An Extended Alternating Direction Method for Three-Block Separable Convex Programming *

Jianchao Bai† Jicheng Li† Hongchao Zhang‡

1School of Mathematics and Statistics, Xi’an Jiaotong University, Xi’an 710049, China
2Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803-4918, USA

Abstract
In order to solve the convex minimization problem with at least two variables, the augmented La-
grangian method is often used and improved to the alternating direction method of multipliers (ADMM) in the Gauss-Seidel or Jacobian fashion. Though various versions of the ADMM were developed for solving the two-block separable convex problem with linear equality constraint, there are a few feasible ways to tackling such problem with three-block objectives. In this paper, we design a novel ADMM combining with these two manners, where the first two subproblems are solved in parallel with some positive-definite proximal terms and the step sizes of updating the Lagrangian multipliers are enlarged. The variational inequality is applied to characterize the solution set of the concerned problem, and the Cauchy-Schwarz inequality is used to analyze some properties of the sequence \( \{w^k - \tilde{w}^k\} \) in a weighted norm, where \( w^k \) and \( \tilde{w}^k \) are respectively the iterative variable and correcting variable. In analyzing the convergence of the proposed method, the lower bound of \( \|w^k - \tilde{w}^k\|_2 \) is discussed separately in several different cases. Then based on the obtained lower bound, the global convergence of the proposed method is proved and the worst-case \( O(1/t) \) convergence rate in an ergodic sense is established.

Keywords Convex programming; Alternating direction method of multipliers; Variational inequality; Cauchy-Schwarz inequality; Complexity

AMS subject classifications 65K10; 90C25

1 Introduction
Throughout this paper, denoted by the symbols \( \mathbb{R}, \mathbb{R}^n, \mathbb{R}^{m \times n} \) be the set of real numbers, the set of \( n \) real column vectors and the set of \( m \times n \) real matrices, respectively. The notation \( \|x\| = \sqrt{\langle x, x \rangle} \) denotes the Euclidean norm of \( x \in \mathbb{R}^n \), which is induced by the inner product \( \langle x, y \rangle = x^T y \) for any vectors \( x, y \) of the same dimension. The symbol \( I_{m \times n} \) stands for the \( m \times n \) matrix whose diagonal entries are one and the others are zero (especially, we denote it as \( I_n \) if \( m = n \)).

We consider the following separable convex programming with three-block objectives, that is,

\[
\begin{align*}
\min & \quad f_1(x_1) + f_2(x_2) + f_3(x_3) \\
\text{s.t.} & \quad A_1x_1 + A_2x_2 + A_3x_3 = b, \\
& \quad x_i \in \mathcal{X}_i, \ i = 1, 2, 3,
\end{align*}
\tag{1.1}
\]

*The work was supported by the National Natural Science Foundation of China(No.11671318) and the Natural Science Foundation of Fujian province(No.2016J01028).
†Corresponding author. E-mail addresses:bjc1987@163.com
where $f_i(x_i) : \mathcal{R}^{n_i} \rightarrow \mathcal{R}(i = 1, 2, 3)$ are closed and proper convex functions (not necessarily smooth); $A_i \in \mathcal{R}^{n_i \times m_i}$ and $b \in \mathcal{R}^n$ are given matrices and vectors, respectively; $\mathcal{X}_i \subset \mathcal{R}^{m_i}(i = 1, 2, 3)$ are closed convex sets. Throughout this article, the solution set of (1.1) is assumed to be nonempty.

The problem (1.1) arises in many applications, such as the sparse inverse covariance estimation problem [11] in finance, the model updating problem [2] in the design of vibration structural dynamic system and bridges, the low rank and sparse representations [13] for image processing and so on. Intuitively, it seems that the problem (1.1) can be solved by the augmented Lagrangian method (ALM, [5]) which minimizes the following augmented Lagrangian function

$$\mathcal{L}_\beta (x_1, x_2, x_3, \lambda) = L(x_1, x_2, x_3, \lambda) + \frac{\beta}{2} \| A_1 x_1 + A_2 x_2 + A_3 x_3 - b \|^2,$$

where $\beta > 0$ is a penalty parameter with respect to the equality constraint and

$$L(x_1, x_2, x_3, \lambda) = f_1(x_1) + f_2(x_2) + f_3(x_3) - \langle \lambda, A_1 x_1 + A_2 x_2 + A_3 x_3 - b \rangle$$

is the Lagrangian function of the problem (1.1) with a Lagrangian multiplier $\lambda \in \mathcal{R}^n$. That is, one may obey the following ALM iterations

$$\begin{cases} (x_1^{k+1}, x_2^{k+1}, x_3^{k+1}) = \arg \min \{ \mathcal{L}_\beta (x_1, x_2, x_3, \lambda) \mid x_i \in \mathcal{X}_i, i = 1, 2, 3 \}, \\ \lambda^{k+1} = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b). \end{cases}$$

In practice, using the ALM still makes it difficult to simultaneously tackle the subproblem of (1.3) with three variables, especially in the case that the objective function is nondifferentiable or more complex. A natural idea to overcome this difficulty is to apply the famous alternating direction method of multipliers (ADMM), which was originally proposed in [4] and can be regarded as a splitting version of the ALM with respect to each variable at a time while fixing other two at the last values, and then update the Lagrange multiplier. In other words, one may follow the following ADMM scheme with Gauss-Seidel decomposition

$$\begin{cases} x_1^{k+1} = \arg \min \{ \mathcal{L}_\beta (x_1, x_2^k, x_3^k, \lambda^k) \mid x_1 \in \mathcal{X}_1 \}, \\ x_2^{k+1} = \arg \min \{ \mathcal{L}_\beta (x_1^{k+1}, x_2, x_3^k, \lambda^k) \mid x_2 \in \mathcal{X}_2 \}, \\ x_3^{k+1} = \arg \min \{ \mathcal{L}_\beta (x_1^{k+1}, x_2^{k+1}, x_3, \lambda^k) \mid x_3 \in \mathcal{X}_3 \}, \\ \lambda^{k+1} = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b). \end{cases}$$

The scheme (1.4a) is actually a serial algorithm which uses the latest information of each variable in each iteration. Though the above technique was proved to be convergent for the two-block separable convex minimization problem (see [6]), as said by Chen et al.,[1], the direct extension of ADMM (1.4a) for the problem (1.1) was not necessarily convergent only if some certain relationships among the matrices $A_i(i = 1, 2, 3)$ were satisfied. Another natural idea is to apply the parallel procedure in the Jacobian fashion to deal with the subproblem of (1.3), that is,

$$\begin{cases} x_1^{k+1} = \arg \min \{ \mathcal{L}_\beta (x_1, x_2^k, x_3^k, \lambda^k) \mid x_1 \in \mathcal{X}_1 \}, \\ x_2^{k+1} = \arg \min \{ \mathcal{L}_\beta (x_1^k, x_2, x_3^k, \lambda^k) \mid x_2 \in \mathcal{X}_2 \}, \\ x_3^{k+1} = \arg \min \{ \mathcal{L}_\beta (x_1^k, x_2^k, x_3, \lambda^k) \mid x_3 \in \mathcal{X}_3 \}, \\ \lambda^{k+1} = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b). \end{cases}$$

However, as shown in [1, 8], neither of the serial algorithm (1.4a) nor the changed parallel algorithm (1.4b) is necessarily convergent. The divergence of the schemes (1.4a)-(1.4b) makes us to pose a question: why not allow some of the subproblems of (1.4) to be solved in a parallel or serial manner? Inspired by such question, He et al.[9] developed a novel ADMM-like splitting method that allowed some of its subproblems to be solved in parallel, while these subproblems should be regularized by some positive-definite proximal terms to ensure the convergence. And in [9], some sparse and low-rank models and image painting problems were tested to verify the efficiency of the method. Recently, a very
meaningful and systematical work of He et al. [10] for solving the two-block separable convex minimization strongly attracts our attentions, where the iterative framework of the method therein is

\[
\begin{align*}
x_1^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta \left( x_1, x_2^k, \lambda^k \right) \mid x_1 \in \mathcal{X}_1 \right\}, \\
\lambda^{k+\frac{1}{2}} &= \lambda^k - r \beta \left( A_1 x_1^{k+1} + A_2 x_2^k - b \right), \\
x_2^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta \left( x_1^{k+1}, x_2, \lambda^{k+\frac{1}{2}} \right) \mid x_2 \in \mathcal{X}_2 \right\}, \\
\lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - s \beta \left( A_1 x_1^{k+1} + A_2 x_2^{k+1} - b \right),
\end{align*}
\]

and the step sizes \( r \) and \( s \) are restricted into the domain

\[
\mathcal{H} = \left\{ \left( r, s \right) \mid s \in \left( 0, \frac{1+\sqrt{5}}{2} \right), \ r \in (-1, 1), \ r + s > 0, |r| < 1 + s - s^2 \right\}.
\]

The main improvement of [10] is that the proposed scheme (1.5) largely extends the range of the step sizes \( r, s \) and it can be regarded as an extended work of [7] with \( r = s \in (0, 1) \). What’s more important, numerical examples about the basis pursuit problem and the total-variational image debarring model were tested to verify the efficiency of (1.5) in CPU time and the number of iterations, compared with the original ADMM in [4]. Now, another question appears: **can the idea of the scheme (1.5) be extended to solve the problem (1.1)?**

Motivated by the aforementioned two questions, we next design an improved ADMM scheme with partial proximal regularization for solving the problem (1.1), and the framework is described as follows:

\[
\begin{align*}
x_1^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta \left( x_1, x_2^k, x_3^k, \lambda^k \right) + \frac{\sigma_1}{2} \left\| A_1 (x_1 - x_1^k) \right\|^2 \mid x_1 \in \mathcal{X}_1 \right\}, \\
x_2^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta \left( x_1^k, x_2, x_3, \lambda^{k+\frac{1}{2}} \right) + \frac{\sigma_2}{2} \left\| A_2 (x_2 - x_2^k) \right\|^2 \mid x_2 \in \mathcal{X}_2 \right\}, \\
\lambda^{k+\frac{1}{2}} &= \lambda^k - r \beta \left( A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b \right), \\
x_3^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta \left( x_1^{k+1}, x_2^{k+1}, x_3, \lambda^{k+\frac{1}{2}} \right) \mid x_3 \in \mathcal{X}_3 \right\}, \\
\lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - s \beta \left( A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b \right),
\end{align*}
\]

where \( \sigma, \tau, s \) are three independent constants that are respectively restricted into

\[
\mathcal{K} = \left\{ (\tau, s) \mid \tau + s > 0, \tau < 1, |\tau| < 1 + s - s^2 \right\}.
\]

The above parameter \( \sigma \) is usually called a proximal coefficient that controls the proximity of the new iterative value to the last one. Clearly, by skipping some constants in the objective functions of the three subproblems, the ADMM (1.6) can be further simplified as

\[
\begin{align*}
x_1^{k+1} &= \arg \min \left\{ f_1(x_1) - \langle \lambda^k, A_1 x_1 \rangle + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k + A_3 x_3^k - b \right\|^2 + \sigma \left\| A_1 (x_1 - x_1^k) \right\|^2 \right\}, \\
x_2^{k+1} &= \arg \min \left\{ f_2(x_2) - \langle \lambda^k, A_2 x_2 \rangle + \frac{\beta}{2} \left\| A_1 x_1^k + A_2 x_2 + A_3 x_3^k - b \right\|^2 + \sigma \left\| A_2 (x_2 - x_2^k) \right\|^2 \right\}, \\
\lambda^{k+\frac{1}{2}} &= \lambda^k - r \beta \left( A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b \right), \\
x_3^{k+1} &= \arg \min \left\{ f_3(x_3) - \langle \lambda^{k+\frac{1}{2}}, A_3 x_3 \rangle + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3 - b \right\|^2 \right\}, \\
\lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - s \beta \left( A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b \right).
\end{align*}
\]

