\frac{1}{\tau} - \frac{3}{4}\right)\left(\beta_k \left\| B(\boldsymbol{y}^k - \boldsymbol{y}^{k+1}) \right\|^2 + \frac{1}{\beta_k} \left\| \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k+1} \right\|^2\right)\end{aligned}$$

and

$$\psi(\boldsymbol{w}^{k-1}, \boldsymbol{w}^k) = \frac{1}{2}\left(\frac{1}{\tau} \left\| \boldsymbol{y}^{k-1} - \boldsymbol{y}^k \right\|_{D_{k-1}}^2 + \frac{\tau - 1}{\tau} \beta_{k-1} \left\| B(\boldsymbol{y}^{k-1} - \boldsymbol{y}^k) \right\|^2\right).$$

Consequently,inequality (A.31) can be simplified as follows

$$\phi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1}) \leq \left(\left\| \boldsymbol{w}^k - \boldsymbol{w}^* \right\|_{H_k}^2 + \psi(\boldsymbol{w}^{k-1}, \boldsymbol{w}^k)\right) - \left(\left\| \boldsymbol{w}^{k+1} - \boldsymbol{w}^* \right\|_{H_k}^2 + \psi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1})\right),$$

which, by rearranging terms and invoking the definition of H_k in (A.17), implies

$$\begin{aligned}&\left\| \boldsymbol{w}^{k+1} - \boldsymbol{w}^* \right\|_{H_k}^2 + \psi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1}) \\ &\leq \left\| \boldsymbol{w}^k - \boldsymbol{w}^* \right\|_{H_k}^2 + \psi(\boldsymbol{w}^{k-1}, \boldsymbol{w}^k) - \phi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1}) \\ &= \left\| \boldsymbol{x}^k - \boldsymbol{x}^* \right\|_{\widetilde{A}_k}^2 + \frac{\beta_k}{\tau\gamma_2} \left\| \boldsymbol{y}^k - \boldsymbol{y}^* \right\|^2 + \frac{1}{\beta_k} \left\| \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^* \right\|^2 + \psi(\boldsymbol{w}^{k-1}, \boldsymbol{w}^k) - \phi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1}) \\ &\leq \frac{\beta_k}{\beta_{k-1}}\left[\left\| \boldsymbol{x}^k - \boldsymbol{x}^* \right\|_{\widetilde{A}_{k-1}}^2 + \frac{\beta_{k-1}}{\tau\gamma_2} \left\| \boldsymbol{y}^k - \boldsymbol{y}^* \right\|^2 + \frac{\beta_{k-1}}{\beta_k^2} \left\| \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^* \right\|^2\right] \\ &\quad + \psi(\boldsymbol{w}^{k-1}, \boldsymbol{w}^k) - \phi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1}) \\ &\leq (\eta_k + 1)\left[\left\| \boldsymbol{x}^k - \boldsymbol{x}^* \right\|_{\widetilde{A}_{k-1}}^2 + \frac{\beta_{k-1}}{\tau\gamma_2} \left\| \boldsymbol{y}^k - \boldsymbol{y}^* \right\|^2 + \frac{1}{\beta_{k-1}} \left\| \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^* \right\|^2\right] \\ &\quad + \psi(\boldsymbol{w}^{k-1}, \boldsymbol{w}^k) - \phi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1}) \\ &\leq (\eta_k + 1)\left(\left\| \boldsymbol{w}^k - \boldsymbol{w}^* \right\|_{H_{k-1}}^2 + \psi(\boldsymbol{w}^{k-1}, \boldsymbol{w}^k)\right) - \phi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1}). \qquad \text{(A.33)}\end{aligned}$$

By invoking the conditions $\gamma_1 \in \left(0, 1/\rho(A^\top A)\right]$ and $\tau \in (1, 4/3)$, it is easy to see that $\phi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1}) \geq 0$. Hence, it follows from (A.33) that

$$\begin{aligned}\left\| \boldsymbol{w}^{k+1} - \boldsymbol{w}^* \right\|_{H_k}^2 + \psi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1}) &\leq (\eta_k + 1)\left(\left\| \boldsymbol{w}^k - \boldsymbol{w}^* \right\|_{H_{k-1}}^2 + \psi(\boldsymbol{w}^{k-1}, \boldsymbol{w}^k)\right) \\ &\leq \widetilde{C}_p\left(\left\| \boldsymbol{w}^1 - \boldsymbol{w}^* \right\|_{H_0}^2 + \psi(\boldsymbol{w}^0, \boldsymbol{w}^1)\right),\end{aligned}$$

which immediately implies that there exists a constant $\widetilde{C} > 0$ such that

$$\left\| \boldsymbol{w}^{k+1} - \boldsymbol{w}^* \right\|_{H_k}^2 + \psi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1}) \leq \widetilde{C}, \quad \forall k \geq 0.$$

As a consequence, it follows from (A.33) that

$$\sum_{k=0}^{\infty} \phi(\boldsymbol{w}^k, \boldsymbol{w}^{k+1})$$

$$\leq \sum_{k=0}^{\infty}\left[(\eta_k+1)\left(\left\|\boldsymbol{w}^k-\boldsymbol{w}^*\right\|_{H_{k-1}}^2+\psi(\boldsymbol{w}^{k-1},\boldsymbol{w}^k)\right)-\left(\left\|\boldsymbol{w}^{k+1}-\boldsymbol{w}^*\right\|_{H_k}^2+\psi(\boldsymbol{w}^k,\boldsymbol{w}^{k+1})\right)\right]$$
$$\leq \sum_{k=0}^{\infty}\eta_k\left(\left\|\boldsymbol{w}^k-\boldsymbol{w}^*\right\|_{H_{k-1}}^2+\psi(\boldsymbol{w}^{k-1},\boldsymbol{w}^k)\right)+\left(\|\boldsymbol{w}^1-\boldsymbol{w}^*\|_{H_0}^2+\psi(\boldsymbol{w}^0,\boldsymbol{w}^1)\right)$$
$$\leq (\widetilde{C}_s+1)\widetilde{C}.$$

which, together with the nonnegativity of $\phi(\boldsymbol{w}^k,\boldsymbol{w}^{k+1})$, implies that

$$\lim_{k\to\infty}\phi(\boldsymbol{w}^k,\boldsymbol{w}^{k+1})=0 \quad \text{and} \quad \lim_{k\to\infty}\|\boldsymbol{w}^k-\boldsymbol{w}^{k+1}\|=0. \tag{A.34}$$

Therefore, we conclude that LR-ADMM (A.2) is convergent. □

Notice that the dynamical strategy on β has a preset upper bound. In other words, after finite iterations, β_k must be a constant. Therefore, to simplify the convergence, we below consider the case of $\beta_k\equiv\beta$ (we correspondingly let $D=D_k$ and $H=H_k$) and show that the sequence $\{\boldsymbol{w}^k\}$ generated by LR-ADMM (A.2) globally converges to a solution of (A.1).

Theorem A.2. *Suppose that $\gamma_1\in\left(0,1/\rho(A^\top A)\right]$, $\gamma_2\in\left(0,1/\rho(B^\top B)\right]$, $\tau\in(1,4/3)$, and $\delta=2\left(\frac{1}{\tau}-\frac{3}{4}\right)$. Then, the sequence $\{\boldsymbol{w}^k\}$ generated by LR-ADMM (A.2) globally converges to a point of the solution set $\mathscr{W}^*$.*

Proof. Under the setting $\beta_k=\beta$, it follows from (A.33) that

$$\begin{aligned}\left\|\boldsymbol{w}^{k+1}-\boldsymbol{w}^*\right\|_H^2 &\leq \left\|\boldsymbol{w}^{k+1}-\boldsymbol{w}^*\right\|_H^2+\frac{1}{2}\left(\frac{1}{\tau}\left\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\right\|_D^2+\frac{\tau-1}{\tau}\beta\left\|B(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\right\|^2\right)\\ &\leq \left\|\boldsymbol{w}^k-\boldsymbol{w}^*\right\|_H^2+\frac{1}{2}\left(\frac{1}{\tau}\left\|\boldsymbol{y}^{k-1}-\boldsymbol{y}^k\right\|_D^2+\frac{\tau-1}{\tau}\beta\left\|B(\boldsymbol{y}^{k-1}-\boldsymbol{y}^k)\right\|^2\right)\\ &\leq \left\|\boldsymbol{w}^1-\boldsymbol{w}^*\right\|_H^2+\frac{1}{2}\left(\frac{1}{\tau}\left\|\boldsymbol{y}^0-\boldsymbol{y}^1\right\|_D^2+\frac{\tau-1}{\tau}\beta\left\|B(\boldsymbol{y}^0-\boldsymbol{y}^1)\right\|^2\right),\end{aligned}$$