There are three main contributions of the paper. One is that we develop a new ADMM (1.8) for solving the three-block separable convex programming problem, and the domain \( \mathcal{H} \) in (1.5b) of the step sizes is enlarged to the domain \( \mathcal{K} \) defined in (1.7) (see the shaded area in Fig.1). Another contribution is that the global convergence property and the worst-case \( O(1/t) \) convergence rate of (1.8) are proved in detail, where the variational inequality and the Cauchy-Schwarz inequality play a significant role in characterizing the solution set of (1.1) and estimating the lower bound of the sequence \( \{w^k - \tilde{w}^k\} \) in a...
This paper is organized as follows. In Section 2, some preliminaries are prepared to reformulate the problem (1.1), including the variational inequality to characterize the solution set of the problem (1.1) and a prediction-correction procedure to interpret the ADMM (1.8). We still prove that the ADMM (1.8) is equivalent to another ADMM version. In Section 3, we first investigate some properties of $\|w^k - w^*\|^2_H$, then we provide the lower bound of $\|w^k - \bar{w}^k\|^2_G$ in different cases, and later we prove that the proposed method is globally convergent in a worst-case $O(1/t)$ convergence rate in an ergodic sense. Section 4 extends the ADMM (1.8) to the multi-block separable convex minimization problem. Finally, the paper is concluded and discussed in Section 5.

2 Preliminaries

In this section, we first uses the variational inequality to characterize the solution set of the problem (1.1). Then we analyze that the ADMM (1.8) can be regarded as a prediction-correction procedure which is split into two steps: the prediction step and the correction step. At the end of this section, we put forward another ADMM scheme and prove that it is equivalent to (1.8).

2.1 Variational reformulation of (1.1)

By the aid of the variational inequality, in this subsection, a unified framework is introduced to characterize the solution set of (1.1). We begin with a lemma that is used throughout the Section 2.

**Lemma 2.1** [10] Let $f(x) : \mathcal{R}^m \rightarrow \mathcal{R}$ and $h(x) : \mathcal{R}^m \rightarrow \mathcal{R}$ be two convex functions defined onto a closed convex set $\Omega \subset \mathcal{R}^m$. In addition, $h(x)$ is differentiable. Assume that the solution set of the
minimization problem \( \min\{f(x) + h(x)\} \) is nonempty. Then we have

\[
x^* = \arg \min_{x \in \Omega} \{f(x) + h(x)\} \iff x^* \in \Omega, f(x) - f(x^*) + (x - x^*, \nabla h(x^*)) \geq 0, \forall x \in \Omega.
\] (2.1)

It is well-known that a quadruple \((x_1^*, x_2^*, x_3^*, \lambda^*)\) is called the saddle point of the Lagrangian function (1.2b) if it satisfies the following saddle-point inequalities

\[ L(x_1^*, x_2^*, x_3^*, \lambda^*) \leq L(x_1^*, x_2^*, x_3^*, \lambda^*) \leq L(x_1^*, x_2^*, x_3^*, \lambda^*), \]

which makes us to solve the following system

\[
\begin{aligned}
x_1^* &= \arg \min \{L(x_1, x_2^*, x_3^*, \lambda) \mid x_1 \in \mathcal{X}_1\}, \\
x_2^* &= \arg \min \{L(x_1^*, x_2, x_3^*, \lambda) \mid x_2 \in \mathcal{X}_2\}, \\
x_3^* &= \arg \min \{L(x_1^*, x_2^*, x_3, \lambda) \mid x_3 \in \mathcal{X}_3\}, \\
\lambda^* &= \arg \max \{L(x_1^*, x_2^*, x_3^*, \lambda) \mid \lambda \in \mathbb{R}^n\}.
\end{aligned}
\] (2.2)

By making use of (2.1), the system (2.2) can be equivalently expressed as

\[
\begin{aligned}
x_1^* &\in \mathcal{X}_1, \quad f_1(x_1) - f_1(x_1^*) + (x_1 - x_1^*, -A_1^T \lambda^*) \geq 0, \quad \forall x_1 \in \mathcal{X}_1, \\
x_2^* &\in \mathcal{X}_2, \quad f_2(x_2) - f_2(x_2^*) + (x_2 - x_2^*, -A_2^T \lambda^*) \geq 0, \quad \forall x_2 \in \mathcal{X}_2, \\
x_3^* &\in \mathcal{X}_3, \quad f_3(x_3) - f_3(x_3^*) + (x_3 - x_3^*, -A_3^T \lambda^*) \geq 0, \quad \forall x_3 \in \mathcal{X}_3, \\
\lambda^* &\in \mathbb{R}^n, \quad \langle \lambda - \lambda^*, A_1(x_1^* - x_1^*) + A_2(x_2^* - x_2^*) + A_3(x_3^* - x_3^*) - b \rangle \geq 0, \quad \forall \lambda \in \mathbb{R}^n,
\end{aligned}
\]

which is further rewritten as a compact variational inequality (VI) form

\[
\text{VI}(f, \mathcal{J}, \mathcal{M}) : f(x) - f(x^*) + \langle w - w^*, \mathcal{J}(w^*) \rangle \geq 0, \quad \forall w \in \mathcal{M},
\] (3.2a)

with

\[
f(x) = f_1(x_1) + f_2(x_2) + f_3(x_3), \quad \mathcal{M} = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathbb{R}^n,
\] (3.2b)

\[
x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda \end{pmatrix}, \quad w = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda \end{pmatrix}, \quad \mathcal{J}(w) = \begin{pmatrix} -A_1^T \lambda \\ -A_2^T \lambda \\ -A_3^T \lambda \\ A_1 x_2 + A_2 x_2 + A_3 x_3 - b \end{pmatrix}.
\] (3.2c)

Notice that the mapping \(\mathcal{J}(w)\) in (2.3c) is monotone because it is affine with a skew-symmetric matrix. Therefore, the following basic inequality holds

\[
\langle w - \hat{w}, \mathcal{J}(w) - \mathcal{J}(\hat{w}) \rangle = 0, \quad \forall w, \hat{w} \in \mathcal{M}.
\] (4.2)

By the nonempty assumption of the solution set of (1.1), the solution set of VI\((f, \mathcal{J}, \mathcal{M})\) (denoted by \(\mathcal{M}^*\) for convenience) is also nonempty and convex, see e.g.[3, Theorem 2.3.5]. The next theorem presents a concise way to characterizing the set \(\mathcal{M}^*\), whose proof is the same as that of Theorem 2.1 [6] and is omitted here.

\textbf{Theorem 2.1} The solution set of VI\((f, \mathcal{J}, \mathcal{M})\) is convex and can be expressed as

\[
\mathcal{M}^* = \bigcap_{\hat{w} \in \mathcal{M}} \{\hat{w} \in \mathcal{M} \mid f(x) - f(\bar{x}) + \langle w - \hat{w}, \mathcal{J}(w) \rangle \geq 0\}.
\] (5.2)

\textbf{Remark 2.1} The above theorem tells us that if

\[
\sup_{w \in \mathcal{K}(\hat{w})} \{f(\bar{x}) - f(x) + \langle \hat{w} - w, \mathcal{J}(w) \rangle \} \leq \epsilon,
\]

then \(\hat{w} \in \mathcal{M}\) is called an \(\epsilon\)-approximate solution point of VI\((f, \mathcal{J}, \mathcal{M})\), where \(\mathcal{K}(\hat{w}) = \{w \in \mathcal{M} \mid \|w - \hat{w}\| \leq 1\}\) and \(\epsilon > 0\) is an accuracy, especially, \(\epsilon = O(1/t)\).
2.2 A prediction-correction interpretation of (1.8)

Now, we give a prediction-correction procedure to interpret the ADMM scheme (1.8), which aims only to provide a convenience for the theoretical analysis of the proposed method. And it is not necessary to take this prediction-correction interpretation to carry out the scheme (1.8) in practice.

For the triple \((x_1^{k+1}, x_2^{k+1}, x_3^{k+1})\) generated by the ADMM (1.8), we define

\[
w^k = \begin{pmatrix} x_1^k \\ x_2^k \\ x_3^k \\ \lambda^k \end{pmatrix}, \quad \tilde{w}^k = \begin{pmatrix} \tilde{x}_1^k \\ \tilde{x}_2^k \\ \tilde{x}_3^k \\ \tilde{\lambda}^k \end{pmatrix},
\]

where

\[
x_1^k = x_1^{k+1}, \quad x_2^k = x_2^{k+1}, \quad x_3^k = x_3^{k+1},
\]

\[
\tilde{\lambda}^k = \lambda^k - \beta \left( A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b \right).
\]

The above auxiliary notation \(\tilde{w}^k\) will be often used in the sequel and make the convergence analysis of the ADMM (1.8) more easier.

The next lemma shows that the optimality condition of the \(x_1\)-subproblem of (1.8) combining with the update of the Lagrangian multiplier \(\lambda\) can be written as a mixed variational inequality.

**Lemma 2.2** For the iterates \(w^k\) and \(\tilde{w}^k\) defined in (2.6a), it holds that

\[
\tilde{w}^k \in \mathcal{M}, \quad f(x) - f(\tilde{x}^k) + \langle w - \tilde{w}^k, J(\tilde{w}^k) + Q(\tilde{w}^k - w^k) \rangle \geq 0, \quad \forall w \in \mathcal{M},
\]

where

\[
Q = \begin{bmatrix}
\sigma \beta A_1^T A_1 & -\beta A_1^T A_2 & 0 & 0 \\
-\beta A_2^T A_1 & \sigma \beta A_2^T A_2 & 0 & 0 \\
0 & 0 & \beta A_3^T A_3 & -\tau A_T^T \\
0 & 0 & -A_3 & \frac{1}{\tau} I_n
\end{bmatrix}.
\]

**Proof** By using (2.1) and \(\tilde{X}^k\) defined in (2.6c), the optimality condition of the \(x_1\)-subproblem of the ADMM (1.8) can be expressed as

\[
x_1^{k+1} \in \mathcal{X}_1, \quad f_1(x_1) - f_1(x_1^{k+1}) + \left( x_1 - x_1^{k+1}, -A_1^T \lambda^k + \beta A_1^T \left( A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b \right) + \sigma \beta A_1^T A_1 (x_1^{k+1} - x_1^k) \right) \geq 0, \quad \forall x_1 \in \mathcal{X}_1,
\]

\[
\implies f_1(x_1) - f_1(x_1^{k+1}) + \left( x_1 - x_1^{k+1}, -A_1^T \lambda^k + \beta A_1^T A_2 (\tilde{x}_2^k - x_2^k) + \sigma \beta A_1^T A_1 (\tilde{x}_1^k - x_1^k) \right) \geq 0 \quad (2.8)
\]