which implies that the sequence $\{\boldsymbol{w}^k\}$ is bounded. Due to the nonsingularity of M_k, we conclude from (A.8) that the sequence $\{\widehat{\boldsymbol{w}}^k\}$ is also bounded. Let $\boldsymbol{w}^\infty$ be a cluster point of $\{\widehat{\boldsymbol{w}}^k\}$, and let $\{\widehat{\boldsymbol{w}}^{k_j}\}$ be the subsequence converging to $\boldsymbol{w}^\infty$. Then, taking $k_j\to\infty$ in (A.20), together with (A.34), we have

$$\theta(\boldsymbol{u})-\theta(\boldsymbol{u}^\infty)+(\boldsymbol{w}-\boldsymbol{w}^\infty)^\top F(\boldsymbol{w}^\infty)\geq 0,\ \ \forall \boldsymbol{w}\in\mathscr{W}, \tag{A.35}$$

which shows that $\boldsymbol{w}^\infty$ is a solution point of (A.5), i.e., $\boldsymbol{w}^\infty\in\mathscr{W}^*$. It follows from (A.35) that

$$\|\boldsymbol{w}^{k+1}-\boldsymbol{w}^\infty\|_H^2\leq\|\boldsymbol{w}^1-\boldsymbol{w}^\infty\|_H^2+\frac{1}{2}\left(\frac{1}{\tau}\|\boldsymbol{y}^0-\boldsymbol{y}^1\|_D^2+\frac{\tau-1}{\tau}\beta\|B(\boldsymbol{y}^0-\boldsymbol{y}^1)\|^2\right).$$

Note that $\boldsymbol{w}^\infty$ is the limit point of $\{\boldsymbol{w}^{k_j}\}$. Together with (A.34), this fact means that the sequence $\{\boldsymbol{w}^k\}$ has at most one cluster point. Hence, the sequence $\{\boldsymbol{w}^k\}$ converges to $\boldsymbol{w}^\infty$ and the proof is complete. □

Hereafter, we aim to show that LR-ADMM has the worst-case $\mathscr{O}(1/t)$ convergence rate measured by the iteration complexity, where t is the iteration counter. We first recall an equivalent characterization of the solution set $\mathscr{W}^*$ of (A.5).

Proposition A.2 ([12, Theorem 2.1]). *The solution set $\mathscr{W}^*$ of (A.5) can be characterized as*

$$\mathscr{W}^* = \bigcap_{\boldsymbol{w}\in\mathscr{W}} \left\{ \widehat{\boldsymbol{w}} \in \mathscr{W} : \theta(\boldsymbol{u}) - \theta(\widehat{\boldsymbol{u}}) + (\boldsymbol{w}-\widehat{\boldsymbol{w}})^\top F(\boldsymbol{w}) \geq 0 \right\}.$$

As a preparation, we below elaborate the definition of ε-approximate solution of VI$(\mathscr{W},F,\theta)$. Concretely, we say that $\widehat{\boldsymbol{w}} \in \mathscr{W}$ is an ε-approximate solution of (A.5) if it satisfies

$$\theta(\widehat{\boldsymbol{u}}) - \theta(\boldsymbol{u}) + (\widehat{\boldsymbol{w}}-\boldsymbol{w})^\top F(\boldsymbol{w}) \leq \varepsilon, \ \ \forall \boldsymbol{w} \in \mathscr{D}_{(\widehat{\boldsymbol{w}})},$$

where $\mathscr{D}_{(\widehat{\boldsymbol{w}})} = \{\boldsymbol{w} \in \mathscr{W} : \|\boldsymbol{w}-\widehat{\boldsymbol{w}}\| \leq 1\}$. The following theorem shows that LR-ADMM (A.2) has an $\mathscr{O}(1/t)$ convergence rate.