Similarly, for the \(x_2\)-subproblem we have

\[
x_2^{k+1} \in \mathcal{X}_2, \quad f_2(x_2) - f_2(x_2^{k+1}) + \left( x_2 - x_2^{k+1}, -A_2^T \lambda^k + \beta A_2^T \left( A_1 x_1^k + A_2 x_2^{k+1} + A_3 x_3^k - b \right) + \sigma \beta A_2^T A_2 (x_2^{k+1} - x_2^k) \right) \geq 0, \quad \forall x_2 \in \mathcal{X}_2,
\]

\[
\implies f_2(x_2) - f_2(x_2^{k+1}) + \left( x_2 - x_2^{k+1}, -A_2^T \lambda^k - \beta A_1^T A_1 (\tilde{x}_1^k - x_1^k) + \sigma \beta A_2^T A_2 (\tilde{x}_2^k - x_2^k) \right) \geq 0 \quad (2.9)
\]

Using the notation \(\tilde{X}^k\) in (2.6c), the iterate \(\lambda^{k+\frac{1}{2}}\) in (1.8) can be rewritten as

\[
\lambda^{k+\frac{1}{2}} = \lambda^k - \tau \left( \tilde{\lambda}^k - \lambda^k \right).
\]
Then by applying Lemma 2.1 to the $x_3$-subproblem together with (2.10), we obtain
\[
\begin{align*}
\lambda_{k+1}^3 &
\in X_3, \quad f_3(x_3) - f_3(x_3^{k+1}) \\
& + \left( x_3 - x_3^{k+1}, -A_3^T\lambda_{k+1}^3 + \beta A_3^T (A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b) \right) \geq 0, \forall x_3 \in X_3,
\end{align*}
\]
\[
\Rightarrow f_3(x_3) - f_3(x_3^*) + \left( x_3 - x_3^*, -A_3^T\lambda_k + \beta A_3^T A_3 (x_3^* - x_3) \right) \geq 0,
\]
\[
\Rightarrow f_3(x_3) - f_3(x_3^*) + \left( x_3 - x_3^*, -A_3^T\lambda_k + \beta A_3^T A_3 (x_3^* - x_3) - \tau A_3^T (\tilde{\lambda}_k - \lambda_k) \right) \geq 0.
\]
Besides, it follows from (2.6c) that
\[
\begin{align*}
(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b) - A_3 (x_3^{k+1} - x_3^k) &= \frac{1}{\beta} \left( \lambda_k - \tilde{\lambda}_k \right) \\
\Rightarrow (A_1\tilde{x}_1^k + A_2\tilde{x}_2^k + A_3\tilde{x}_3^k - b) - A_3 (\tilde{x}_3^k - x_3^k) &= \frac{1}{\beta} \left( \tilde{\lambda}_k - \lambda_k \right) = 0,
\end{align*}
\]
which is rewritten as
\[
\begin{align*}
\tilde{\lambda}_k &\in \mathbb{R}^n, \quad \left( \lambda - \tilde{\lambda}_k, (A_1\tilde{x}_1^k + A_2\tilde{x}_2^k + A_3\tilde{x}_3^k - b) - A_3 (\tilde{x}_3^k - x_3^k) + \frac{1}{\beta} \left( \tilde{\lambda}_k - \lambda_k \right) \right) \geq 0, \forall \lambda \in \mathbb{R}^n.
\end{align*}
\]
Combining (2.8)-(2.9) and (2.11)-(2.12), the proof of the inequality (2.7) is completed. \hfill \blacksquare

**Lemma 2.3** For the sequence $\{w^{k+1}\}$ generated by the ADMM (1.8) and the iterate $\tilde{w}^k$ defined in (2.6a), the following relationship
\[
w^{k+1} = w^k - M(w^k - \tilde{w}^k)
\]
holds with
\[
M = \begin{bmatrix}
I_{m_1} & 0 & 0 & 0 \\
0 & I_{m_2} & 0 & 0 \\
0 & 0 & I_{m_3} & 0 \\
0 & 0 & -s\beta A_3 & (\tau + s)I_n
\end{bmatrix}.
\]

**Proof** By the update of $\lambda^{k+1}$ in (1.8) and Eq.(2.10), it follows that
\[
\begin{align*}
\lambda^{k+1} &= \lambda^k + \frac{1}{\beta} - s\beta \left( A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b \right) \\
&= \lambda^k + \frac{1}{\beta} - s\beta \left( A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b \right) - s\beta A_3 \left( x_3^{k+1} - x_3^k \right) \\
&= \lambda^k - \tau \left( \lambda^k - \tilde{\lambda}_k \right) - s \left( \lambda^k - \tilde{\lambda}_k \right) - s\beta A_3 \left( x_3^k - x_3^k \right) \\
&= \lambda^{k+1} - \left[ -s\beta A_3 \left( x_3^k - x_3^k \right) + (\tau + s) \left( \lambda^k - \tilde{\lambda}_k \right) \right].
\end{align*}
\]
The above equality together with $x_i^{k+1} = \tilde{x}_i^k (i = 1, 2, 3)$ implies that
\[
\begin{bmatrix}
x_1^{k+1} \\
x_2^{k+1} \\
x_3^{k+1} \\
\lambda^{k+1}
\end{bmatrix} = \begin{bmatrix}
x_1^k \\
x_2^k \\
x_3^k \\
\lambda^k
\end{bmatrix} - \begin{bmatrix}
I_{m_1} & 0 & 0 & 0 \\
0 & I_{m_2} & 0 & 0 \\
0 & 0 & I_{m_3} & 0 \\
0 & 0 & -s\beta A_3 & (\tau + s)I_n
\end{bmatrix} \begin{bmatrix}
x_1^k - x_1^k \\
x_2^k - x_2^k \\
x_3^k - x_3^k \\
\lambda^k - \lambda^k
\end{bmatrix},
\]
that is, the relationship (2.13a) holds. \hfill \blacksquare

Lemma 2.2-2.3 show that the ADMM (1.8) can be interpreted as a prediction-correction framework as we mentioned before, where the iterates $w^{k+1}$ and $\tilde{w}^k$ are formally called the predictive variable and correcting variable, respectively.
2.3 Relationship between (1.8) and another scheme

This subsection presents an improvement of the algorithms (1.4a)-(1.4b), where each subproblem is regularized by a proximal regularization and the Lagrangian multiplier is updated once time in each iteration, that is the following scheme:

\[
\begin{align*}
\{x_1^{k+1}\} &= \arg\min_{x_1 \in X_1} \left\{ \mathcal{L}_\beta \left(x_1, x_2^k, x_3^k, \lambda^k \right) + \frac{\beta}{2} \|A_1(x_1 - x_1^k)\|^2 \right\}, \\
\{x_2^{k+1}\} &= \arg\min_{x_2 \in X_2} \left\{ \mathcal{L}_\beta \left(x_2^k, x_2, x_3^k, \lambda^k \right) + \frac{\beta}{2} \|A_2(x_2 - x_2^k)\|^2 \right\}, \\
\{x_3^{k+1}\} &= \arg\min_{x_3 \in X_3} \left\{ f_3(x_3) - \langle \lambda^k, \lambda_3x_3 \rangle + \alpha(x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b) \right\}, \\
&\quad \lambda^{k+1} = \lambda^k - \beta \left( A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b \right),
\end{align*}
\]

where \(\mu = \tau + 1, \sigma > 1\) and \(\tau, \sigma\) are independent parameters restricted to the domain \(K\) in (1.7).

Note that there are three characteristics of the above ADMM scheme: (i) the first two subproblems of (2.14) are regularized by respectively adding a proximal term, which is the same as that of (1.8); (ii) the parameter \(\mu = \tau + 1\) instead of the constant one and the \(x_3\)-subproblem is regularized by subtracting a proximal term; (iii) the step size of \(\lambda\), that is \(s\), is not the constant one but an enlarged domain. Although the scheme (2.14) is different from (1.8) on the surface, we will prove that both of them are actually equivalent.

**Lemma 2.4** The ADMM (1.8) is equivalent to the scheme (2.14) with \(\mu = \tau + 1\).

**Proof** Clearly, we only need to prove that the optimality condition of the \(x_3\)-subproblem in (1.8) are the same as that of (2.14).

By applying (2.1), the optimality condition of the \(x_3\)-subproblem in (2.14) is \(x_3^{k+1} \in X_3\) and

\[
\begin{align*}
&\quad f_3(x_3) - f_3(x_3^{k+1}) + \\
&\quad \langle x_3 - x_3^{k+1}, -A_3^T\lambda^{k+1} + \mu \beta A_3^T(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b) - \tau \beta A_3^TA_3(x_3^{k+1} - x_3^k) \rangle \geq 0, \quad \forall x_3 \in X_3.
\end{align*}
\]

Besides, using the update of \(\lambda^{k+1}\) in (1.8), the optimality condition of the \(x_3\)-subproblem in (1.8) can be further rewritten as

\[
\begin{align*}
&\quad f_3(x_3) - f_3(x_3^{k+1}) + \\
&\quad \langle x_3 - x_3^{k+1}, -A_3^T\left( \lambda^k - \tau \beta (A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b) + \beta \lambda A_3^T(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b) \right) \rangle \geq 0,
\end{align*}
\]

\[
\begin{align*}
\because f_3(x_3) - f_3(x_3^{k+1}) + \\
&\quad \langle x_3 - x_3^{k+1}, -A_3^T\lambda^k + (\tau + 1) \beta A_3^T(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b) - \tau \beta A_3^TA_3(x_3^{k+1} - x_3^k) \rangle \geq 0.
\end{align*}
\]

Obviously, (2.15) is just (2.16) by setting \(\mu = \tau + 1\). ■

**Remark 2.2** When taking \((\tau, s) = (0, 1)\) \(\in K\), then (2.14) becomes a concise scheme

\[
\begin{align*}
\{x_1^{k+1}\} &= \arg\min_{x_1 \in X_1} \left\{ \mathcal{L}_\beta \left(x_1, x_2^k, x_3^k, \lambda^k \right) + \frac{\beta}{2} \|A_1(x_1 - x_1^k)\|^2 \right\}, \\
\{x_2^{k+1}\} &= \arg\min_{x_2 \in X_2} \left\{ \mathcal{L}_\beta \left(x_2^k, x_2, x_3^k, \lambda^k \right) + \frac{\beta}{2} \|A_2(x_2 - x_2^k)\|^2 \right\}, \\
\{x_3^{k+1}\} &= \arg\min_{x_3 \in X_3} \left\{ \mathcal{L}_\beta \left(x_3^{k+1}, x_2^{k+1}, x_3, \lambda^k \right) \right\}, \\
&\quad \lambda^{k+1} = \lambda^k - \beta \left( A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b \right).
\end{align*}
\]
3 Convergence analysis of ADMM (1.8)

The key to prove the convergence of the ADMM (1.8) is to verify that the cross term of (2.7) converges to zero, that is,

$$\lim_{k \to \infty} \langle w - \tilde{w}^k, Q(w^k - \tilde{w}^k) \rangle = 0, \ \forall w \in M,$$

where $w^k$ and $\tilde{w}^k$ are defined in (2.6a), in other words, the sequence $\{w^k - w^*\}$ would be contractive in some weighted norm. In this section, we first investigate some properties of the sequence $\{\|w^k - w^*\|_H^2\}$ and estimate the lower bound of $\|w^k - \tilde{w}^k\|^2_2$. Then by using these properties, several theorems are provided to illustrate the global convergence property and the convergence rate of the ADMM (1.8).

3.1 Properties of $\{\|w^k - w^*\|_H^2\}$

We investigate whether the sequence $\{w^k - w^*\}$ is contractive or not. Using the property of $J(w)$, that is (2.4), the variational inequality (2.7a) becomes

$$\tilde{w}^k \in M, \ f(x) - f(\tilde{w}^k) + \langle w - \tilde{w}^k, J(w) \rangle \geq \langle w - \tilde{w}^k, Q(w^k - \tilde{w}^k) \rangle, \ \forall w \in M. \ (3.1)$$

Note that the matrix $M$ in (2.13b) is nonsingular. By making use of the relationship (2.13), the right-hand term of (3.1) is further changed to

$$\langle w - \tilde{w}^k, Q(w^k - \tilde{w}^k) \rangle = \langle w - \tilde{w}^k, H(w^k - w^{k+1}) \rangle. \ (3.2)$$

where

$$H = QM^{-1}. \ (3.3)$$

It is noteworthy that the above relationship (3.2) will play a fundamental role in analyzing the convergence of the proposed method in the sequel. The next lemma shows that the matrix $H$ is symmetric positive definite in the discussed domain.

Lemma 3.1 The matrix $H$ defined in (3.3) is symmetric positive definite for any $\sigma \in (1, +\infty)$ and $(\tau, s) \in K$ when the matrices $A_i (i = 1, 2, 3)$ are full column rank.

Proof For the matrix $M$ defined in (2.13b), it is easy to obtain its inverse

$$M^{-1} = \begin{bmatrix} I_{m_1} & 0 & 0 & 0 \\ 0 & I_{m_2} & 0 & 0 \\ 0 & 0 & I_{m_3} & 0 \\ 0 & 0 & \frac{\sigma \beta}{\tau + s} A_3 & \frac{1}{\tau + s} I_n \end{bmatrix}. \ (3.4a)$$

Thus by the relationship (3.3), we have

$$H = \begin{bmatrix} \sigma \beta A_1^T A_1 & -\beta A_1^T A_2 & 0 & 0 \\ -\beta A_1^T A_2 & \sigma \beta A_2^T A_2 & 0 & 0 \\ 0 & 0 & \beta A_3^T A_3 & -\tau A_3^T \\ 0 & 0 & -A_3^T & \frac{1}{\tau + s} I_n \end{bmatrix} \begin{bmatrix} I_{m_1} & 0 & 0 & 0 \\ 0 & I_{m_2} & 0 & 0 \\ 0 & 0 & I_{m_3} & 0 \\ 0 & 0 & \frac{\sigma \beta}{\tau + s} A_3 & \frac{1}{\tau + s} I_n \end{bmatrix} \begin{bmatrix} \sigma \beta A_1^T A_1 & -\beta A_1^T A_2 & 0 & 0 \\ -\beta A_1^T A_2 & \sigma \beta A_2^T A_2 & 0 & 0 \\ 0 & 0 & \beta A_3^T A_3 & -\tau A_3^T \\ 0 & 0 & -A_3^T & \frac{1}{\tau + s} A_3 \\ 0 & 0 & \beta A_3^T A_3 & -\tau A_3^T \\ 0 & 0 & -A_3^T & \frac{1}{\tau + s} A_3 \end{bmatrix}. \ (3.4b)$$

Clearly, the matrix $H$ is symmetric and can be decomposed into $H = D^T H_0 D$, where

$$D = \begin{bmatrix} A_1 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 \\ 0 & 0 & A_3 & 0 \\ 0 & 0 & 0 & I_n \end{bmatrix}, \ H_0 = \begin{bmatrix} \sigma \beta I_n & -\beta I_n & 0 & 0 \\ -\beta I_n & \sigma \beta I_n & 0 & 0 \\ 0 & 0 & (1 - \frac{\sigma \beta}{\tau + s}) I_n & -\tau I_n \\ 0 & 0 & -\tau I_n & (1 - \frac{\sigma \beta}{\tau + s}) I_n \end{bmatrix}. \ (3.4b)$$
Then, the matrix $H$ is positive definite if and only if $H_0$ is positive definite. In other words, all the ordered-main subdeterminants of $H_0$ should be positive for any given $\beta > 0$, equivalently,

$$
\begin{align*}
\sigma > 0, \\
\sigma^2 - 1 > 0, \\
(\sigma^2 - 1)(\tau + s)(\tau + s - \tau s) > 0, \\
(\sigma^2 - 1)(\tau + s)(\tau + s - \tau s - \tau^2) > 0.
\end{align*}
$$
\hspace{1cm} (3.5)

Obviously, the above obtained range of $\sigma, \tau, s$ in (3.5) has been in the domain of (1.7).

**Theorem 3.1** Let the matrices $Q, M, H$ be defined in (2.7b), (2.13b) and (3.4a), respectively. For the sequence $\{w^{k+1}\}$ generated by the ADMM (1.8) and the iterate $\tilde{w}^k$ defined in (2.6a), we have

$$
\begin{align*}
f(x) - f(\tilde{x}^k) + \langle w - \tilde{w}^k, J(w) \rangle &\geq \frac{1}{2} \{\|w - w^{k+1}\|^2_H - \|w - w^k\|^2_H + \|w^k - \tilde{w}^k\|^2_G\}, \forall w \in \mathcal{M},
\end{align*}
$$
\hspace{1cm} (3.6a)

where

$$
G = Q + Q^T - M^T H M.
$$
\hspace{1cm} (3.6b)

**Proof** Applying the identity

$$
2(a - b, H(c - d)) = \|a - d\|^2_H - \|a - c\|^2_H + \|c - b\|^2_H - \|d - b\|^2_H
$$

with substitutions $a = w$, $b = \tilde{w}^k$, $c = w^k$, $d = w^{k+1}$, we get

$$
2 \langle w - \tilde{w}^k, H(w^{k+1} - w^k) \rangle = \|w - w^{k+1}\|^2_H - \|w - w^k\|^2_H + \|w^k - \tilde{w}^k\|^2_H - \|w^{k+1} - \tilde{w}^k\|^2_H.
$$
\hspace{1cm} (3.7)

Meanwhile, for the last two terms of the right-hand side of (3.7), it follows that

$$
\begin{align*}
\|w^k - \tilde{w}^k\|^2_H - \|w^{k+1} - \tilde{w}^k\|^2_H &= \|w^k - \tilde{w}^k\|^2_H - \|w^k - w^{k+1}\|^2_H \\
&= \langle w^{k+1} - \tilde{w}^k, M(w^k - \tilde{w}^k) \rangle \\
&= \langle w^{k+1} - \tilde{w}^k, M(w^k - \tilde{w}^k) \rangle \\
&= \langle w^{k+1} - \tilde{w}^k, (Q + Q^T - M^T H M)(w^k - \tilde{w}^k) \rangle
\end{align*}
$$
\hspace{1cm} (3.8)

where the fourth equality of (3.8) uses the Eq.(3.3). Combining (3.1)-(3.2) and (3.7)-(3.8), we complete the proof of the assertion (3.6).

**Theorem 3.2** For the iterate $\tilde{w}^k$ defined in (2.6a), the sequence $\{w^{k+1}\}$ generated by the ADMM (1.8) satisfies

$$
\|w^{k+1} - w^*\|^2_H \leq \|w^k - w^*\|^2_H + \|w^k - \tilde{w}^k\|^2_G, \forall w^* \in \mathcal{M}^*.
$$
\hspace{1cm} (3.9)

**Proof** By via of taking $w = w^*$ in (3.6a), we have

$$
\frac{1}{2} \{\|w^k - w^*\|^2_H - \|w^k - \tilde{w}^k\|^2_H\} \geq \frac{1}{2} \|w^{k+1} - w^*\|^2_H + f(\tilde{x}^k) - f(x^*) + \langle \tilde{w}^k - w^*, J(w^*) \rangle.
$$

The above inequality together with (2.3a) immediately implies the inequality (3.9).

Note from (3.9) that the sequence $\{w^k - w^*\}$ is not always contractive since the matrix $G$ defined in (3.6b) is not always positive definite when the parameters $\tau$ and $s$ are restricted into $\mathcal{K}$. Hence, we need to discuss the convergence of the proposed method in different cases. For analysis convenience, let

$$
\mathcal{K} = \bigcup_{i=1}^{5} \mathcal{K}_i, \quad \mathcal{K}_i \cap \mathcal{K}_j = \emptyset, \forall i \neq j, \ i, j = 1, 2, 3, 4, 5.
$$
where \( K_i \) be partitioned in a similar way as [10], that is,

\[
K_1 = \{(\tau, s) \mid -s < \tau < 1, s < 1\},
\]

(3.10a)

\[
K_2 = \{(\tau, s) \mid -1 < \tau < 1, s = 1\},
\]

(3.10b)

\[
K_3 = \left\{(\tau, s) \mid \tau = 0, 1 < s < \frac{1 + \sqrt{5}}{2}\right\},
\]

(3.10c)

\[
K_4 = \{(\tau, s) \mid 0 < \tau < 1 + s - s^2, s > 1\},
\]

(3.10d)

\[
K_5 = \{(\tau, s) \mid s^2 - s - 1 < \tau < 0, s > 1\}.
\]

(3.10e)

### 3.2 Lower bound of \( \|w^k - \tilde{w}^k\|^2_G \) in \( K_1 \cup K_2 \)

In this subsection, we discuss some properties of \( \|w^k - \tilde{w}^k\|^2_G \) in the domain \( K_1 \cup K_2 \), where \( w_k \) is the \( k \)th iteration of the ADMM (1.8) and \( \tilde{w}^k \) is defined in (2.6a). Moreover, the lower bound of \( \|w^k - \tilde{w}^k\|^2_G \) are given in the aforementioned Theorem 3.4.