Theorem A.3. *Suppose that $\gamma_1 \in \left(0, 1/\rho(A^\top A)\right]$, $\gamma_2 \in \left(0, 1/\rho(B^\top B)\right]$, $\tau \in (1,4/3)$, and $\delta = 2\left(\frac{1}{\tau}-\frac{3}{4}\right)$. Let $\{\boldsymbol{w}^k\}$ be the sequence generated by LR-ADMM (A.2). Then,*

$$\begin{aligned}
&\theta(\widehat{\boldsymbol{u}}_t) - \theta(\boldsymbol{u}) + (\widehat{\boldsymbol{w}}_t-\boldsymbol{w})^\top F(\boldsymbol{w}) \\
\leq &\frac{1}{2t}\left(\|\boldsymbol{w}-\boldsymbol{w}^1\|_H^2 + \frac{1}{2}\left(\frac{1}{\tau}\|\boldsymbol{y}^0-\boldsymbol{y}^1\|^2 + \frac{\tau-1}{\tau}\beta\|B(\boldsymbol{y}^0-\boldsymbol{y}^1)\|^2 \right)\right),
\end{aligned} \tag{A.36}$$

where

$$\widehat{\boldsymbol{u}}_t = \frac{1}{t}\sum_{k=1}^t \widehat{\boldsymbol{u}}^k \ \ \text{and} \ \ \widehat{\boldsymbol{w}}_t = \frac{1}{t}\sum_{k=1}^t \widehat{\boldsymbol{w}}^k. \tag{A.37}$$

Proof. It first follows from (A.30) that

$$\begin{aligned}
&\theta(\widehat{\boldsymbol{u}}^k) - \theta(\boldsymbol{u}) + (\widehat{\boldsymbol{w}}^k-\boldsymbol{w})^\top F(\boldsymbol{w}) \\
\leq &\left(\frac{1}{2}\left\|\boldsymbol{w}-\boldsymbol{w}^k\right\|_H^2 + \frac{1}{4}\left(\frac{1}{\tau}\left\|\boldsymbol{y}^{k-1}-\boldsymbol{y}^k\right\|_D^2 + \frac{\tau-1}{\tau}\beta\left\|B(\boldsymbol{y}^{k-1}-\boldsymbol{y}^k)\right\|^2 \right)\right) \\
&- \left(\frac{1}{2}\left\|\boldsymbol{w}-\boldsymbol{w}^{k+1}\right\|_H^2 + \frac{1}{4}\left(\frac{1}{\tau}\left\|\boldsymbol{y}^{k}-\boldsymbol{y}^{k+1}\right\|_D^2 + \frac{\tau-1}{\tau}\beta\left\|B(\boldsymbol{y}^{k}-\boldsymbol{y}^{k+1})\right\|^2 \right)\right).
\end{aligned}$$

Summing the above inequality over $k = 1,2,...,t$ leads to

$$\begin{aligned}
&\sum_{k=1}^t \theta(\widehat{\boldsymbol{u}}^k) - t\theta(\boldsymbol{u}) + \left(\sum_{k=1}^t \widehat{\boldsymbol{w}}^k - t\boldsymbol{w} \right)^\top F(\boldsymbol{w}) \\
\leq &\left(\frac{1}{2}\|\boldsymbol{w}-\boldsymbol{w}^1\|_H^2 + \frac{1}{4}\left(\frac{1}{\tau}\|\boldsymbol{y}^0-\boldsymbol{y}^1\|_D^2 + \frac{\tau-1}{\tau}\beta\|B(\boldsymbol{y}^0-\boldsymbol{y}^1)\|^2 \right)\right).
\end{aligned} \tag{A.38}$$

Note that $\theta(\boldsymbol{u})$ is convex. Consequently, it follows from the definitions of $\widehat{\boldsymbol{u}}_t$ and $\widehat{\boldsymbol{w}}_t$ in (A.37) that

$$t\theta(\widehat{\boldsymbol{u}}_t) \leq \sum_{k=1}^t \theta(\widehat{\boldsymbol{u}}^k) \qquad \text{and} \qquad t\widehat{\boldsymbol{w}}_t = \sum_{k=1}^t \widehat{\boldsymbol{w}}^k.$$