**Lemma 3.2** For the sequence \( \{w^{k+1}\} \) generated by the ADMM (1.8) and the iterate \( \tilde{w}^k \) defined in (2.6a), we have

\[
\|w^k - \tilde{w}^k\|^2_G = \sigma \beta \|A_1 (x_1^k - x_1^{k+1})\|^2 + \sigma \beta \|A_2 (x_2^k - x_2^{k+1})\|^2 + (1 - \tau) \beta \|A_3 (x_3^k - x_3^{k+1})\|^2
\]

\[
+ (2 - \tau - s) \beta \|A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b\|^2 - 2 \beta (x_1^k - x_1^{k+1})^T A_1^T A_2 (x_2^k - x_2^{k+1})
\]

\[
+ 2 (1 - \tau) \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b)^T A_3 (x_3^k - x_3^{k+1}).
\]

(3.11)

**Proof** By a simple calculation, we obtain

\[
M^T H M = M^T Q = \sigma \beta A_1^T A_1 - \beta A_1^T A_2 0 0
\]

\[
- \beta A_2^T A_1 \sigma \beta A_2^T A_2 0 0
\]

\[
0 0 (1 + s) \beta A_1^T A_3 - (\tau + s) A_3^T 0
\]

\[
0 0 - (\tau + s) A_3^T \frac{2s}{\tau s} I_n.
\]

and then

\[
G = Q + Q^T - M^T H M
\]

\[
= \begin{bmatrix}
\sigma \beta A_1^T A_1 & -\beta A_1^T A_2 & 0 & 0 \\
-\beta A_2^T A_1 & \sigma \beta A_2^T A_2 & 0 & 0 \\
0 & 0 & (1 - s) \beta A_1^T A_3 & (s - 1) A_3^T \\
0 & 0 & 0 & (s - 1) A_3^T \frac{2s}{\tau s} I_n
\end{bmatrix}.
\]

(3.12)

Hence, it holds that

\[
\|w^k - \tilde{w}^k\|^2_G = \langle w^k - \tilde{w}^k, G(w^k - \tilde{w}^k) \rangle
\]

\[
= \sigma \beta \|A_1 (x_1^k - \tilde{x}_1^k)\|^2 + \sigma \beta \|A_2 (x_2^k - \tilde{x}_2^k)\|^2 + (1 - s) \beta \|A_3 (x_3^k - \tilde{x}_3^k)\|^2
\]

\[
+ \frac{2 - \tau - s}{\beta} \|\lambda^k - \tilde{\lambda}^k\|^2 + 2 (s - 1) (\lambda^k - \tilde{\lambda}^k)^T A_3 (x_3^k - \tilde{x}_3^k)
\]

\[
- 2 \beta (x_1^k - \tilde{x}_1^k)^T A_1^T A_2 (x_2^k - \tilde{x}_2^k) - 2 \beta (x_1^k - \tilde{x}_1^k)^T A_1^T A_3 (x_3^k - \tilde{x}_3^k).
\]

(3.13)

By the notation \( \lambda^k \) in (2.6c), we have

\[
\lambda^k - \tilde{\lambda}^k = \beta \left[ (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b) + A_3 (x_3^k - x_3^{k+1}) \right].
\]

(3.14a)
Thus, the cross term

$$(\lambda^k - \lambda^*)^T A_3 (x_3^k - \bar{x}_3^k) = \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b)^T A_3 (x_3^k - \bar{x}_3^k) + \beta \| A_3 (x_3^k - \bar{x}_3^k) \|^2$$

(3.14b)

and

$$\| \lambda^k - \bar{\lambda}^k \|^2 = \beta^2 \| A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b \|^2 + \beta^2 \| A_3 (x_3^k - x_3^{k+1}) \|^2 + 2 \beta^2 (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b)^T A_3 (x_3^k - x_3^{k+1})$$

(3.14c)

Using the notations $\bar{x}_i^k = x_i^{k+1} (i = 1, 2, 3)$ in (2.6b) and substituting (3.14b)-(3.14c) into (3.13), we obtain the equality (3.11).

**Lemma 3.3** The sequence $\{ w^{k+1} \}$ generated by the ADMM (1.8) satisfies

$$(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b)^T A_3 (x_3^k - x_3^{k+1}) \geq \frac{1-\tau-s}{1+\tau} (A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b)^T A_3 (x_3^k - x_3^{k+1})$$

(3.15)

**Proof** By Lemma 2.1, the optimality condition of the $x_3$-subproblem of (1.8) is

$$x_3^{k+1} \in A_3, \quad f_3(x_3) - f_3(x_3^{k+1}) + \langle x_3 - x_3^{k+1}, -A_3^T \lambda^{k+\frac{1}{2}} + \beta A_3^T (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b) \rangle \geq 0, \quad \forall x_3 \in A_3,$$

(3.16a)

$$\Rightarrow f_3(x_3) - f_3(x_3^k) + \langle x_3 - x_3^k, -A_3^T \lambda^k - \frac{1}{2} - \beta A_3^T (A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b) \rangle \geq 0.$$  

(3.16b)

Setting $x_3 = x_3^k$ in (3.16a) and $x_3 = x_3^{k+1}$ in (3.16b), and adding them together, we have

$$\langle x_3^k - x_3^{k+1}, A_3^T (\lambda^{k+\frac{1}{2}} - \lambda^k - \frac{1}{2}) + \beta A_3^T (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b) - \beta A_3^T (A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b) \rangle \geq 0.$$  

(3.17)

From the update of $\lambda^{k+\frac{1}{2}}$ in (1.8), i.e.,

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \tau \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b)$$

and the previous iteration

$$\lambda^k = \lambda^{k-\frac{1}{2}} - s \beta (A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b),$$

we deduce that

$$\lambda^{k-\frac{1}{2}} - \lambda^{k+\frac{1}{2}} = \tau \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b) + \tau \beta A_3 (x_3^k - x_3^{k+1}) + s \beta (A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b).$$  

(3.18)

Substituting (3.18) into (3.17) with a simplification, the proof is completed.

The aforementioned Lemmas 3.2-3.3 immediately result in the following Theorem 3.3 which plays an important role in analyzing the convergence of the ADMM (1.8) in the sequel.

**Theorem 3.3** For the sequence $\{ w^{k+1} \}$ generated by the ADMM (1.8) and the iterate $\bar{w}^k$ defined in (2.6a), we have

$$\| w^k - \bar{w}^k \|_G^2 \geq \sigma \beta \| A_1 (x_1^k - x_1^{k+1}) \|^2 + \sigma \beta \| A_2 (x_2^k - x_2^{k+1}) \|^2 + \frac{(1-\tau-s)^2 \beta \| A_3 (x_3^k - x_3^{k+1}) \|^2}{1+\tau}$$

$$+ (2-\tau-s) \beta \| A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b \|^2 + \frac{(2\tau-1)(1-s) \beta (A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b)^T A_3 (x_3^k - x_3^{k+1}) - 2 \beta (x_1^k - x_1^{k+1})^T A_1^T A_2 (x_2^k - x_2^{k+1})}{2\tau}.$$  

(3.19)
Theorem 3.4 Let the sequence \( \{w^{k+1}\} \) be generated by the ADMM (1.8) and the iterate \( \tilde{w}^k \) be defined in (2.6a). For any \( \sigma > 1 \) and \((\tau, s) \in K_1 \cup K_2\), there exist constants \( \xi_i > 0 (i = 1, 2, 3, 4) \) such that
\[
\|w^k - \tilde{w}^k\|^2_G \geq \xi_1 \|A_1 (x_1^k - x_1^{k+1})\|^2 + \xi_2 \|A_2 (x_2^k - x_2^{k+1})\|^2 + \xi_3 \|A_3 (x_3^k - x_3^{k+1})\|^2 + \xi_4 \|A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b\|^2.
\] (3.20)

Proof The assertion (3.20) can be proved separately in two different cases.

(Case I) For any
\[
\sigma > 1, \ (\tau, s) \in K_1 = \{ (\tau, s) | -s < \tau < 1, s < 1 \},
\]
the matrix \( G \) defined in (3.12) can be decomposed into \( G = D^T G_0 D \), where \( D \) is defined in (3.14b) and \( G_0 \) is given by
\[
G_0 = \begin{bmatrix}
\sigma \beta I_n & -\beta I_n & 0 & 0 \\
-\beta I_n & \sigma \beta I_n & 0 & 0 \\
0 & 0 & (s-1)\beta I_n & (s-1)I_n \\
0 & 0 & (s-1)I_n & \frac{2-(\tau+s)}{\beta^2} I_n
\end{bmatrix}.
\] (3.21)

Regardless of the given positive parameter \( \beta \), it is easy to compute that all the ordered-main subdeterminants of the matrix \( G_0 \) are positive for \( \sigma > 1 \) and \((\tau, s) \in K_1 \), that is,
\[
\begin{aligned}
&\sigma > 0, \\
&\sigma^2 - 1 > 0, \\
&(\sigma^2 - 1)(1-s) > 0, \\
&(\sigma^2 - 1)(1-s)(1-\tau) > 0.
\end{aligned}
\]

Hence, both \( G \) and \( G_0 \) are positive definite. Besides, by (3.14a) and (2.6b), we have
\[
\begin{pmatrix}
A_1(x_1^k - \tilde{x}_1^k) \\
A_2(x_2^k - \tilde{x}_2^k) \\
A_3(x_3^k - \tilde{x}_3^k) \\
\lambda^k - \tilde{\lambda}_1^k
\end{pmatrix}
= \begin{bmatrix}
I_n & 0 & 0 & 0 \\
0 & I_n & 0 & 0 \\
0 & 0 & I_n & 0 \\
0 & 0 & \beta I_n & \beta I_n
\end{bmatrix}
\begin{pmatrix}
A_1(x_1^k - x_1^{k+1}) \\
A_2(x_2^k - x_2^{k+1}) \\
A_3(x_3^k - x_3^{k+1}) \\
A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b
\end{pmatrix}
\]
\[
\Rightarrow \|w^k - \tilde{w}^k\|^2_G = \left\| \begin{pmatrix}
A_1(x_1^k - x_1^{k+1}) \\
A_2(x_2^k - x_2^{k+1}) \\
A_3(x_3^k - x_3^{k+1}) \\
A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b
\end{pmatrix}_{G_0} \right\|^2,
\] (3.22)

where the matrix \( \tilde{G}_0 = L^T G_0 L \) is positive definite and the lower triangular matrix
\[
L = \begin{bmatrix}
I_n & 0 & 0 & 0 \\
0 & I_n & 0 & 0 \\
0 & 0 & I_n & 0 \\
0 & 0 & \beta I_n & \beta I_n
\end{bmatrix}
\]
is nonsingular.

Consequently, the assertion (3.20) holds followed by (3.22).

(Case II) For any
\[
\sigma > 1, \ (\tau, s) \in K_2 = \{ (\tau, s) | 1 < \tau < 1, s = 1 \},
\]
by using the Cauchy-Schwarz inequality (see e.g.[14], Page xvii), there must exists a positive number \( \eta \in (1/\sigma, \sigma) \) (especially, \( \eta = 1 \)) such that
\[
-2\beta (x_1^k - x_1^{k+1})^T A_1 A_2 (x_2^k - x_2^{k+1}) \geq -\eta \beta \|A_1 (x_1^k - x_1^{k+1})\|^2 - \frac{\beta}{\eta} \|A_2 (x_2^k - x_2^{k+1})\|^2.
\] (3.23)
Setting \( s = 1 \) in (3.19) together with the above inequality, we get

\[
\|w^k - \tilde{w}^k\|_G^2 \geq (\sigma - \eta)\beta \|A_1 (x_1^k - x_1^k+1)\|^2 + (\sigma - \frac{1}{\eta})\beta \|A_2 (x_2^k - x_2^k+1)\|^2 + \frac{(1-\tau)^2}{1+\tau} \beta \|A_3 (x_3^k - x_3^k+1)\|^2 \\
+ (1-\tau)\beta \|A_1 x_1^k+1 + A_2 x_2^k+1 + A_3 x_3^k+1 - b\|^2.
\]

That is, the assertion (3.20) holds with

\[
\begin{align*}
\xi_1 &= (\sigma - \eta)\beta > 0, \\
\xi_2 &= (\sigma - 1/\eta)\beta > 0, \\
\xi_3 &= (1-\tau)^2 2 \beta > 0, \\
\xi_4 &= (1-\tau)\beta > 0.
\end{align*}
\]

**Remark 3.1** When taking \( \tau = s \) and restricting \( s \in (0, 1) \subset \mathcal{K}_1 \), the ADMM (1.8) can be regarded as an extension of the symmetric ADMM scheme in [7] that considers the two-block separable convex programming. Also, it is an extension of the recent work of He et al.[10], but the problem and the range of the parameters are different in this paper.

### 3.3 Lower bound of \( \|w^k - \tilde{w}^k\|_G^2 \) in \( \mathcal{K}_3 \cup \mathcal{K}_4 \cup \mathcal{K}_5 \)

We first give a lower bound of \( \|w^k - \tilde{w}^k\|_G^2 \) described in the following theorem.

**Theorem 3.5** Let the sequence \( \{w^{k+1}\} \) be generated by the ADMM (1.8) and the iterate \( \tilde{w}^k \) be defined in (2.6a). For any \( \sigma > 1 \) and \( (\tau, s) \in \mathcal{K}_3 \cup \mathcal{K}_4 \cup \mathcal{K}_5 \), there exist constants \( \xi_i > 0 (i = 1, 2, 3, 4, 5) \) such that

\[
\|w^k - \tilde{w}^k\|_G^2 \geq \xi_1 \|A_1 (x_1^k - x_1^k+1)\|^2 + \xi_2 \|A_2 (x_2^k - x_2^k+1)\|^2 + \xi_3 \|A_3 (x_3^k - x_3^k+1)\|^2 \\
+ \xi_4 \left( \|A_1 x_1^k+1 + A_2 x_2^k+1 + A_3 x_3^k+1 - b\|^2 - \|A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b\|^2 \right) \\
+ \xi_5 \|A_1 x_1^k+1 + A_2 x_2^k+1 + A_3 x_3^k+1 - b\|^2.
\]

**Proof** We also analyze (3.24) separately in three different cases. Because the inequalities (3.23) and (3.19) hold for any \( \sigma > 1 \) and \( (\tau, s) \in \mathcal{K} \), we thus have

\[
\|w^k - \tilde{w}^k\|_G^2 \geq (\sigma - \eta)\beta \|A_1 (x_1^k - x_1^k+1)\|^2 + (\sigma - \frac{1}{\eta})\beta \|A_2 (x_2^k - x_2^k+1)\|^2 + \frac{(1-\tau)^2}{1+\tau} \beta \|A_3 (x_3^k - x_3^k+1)\|^2 \\
+ (2-\tau-s)\beta \|A_1 x_1^k+1 + A_2 x_2^k+1 + A_3 x_3^k+1 - b\|^2 \\
+ \frac{2(1-\tau)(1-\tau)}{1+\tau} \beta (A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b)^T A_3 (x_3^k - x_3^k+1).
\]

(3.25)

(Case 1) For any

\[ (\tau, s) \in \mathcal{K}_3 = \left\{ (\tau, s) | \tau = 0, 1 < s < \frac{1+\sqrt{5}}{2} \right\}, \]

we know \( 1 - s < 0 \). If there exists \( N_1 > 0 \), then by the Cauchy-Schwarz inequality, there is

\[
2(1-s)\beta (A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b)^T A_3 (x_3^k - x_3^k+1) \geq -N_1 \beta \|A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b\|^2 \\
- \frac{(1-s)^2}{N_1} \beta \|A_3 (x_3^k - x_3^k+1)\|^2.
\]

(3.26)

Substituting (3.26) into (3.25) together with \( (\tau, s) \in \mathcal{K}_3 \), we get

\[
\|w^k - \tilde{w}^k\|_G^2 \geq (\sigma - \eta)\beta \|A_1 (x_1^k - x_1^k+1)\|^2 + (\sigma - \frac{1}{\eta})\beta \|A_2 (x_2^k - x_2^k+1)\|^2 \\
+ \left( 1 - \frac{(1-s)^2}{N_1} \right) \beta \|A_3 (x_3^k - x_3^k+1)\|^2 + (2-\tau-s)\beta \|A_1 x_1^k+1 + A_2 x_2^k+1 + A_3 x_3^k+1 - b\|^2 \\
+ N_1 \beta \left( \|A_1 x_1^k+1 + A_2 x_2^k+1 + A_3 x_3^k+1 - b\|^2 - \|A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b\|^2 \right),
\]

(3.27)
where \( \eta \in (1/\sigma, \sigma) \) for any \( \sigma > 1 \). Now, we prove that the positive number \( N_1 \) exists. By (3.27), we need
\[
\begin{align*}
\xi_1 &= (\sigma - \eta) \beta > 0, \\
\xi_2 &= (\sigma - 1/\eta) \beta > 0, \\
\xi_3 &= \left(1 - \frac{(1-s)^2}{N_1^2}\right) \beta > 0, \quad \iff (1-s)^2 < N_1 < 2 - s, \forall \beta > 0. \\
\xi_4 &= N_1 \beta > 0, \\
\xi_5 &= (2 - N_1 - s) \beta > 0,
\end{align*}
\]
By the intermediate-value theorem in mathematical analysis, \( N_1 \) exists if and only if
\[
(1-s)^2 < 2 - s \iff 0 < 1 + s - s^2
\]
\[
\iff s \in \left(\frac{1-\sqrt{5}}{2}, 1\right) \cup \left(1, \frac{1+\sqrt{5}}{2}\right).
\]
Thus only if we take \( N_1 = \{(1-s)^2 + 2 - s\}/2 \), the assertion (3.24) is proved.

**(Case II)** For any \((\tau, s) \in \mathcal{K}_4\), the assertion (3.24) holds. Since
\[
\mathcal{K}_4 = \{(\tau, s) | 0 < \tau < 1 + s - s^2, s > 1\},
\]
if setting
\[
N_2 = \tau + s + (1-s)^2,
\]
then, obviously, \( N_2 - (\tau + s) > 0 \). Since \( 1 - s < 0 \), by the Cauchy-Schwarz inequality, we get
\[
\begin{align*}
&\frac{2(1-r)(1-s)}{1+r}(A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b)^T A_3 (x_3^k - x_3^{k+1}) \\
&\geq - (N_2 - \tau - s) \beta \| A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b \|^2 - \frac{(1-r)^2(1-s)^2}{(N_2 - \tau - s)(1+r)} \beta \| A_3 (x_3^k - x_3^{k+1}) \|^2.
\end{align*}
\]
Substituting (3.29) into (3.25), it holds that
\[
\begin{align*}
\| u^k - \tilde{u}^k \|^2 &\geq (\sigma - \eta) \beta \| A_1 (x_1^k - x_1^{k+1}) \|^2 + (\sigma - 1/\eta) \beta \| A_2 (x_2^k - x_2^{k+1}) \|^2 \\
&\quad + \frac{(1-r)^2}{1+r} \left(1 - \frac{(1-s)^2}{(N_2 - \tau - s)(1+r)} \right) \beta \| A_3 (x_3^k - x_3^{k+1}) \|^2 + (2 - N_2) \beta ^2 \| A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b \|^2 \\
&\quad + (N_2 - \tau - s) \beta \left(\| A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b \|^2 - \| A_1 x_1^k + A_2 x_2^k + A_3 x_3^k - b \|^2\right),
\end{align*}
\]
where \( \eta \in (1/\sigma, \sigma) \) for any \( \sigma > 1 \). Now we prove that all the coefficients of the right-hand term of (3.30) are positive, i.e.,
\[
\begin{align*}
\xi_1 &= (\sigma - \eta) \beta > 0, \\
\xi_2 &= (\sigma - 1/\eta) \beta > 0, \\
\xi_3 &= \frac{(1-r)^2}{1+r} \left(1 - \frac{(1-s)^2}{(N_2 - \tau - s)(1+r)} \right) \beta > 0, \quad \forall \beta > 0. \\
\xi_4 &= (N_2 - \tau - s) \beta > 0, \\
\xi_5 &= (2 - N_2) \beta > 0,
\end{align*}
\]
Clearly, by the known conditions, \( \xi_1, \xi_2, \xi_4 \) have been positive. For the domain \( \mathcal{K}_4 \), we know
\[
\tau < 1 + s - s^2 \implies 1 > \tau + s^2 - s
\]
\[
\implies 2 > \tau + s^2 - s + 1
\]
\[
\implies 2 > \tau + s + (1-s)^2 = N_2,
\]
which implies \( \xi_5 > 0 \). Besides, by the definition of \( N_2 \) in (3.28), we obtain
\[
\xi_3 = \frac{(1-r)^2}{1+r} \left(1 - \frac{(1-s)^2}{(N_2 - \tau - s)(1+r)} \right) \beta = \frac{\tau (1-r)^2}{(1+r)^2} > 0, \quad \forall (\tau, s) \in \mathcal{K}_4.
\]

**(Case III)** For any \((\tau, s) \in \mathcal{K}_4\), the assertion (3.24) holds. The proof is similar to that of Case II (see also Lemma 5.11 [10] for more details), which is omitted here. ■
Remark 3.2 The proof of Theorem 3.5 implies why we choose $|\tau| < 1 + s - s^2$. The difference between Theorem 3.4 and Theorem 3.5 is that the lower bound of $\|w^k - \bar{w}^k\|^2_{L^2}$ is different due to the variant choice of $(\tau, s)$, but it does not affect the convergence of the proposed method.

3.4 Convergence analysis

In this subsection, we will use the obtained Theorems 3.1-3.2 and Theorems 3.4-3.5 to establish the global convergence and the worst-case $O(1/t)$ convergence rate of the ADMM (1.8).

The following corollary can be deduced easily from Theorem 3.2 together with Theorems 3.4-3.5.

Corollary 3.1 Let the sequence $\{w^{k+1}\}$ be generated by the ADMM (1.8). Then for any $w^* \in M^*$ and $\sigma > 1$, the following two inequalities hold:

(i) For any $\tau, s \in \mathcal{K}_1 \cup \mathcal{K}_2$, there exist constants $\xi_i > 0 (i = 1, 2, 3, 4)$ such that

$$
\|w^{k+1} - w^*\|^2_H \leq \|w^k - w^*\|^2_H - \left(\xi_1 \|A_1(x^k_1 - x^{k+1}_1)\|^2 + \xi_2 \|A_2(x^k_2 - x^{k+1}_2)\|^2 - \xi_3 \|A_3(x^k_3 - x^{k+1}_3)\|^2 + \xi_4 \|A_4(x^k_4 + A_2x^{k+1}_2 + A_3x^{k+1}_3 - b)\|^2\right);
$$

(ii) For any $(\tau, s) \in \mathcal{K}_3 \cup \mathcal{K}_4 \cup \mathcal{K}_5$, there exist constants $\xi_i > 0 (i = 1, 2, 3, 4, 5)$ such that

$$
\|w^{k+1} - w^*\|^2_H + \xi_1 \|A_1x^{k+1}_1 + A_2x^{k+1}_2 + A_3x^{k+1}_3 - b\|^2 - \left(\xi_2 \|A_1(x^k_1 - x^{k+1}_1)\|^2 + \xi_3 \|A_2(x^k_2 - x^{k+1}_2)\|^2 + \xi_4 \|A_3(x^k_3 - x^{k+1}_3)\|^2\right)
$$

$$
- \xi_5 \|A_1x^{k+1}_1 + A_2x^{k+1}_2 + A_3x^{k+1}_3 - b\|^2.
$$

Remark 3.3 The above corollary illustrates that for any $\sigma > 1$ the sequence $\{\|w^k - w^*\|^2_H\}$ is contractive for the parameters $(\tau, s) \in \mathcal{K}_1 \cup \mathcal{K}_2$, i.e.,

$$
\|w^{k+1} - w^*\|^2_H \leq \|w^k - w^*\|^2_H, \ w^* \in M^*.
$$

And the sequence $\{\|w^k - w^*\|^2_H + \xi_1 \|A_1x^{k+1}_1 + A_2x^{k+1}_2 + A_3x^{k+1}_3 - b\|^2\}$ is also contractive for the parameters $(\tau, s) \in \mathcal{K}_3 \cup \mathcal{K}_4 \cup \mathcal{K}_5$. These two styles of contractive properties guarantee the convergence of the ADMM (1.8).