By substituting the above two inequalities into (A.38), we obtain the inequality (A.36). Then for a given compact set $\mathscr{D}_{(\widehat{\boldsymbol{w}})} \subset \mathscr{W}$, after t iterations of LR-ADMM (A.2), the point $\widehat{\boldsymbol{w}}_t$ defined in (A.37) satisfies

$$\sup_{w\in\mathscr{D}_{(\widehat{\boldsymbol{w}})}} \left\{ \theta(\widehat{\boldsymbol{u}}_t) - \theta(\boldsymbol{u}) + (\widehat{\boldsymbol{w}}_t-\boldsymbol{w})^\top F(\boldsymbol{w}) \right\} \leq \frac{\kappa}{2t} = \mathscr{O}(1/t),$$

where

$$\kappa := \sup\left\{ \left\| \boldsymbol{w} - \boldsymbol{w}^1 \right\|_H^2 + + \frac{1}{2}\left(\tau \left\| \boldsymbol{y}^0 - \boldsymbol{y}^1 \right\|^2 + (1-\tau)\beta \left\| B(\boldsymbol{y}^0 - \boldsymbol{y}^1) \right\|^2 \right) | \boldsymbol{w} \in \mathscr{D}_{(\widehat{\boldsymbol{w}})} \right\}.$$

Therefore, $\widehat{\boldsymbol{w}}_t$ is an approximate solution of (A.5) with an accuracy of $\mathscr{O}(1/t)$, which means that LR-ADMM (A.2) has the worst-case $\mathscr{O}(1/t)$ convergence rate in the ergodic sense. □

Acknowledgments
This research was supported in part by Ningbo Natural Science Foundation (No. 2023J014), Zhejiang Provincial Natural Science Foundation of China (No. LZ24A010001), and National Natural Science Foundation of China (No. 12371303).

References

[1] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (2009), 183-202.
[2] S. Becker, J. Bobin, E. Candés, NESTA: A fast and accurate first-order method for sparse recovery, SIAM J. Imaging Sci. 4 (2011), 1-39.
[3] E. Candés, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2006), 489-509.
[4] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (2001), 129-159.
[5] D. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (2006), 1289-1306.
[6] J. Eckstein, D. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, Math. Program. 55 (1992), 293-318.
[7] B. Efron, T. Hastie, I.Johnstone, R.Tibshirani, Least angle regression, Ann. Stat. 32 (2004), 407–451.
[8] M. Figueiredo, R. Nowak, S. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, IEEE J. Sel. Top. Sign. Proces. 1 (2007), 586-598.
[9] D. Gabay, B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximations, Comput. Math. Appl. 2 (1976), 16-40.
[10] R. Glowinski, A. Marrocco, Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problèmes non linéaires, R.A.I.R.O. R2 (1975), 41-76.
[11] B. He, F. Ma, X. Yuan, Optimally linearizing the alternating direction method of multipliers for convex programming, Comput. Optim. Appl. 75 (2020), 361–388.
[12] B. He, X. Yuan, On the $O(1/n)$ convergence rate of Douglas-Rachford alternating direction method, SIAM J. Numer. Anal. 50 (2012), 700-709.
[13] B. He, X. Yuan, On nonergodic convergence rate of Douglas-Rachford alternating direction method of multipliers, Numer. Math. 130 (2015), 567-577.
[14] D. Lorenz, M. Pfetsch, A. Tillmann, Solving basis pursuit: heuristic optimality check and solver comparison, ACM Trans. Math. Softw. 41 (2015), Article No. 8.

[15] J. Nocedal, S. Wright, Numerical Optimization, 2nd ed., Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.
[16] M. Osborne, B. Presnell, B. Turlach, A new approach to variable selection in least squares problems, IMA J. Numer. Anal. 20 (2000), 389-403.
[17] M. Osborne, B. Presnell, B. Turlach, On the LASSO and its dual, J. Comput. Graph. Statist. 9 (2000), 319-337.
[18] E. van den Berg, M. Friedlander, Probing the pareto frontier for basis pursuit solutions, SIAM J. Sci. Comput. 31 (2008), 890-912.
[19] Z. Wen, W. Yin, Y. Zhang, A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization and continuation, SIAM J. Sci. Comput. 32 (2010), 1832-1857.
[20] J. Yang, X. Yuan, Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization, Math. Comput. 82 (2013), 301-329.
[21] J. Yang, Y. Zhang, Alternating direction algorithms for ℓ_1-problems in compressive sensing, SIAM J. Sci. Comput. 332 (2011), 250-278.
[22] Z. Zhu, F. Chen, J. Zhang, Z. Wen, A unified primal-dual algorithm framework for inequality constrained problems, J. Sci. Comput. 97 (2023), Article No. 39.