The following result is obtained directly from Theorem 3.1 and Theorems 3.4-3.5.

Corollary 3.2 Let the sequence $\{w^{k+1}\}$ be generated by the ADMM (1.8) and the iterate $\bar{w}^k$ be defined in (2.6a). Then the following two inequalities hold:

(i) For any $(\tau, s) \in \mathcal{K}_1 \cup \mathcal{K}_2$, we have

$$
f(x) - f(\bar{x}^k) + \langle w - \bar{w}^k, J(w) \rangle + \frac{1}{2} \|w - w^k\|^2_H \geq \frac{1}{2} \|w - w^{k+1}\|^2_H, \ \forall w \in M; \tag{3.32a}
$$

(ii) For any $(\tau, s) \in \mathcal{K}_3 \cup \mathcal{K}_4 \cup \mathcal{K}_5$, there exists a constant $\xi_i > 0$ such that

$$
f(x) - f(\bar{x}^k) + \langle w - \bar{w}^k, J(w) \rangle + \frac{1}{2} \left(\|w - w^k\|^2_H + \xi_1 \|A_1x^k_1 + A_2x^k_2 + A_3x^k_3 - b\|^2\right)
$$

$$
\geq \frac{1}{2} \left(\|w - w^{k+1}\|^2_H + \xi_1 \|A_1x^{k+1}_1 + A_2x^{k+1}_2 + A_3x^{k+1}_3 - b\|^2\right), \ \forall w \in M. \tag{3.32b}
$$
Theorem 3.6 Let the sequence \( \{w^{k+1}\} \) be generated by the ADMM (1.8) and the iterate \( \tilde{w}^k \) be defined in (2.6a). Then for any \((\tau, s) \in K_1\),

(i) it holds that

\[
\lim_{k \to \infty} \left( \|A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b\|^2 + \sum_{i=1}^{3} \|A_i(x_i^k - x_i^{k+1})\|^2 \right) = 0; \tag{3.33}
\]

(ii) the sequences \( \{w^k\} \) and \( \{\tilde{w}^k\} \) are bounded;

(iii) any accumulation point of \( \{\tilde{w}^k\} \) is a solution point of VI\((f, J, M)\);

(iv) there exists \( w^\infty \in M^* \) such that \( \lim_{k \to \infty} \tilde{w}^k = w^\infty \).

Proof Summing the inequality (3.31a) over \( k = 0, 1, 2, \ldots, \infty \), we have

\[
\sum_{k=0}^{\infty} \left( \xi_4 \|A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b\|^2 + \sum_{i=1}^{3} \xi_i \|A_i(x_i^k - x_i^{k+1})\|^2 \right) \leq \|w^0 - w^*\|^2_H,
\]

which shows that the assertion (3.33) holds. In a similar way, it follows from Theorem 3.2 that

\[
\sum_{k=0}^{\infty} \|w^k - \tilde{w}^k\|^2_G \leq \|w^0 - w^*\|^2_H \implies \lim_{k \to \infty} w^k - \tilde{w}^k = 0, \forall (\tau, s) \in K_1, \tag{3.34}
\]

where the above assertion uses the positivity of matrices \( G \) and \( H \) in the domain \( K_1 \). Hence, by (3.34) the second assertion holds. Taking the limit of both sides of (2.7a) together with (3.34), we have

\[
\lim_{k \to \infty} \left\{ f(x) - f(\tilde{x}^k) + \langle w - \tilde{w}^k, J(\tilde{w}^k) \rangle \right\} \geq 0, \forall \tilde{w}^k \in M,
\]

which verifies that \( \lim_{k \to \infty} \tilde{w}^k \) is a solution point of VI\((f, J, M)\), i.e., the assertion (iii) holds.

Let \( w^\infty \) be an accumulation point of \( \{\tilde{w}^k\} \). Then the assertion (iii) shows that \( w^\infty \in M^* \), which together with (3.9) implies that

\[
\|w^{k+1} - w^\infty\|^2_H \leq \|w^k - w^\infty\|^2_H - \|w^k - \tilde{w}^k\|^2_G.
\]

Combining the above inequality and (3.34), the proof of the assertion (iv) is completed. \( \blacksquare \)

Theorem 3.7 Let the sequence \( \{w^{k+1}\} \) be generated by the ADMM (1.8) and the iterate \( \tilde{w}^k \) be defined in (2.6a). Then the assertion (3.33) holds for any \((\tau, s) \in K_2 \cup K_3 \cup K_4 \cup K_5\). Moreover, if the matrices \( A_i(i = 1, 2, 3) \) are assumed to be full column rank, then the sequence \( \{\tilde{w}^k\} \) converges to a solution point of \( w^\infty \in M^* \).

Proof By (3.31), we can prove this theorem in a similar way of proving Theorem 6.3 [10]. \( \blacksquare \)

Remark 3.4 Both Theorem 3.6 and Theorem 3.7 show that the ADMM (1.8) is globally convergent for the step sizes \((\tau, s)\) in the domain \( K \), but the former uses the positivity of matrix \( G \) and the latter uses the full column rank assumption of the matrices \( A_i(i = 1, 2, 3) \).

Theorem 3.8 Let the sequence \( \{w^{k+1}\} \) be generated by the ADMM (1.8) and the iterate \( \tilde{w}^k \) be defined in (2.6a). For any integer \( t > 0 \), let

\[
\hat{w}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{w}^k. \tag{3.35}
\]

(i) For any \((\tau, s) \in K_1 \cup K_2\), we have

\[
f(\hat{x}_t) - f(x) + \langle \hat{w}_t - w, J(w) \rangle \leq \frac{1}{2(t+1)} \|w^0 - w\|^2_H, \forall w \in M; \tag{3.36a}
\]
(ii) For any \((\tau, s) \in \mathcal{K}_3 \cup \mathcal{K}_4 \cup \mathcal{K}_5\), we have

\[
    f(\dot{x}_t) - f(x) + \langle \dot{w}_t - w, \mathcal{J}(w) \rangle \leq \frac{1}{2(t+1)} \left( \|w^0 - w\|_H^2 + \xi_1 \|A_1x_1^0 + A_2x_2^0 + A_3x_3^0 - b\|^2 \right), \ \forall w \in \mathcal{M}.
\]

**Proof** Notice that the inequalities (3.32a)-(3.32b) can be respectively rewritten as

\[
    \begin{align*}
    &f(\tilde{x}^k) - f(x) + \langle \tilde{w}^k - w, \mathcal{J}(w) \rangle + \frac{1}{2} \|w - w^{k+1}\|_H^2 \leq \frac{1}{2} \|w - w^k\|_H^2, \\
    &f(\tilde{x}^k) - f(x) + \langle \tilde{w}^k - w, \mathcal{J}(w) \rangle + \frac{1}{2} \left( \|w - w^{k+1}\|_H^2 + \xi_1 \|A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b\|^2 \right) \\
    &\leq \frac{1}{2} \left( \|w - w^k\|_H^2 + \xi_1 \|A_1x_1^k + A_2x_2^k + A_3x_3^k - b\|^2 \right), \ \forall w \in \mathcal{M},
    \end{align*}
\]

\[
    \begin{align*}
    &\sum_{k=0}^{t} f(\tilde{x}^k) - (t+1)f(x) + \left( \sum_{k=0}^{t} \tilde{w}^k - (t+1)w, \mathcal{J}(w) \right) \leq \frac{1}{2(t+1)} \|w - w^0\|_H^2, \\
    &\sum_{k=0}^{t} f(\tilde{x}^k) - (t+1)f(x) + \left( \sum_{k=0}^{t} \tilde{w}^k - (t+1)w, \mathcal{J}(w) \right) \\
    &\leq \frac{1}{2(t+1)} \left( \|w - w^0\|_H^2 + \xi_1 \|A_1x_1^0 + A_2x_2^0 + A_3x_3^0 - b\|^2 \right), \ \forall w \in \mathcal{M}.
    \end{align*}
\]

Because \(f(x)\) is convex and \(\dot{x}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{x}^k\), therefore, it follows that

\[
    f(\dot{x}_t) \leq \frac{1}{t+1} \sum_{k=0}^{t} f(\tilde{x}^k).
\]

Substituting it respectively into the two inequalities of (3.37) and using the definition of \(\dot{w}_t\) in (3.35), the proof is immediately completed. □

**Remark 3.5** The above Theorems 3.6-3.7 imply that if the matrices \(A_i (i = 1, 2, 3)\) are assumed to be full column rank, then we can choose an easily implementable stopping criterion when carrying out some numerical experiments, that is,

\[
    \max \left\{ \|x_1^k - x_1^{k+1}\|_\infty, \|x_2^k - x_2^{k+1}\|_\infty, \|x_3^k - x_3^{k+1}\|_\infty, \|A_1x_1^k + A_2x_2^k + A_3x_3^k - b\|_\infty \right\} \leq \text{tol}.
\]

Otherwise, we can use the following stopping criterion

\[
    \max \left\{ \|A_1(x_1^k - x_1^{k+1})\|_\infty, \|A_2(x_2^k - x_2^{k+1})\|_\infty, \|A_3(x_3^k - x_3^{k+1})\|_\infty, \|A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b\|_\infty \right\} \leq \text{tol},
\]

where \(\text{tol}\) is the given tolerance error and \(w^{k+1}\) is the \((k+1)\)th iteration of the ADMM (1.8).

**Remark 3.6** Theorem 3.8 tells us that for any given compact set \(\mathcal{K} \subset \mathcal{M}\), if taking

\[
    \sup_{w \in \mathcal{K}} \|w^0 - w\|_H^2 = \eta
\]

and

\[
    \sup_{w \in \mathcal{K}} \left( \|w^0 - w\|_H^2 + \xi_1 \|A_1x_1^0 + A_2x_2^0 + A_3x_3^0 - b\|^2 \right) = \eta,
\]
the vector \( \hat{w}_t \) defined in (3.35) must satisfy

\[
f(\hat{x}_t) - f(x) + \langle \hat{w}_t - w, J(w) \rangle \leq \frac{\eta}{2(t + 1)},
\]

which shows that the proposed method converges in a worst-case \( O(1/t) \) rate in an ergodic sense.

4 Extensions to the multi-block case

Noting that in Sections 2-3 the upper-left 2 by 2 blocks in matrices \( Q, M \) and \( G \) are positive definite for \( \sigma > 1 \), and the treatment of the lower-right 2 by 2 blocks are almost the same as in [10]. These marked observations motivate us to further study the following multi-block separable convex minimization problem

\[
\min \left\{ \sum_{i=1}^{p} f_i(x_i) \right\}, \quad \forall x \in X, \quad i = 1, 2, 3, \ldots, p,
\]

where \( f_i(x_i) \) are closed and proper convex functions (not necessarily smooth); \( A \in \mathbb{R}^{n \times m_i} \) and \( b \in \mathbb{R}^n \) are given matrices and vectors, respectively; \( X \subset \mathbb{R}^{m_i} (i = 1, 2, 3, \ldots, p) \) are closed convex sets.

In a similar way as (1.8), we can design the corresponding ADMM scheme for (4.1), that is, the first \( p - 1 \) subproblems are solved in a parallel manner with some positive-definite proximal terms:

\[
\begin{align*}
  x_1^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta (x_1, x_2, x_3, \ldots, x_p, \lambda) + \frac{\sigma}{2} \| A_1(x_1 - x_1^k) \|^2 \mid x_1 \in X_1 \right\}, \\
  x_2^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta (x_1^k, x_2, x_3, \ldots, x_p, \lambda) + \frac{\sigma}{2} \| A_2(x_2 - x_2^k) \|^2 \mid x_2 \in X_2 \right\}, \\
  &\ldots \\
  x_i^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta (x_1^k, x_2^k, \ldots, x_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_p, \lambda) + \frac{\sigma}{2} \| A_i(x_i - x_i^k) \|^2 \mid x_i \in X_i \right\}, \\
  &\ldots \\
  x_{p-1}^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta (x_1^k, x_2^k, \ldots, x_{p-2}^k, x_{p-1}^k, x_p, \lambda) + \frac{\sigma}{2} \| A_{p-1}(x_{p-1} - x_{p-1}^k) \|^2 \mid x_{p-1} \in X_{p-1} \right\}, \\
  \lambda^{k+1} &= \lambda^k - \tau \beta \left( A_1 x_1^{k+1} + A_2 x_2^{k+1} + \ldots + A_{p-1} x_{p-1}^{k+1} + A_p x_p^{k+1} - b \right), \\
  x_p^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta (x_1^k, x_2^k, \ldots, x_{p-1}^k, x_p, \lambda^{k+1}) \mid x_p \in X_p \right\}, \\
  \lambda^{k+1} &= \lambda^{k+1} - s \beta \left( A_1 x_1^{k+1} + A_2 x_2^{k+1} + \ldots + A_p x_p^{k+1} - b \right),
\end{align*}
\]

where

\[
\mathcal{L}_\beta (x_1, x_2, x_3, \ldots, x_p, \lambda) = \sum_{i=1}^{p} f_i(x_i) - \langle \lambda, \sum_{i=1}^{p} A_i x_i - b \rangle + \frac{\beta}{2} \sum_{i=1}^{p} \| A_i x_i - b \|^2
\]

is the augmented Lagrangian function of the problem (4.1) and \( \sigma, \tau, s \) are three independent constants that are respectively restricted into

\[
\sigma \in (p - 2, +\infty), \quad (\tau, s) \in \mathcal{K} = \{(r, s) \mid r + s > 0, r < 1, |r| < 1 + s - s^2\}.
\]

For the saddle point \( (x_1^*, x_2^*, \ldots, x_p^*, \lambda^*) \) belonging to the solution set \( \mathcal{M}^* \) of the problem (4.1), it satisfies the following variational inequality

\[
\forall (f, J, \mathcal{M}) : f(x) - f(x^*) + \langle w - w^*, J(w^*) \rangle \geq 0, \quad \forall w \in \mathcal{M},
\]

where \( f(x) = \sum_{i=1}^{p} f_i(x_i), \mathcal{M} = X_1 \times X_2 \times \cdots \times X_p \times \mathbb{R}^n \) and

\[
\begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_p \\
\lambda
\end{pmatrix}, \quad
\begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_p
\end{pmatrix}, \quad J(w) =
\begin{pmatrix}
-A_1^T \lambda \\
-A_2^T \lambda \\
\vdots \\
-A_p^T \lambda \\
A_1 x_2 + A_2 x_2 + \ldots + A_p x_p - b
\end{pmatrix}.
\]
Define

\[ w^k = \begin{pmatrix} x^k_1 \\ x^k_2 \\ \vdots \\ x^k_p \end{pmatrix}, \quad \tilde{w}^k = \begin{pmatrix} \tilde{x}^k_1 \\ \tilde{x}^k_2 \\ \vdots \\ \tilde{x}^k_p \end{pmatrix}, \]

where \( \tilde{x}^k = (\tilde{x}^k_1, \tilde{x}^k_2, \ldots, \tilde{x}^k_p) = (x^k_1 + 1, x^k_2 + 1, \ldots, x^k_p + 1) \) and

\[ \tilde{\lambda}^k = \lambda^k - \beta (A_1 x^k_1 + A_2 x^k_2 + \ldots + A_{p-1} x^k_{p-1} + A_p x^k_p - b). \]

Using Lemma 2.1 and (4.5), the optimal condition of the \( x_i \)-subproblem \( (i = 1, 2, \ldots, p-1) \) of the ADMM (4.2) can be expressed as

\[
x_{i}^{k+1} \in \mathcal{X}_i, \quad f_i(x_i) - f_i(x_{i}^{k+1}) + \langle x_i - \tilde{x}_{i}^{k}, -A_i^T \tilde{\lambda}^k - \beta A_i^T A_i (\tilde{x}_{i}^{k} - x_{i}^{k}) \rangle + \sum_{j=1, j \neq i}^{p-1} A_j (\tilde{x}_{j}^{k} - x_{j}^{k}) \geq 0, \forall x_i \in \mathcal{X}_i,
\]

(4.6a)

Similarly, the optimal condition of the \( x_p \)-subproblem of the ADMM (4.2) is \( x_{p}^{k+1} \in \mathcal{X}_p \) and

\[
f_p(x_p) - f_p(\tilde{x}_{p}^{k}) + \langle x_p - \tilde{x}_{p}^{k}, -A_p^T \tilde{\lambda}^k - \beta A_p^T A_p (\tilde{x}_{p}^{k} - x_{p}^{k}) - \tau \lambda^k \rangle \geq 0, \forall x_p \in \mathcal{X}_p.
\]

(4.6b)

Combining (4.6a) and (4.6b), the sequence \( \{w^k\} \) in the ADMM (4.2) satisfies

\[ \tilde{w}^k \in \mathcal{M}, \quad f(x) - f(\tilde{x}^k) + \langle w, \tilde{w}^k, J(\tilde{w}^k) + Q(\tilde{w}^k - w^k) \rangle \geq 0, \quad \forall w \in \mathcal{M}, \]

(4.7a)

where

\[
Q = \begin{bmatrix}
\sigma \beta A_1^T A_1 & -\beta A_1^T A_2 & \cdots & -\beta A_1^T A_{p-1} & 0 & 0 \\
-\beta A_2^T A_1 & \sigma \beta A_2^T A_2 & \cdots & -\beta A_2^T A_{p-1} & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & 0 \\
-\beta A_{p-1} A_1 & -\beta A_{p-1}^T A_2 & \cdots & \sigma \beta A_{p-1}^T A_{p-1} & 0 & 0 \\
0 & 0 & \cdots & \beta A_p^T A_p - \tau \lambda^k & 0 & 0 \\
0 & 0 & \cdots & -A_p & \frac{1}{\beta} I_n \\
\end{bmatrix}.
\]

(4.7b)

The following Lemma is easy to be obtained in a similar way as the proof of Lemma 2.3.

**Lemma 4.1** For the sequence \( \{w^{k+1}\} \) generated by the ADMM (4.2) and the iterate \( \tilde{w}^k \) defined in (4.5), the following relationship

\[ w^{k+1} = w^k - M(w^k - \tilde{w}^k) \]

holds with

\[
M = \begin{bmatrix}
I_{m_1} & 0 & 0 & 0 & 0 & 0 \\
0 & I_{m_2} & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & 0 & 0 \\
0 & 0 & 0 & I_{m_{p-1}} & 0 & 0 \\
0 & 0 & 0 & 0 & I_{m_p} & 0 \\
0 & 0 & 0 & 0 & -s \beta A_p & (-s + \tau) I_n \\
\end{bmatrix}.
\]

(4.8a)
Clearly, the above variational characterization (4.7) and prediction-correction interpretation (4.8) are the extended versions of (2.7) and (2.13) for the three-block separable convex minimization problem. Hence, the convergence analysis of the ADMM (4.2) is similar to that of the ADMM (1.8) in Section 3, which is omitted for succinctness.

5 Conclusion and discussion

Since Chen et al.[1] analyzed that the direct extension of the ADMM with a Gauss-Seidel for solving the three-block separable convex minimization problem is not necessarily convergence, there has been a constantly increasing interest in developing and improving the theory of the ADMM for such problem. In this paper, the idea of combining the parallel and serial algorithms is applied to improve the traditional ADMM, and a novel ADMM scheme (1.8) is developed to solve the convex programming with three-block objectives, where the first two subproblems are regularized by some positive-definite proximal terms. Moreover, the domain of the step size is enlarged, but the global convergence of the method with a worst-case $O(1/t)$ convergence rate still holds. Also, the proposed method is extended to solve the multi-block separable convex minimization problem.

It is worthwhile to think further that the ADMM (1.8) is designed by improving the first two subproblems and adding a step of updating the Lagrangian multiplier. The following general iteration scheme is also worthwhile to discuss deeper, that is,

$$
\begin{align*}
    x^{k+1}_1 &= \arg \min \left\{ \mathcal{L}_\beta(x_1, x_2, x_3, \ldots, x_p, \lambda^k) \mid x_1 \in \mathcal{X}_1 \right\}, \\
    \lambda^{k+\frac{1}{2}} &= \lambda^k - \tau \beta \left( A_1 x^{k+1}_1 + A_2 x_2 + A_3 x_3 + \cdots + A_p x_p - b \right), \\
    x^{k+1}_2 &= \arg \min \left\{ \mathcal{L}_\beta \left( x^{k+1}_1, x_2, x_3, \ldots, x_p, \lambda^{k+\frac{1}{2}} \right) + \frac{\sigma \beta}{2} \left\| A_2 (x_2 - x_2^k) \right\|^2 \mid x_2 \in \mathcal{X}_2 \right\}, \\
    x^{k+1}_3 &= \arg \min \left\{ \mathcal{L}_\beta \left( x^{k+1}_1, x_2^k, x_3, \ldots, x_p^k, \lambda^{k+\frac{1}{2}} \right) + \frac{\sigma \beta}{2} \left\| A_3 (x_3 - x_3^k) \right\|^2 \mid x_3 \in \mathcal{X}_3 \right\}, \\
    \vdots \\
    x^{k+1}_p &= \arg \min \left\{ \mathcal{L}_\beta \left( x^{k+1}_1, x_2^k, x_3^k, \ldots, x_{p-1}^k, x_p, \lambda^{k+\frac{1}{2}} \right) + \frac{\sigma \beta}{2} \left\| A_p (x_p - x_p^k) \right\|^2 \mid x_p \in \mathcal{X}_p \right\}, \\
    \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - s \beta \left( A_1 x^{k+1}_1 + A_2 x_2^{k+1} + A_3 x_3^{k+1} + \cdots + A_p x_p^{k+1} - b \right).
\end{align*}
$$

Maybe the convergence of the above scheme need to change the range of $\sigma$ and $(\tau, s)$. Besides, another ADMM scheme that uses the convex combination of the last iteration of each variables (see e.g.[12]) will be investigated in the future work.

References


