A class of parallel splitting method inspired by pseudo search direction for separable convex programming

Shengjie Xu

Jun 8, 2020

Abstract. We present a class of parallel splitting augmented Lagrange method (ALM) based algorithm for solving the separable convex minimization problem with linear constraints, whose objective function is the sum of \( m \) individual subfunctions without coupled variables. In real computations, it is well known that an algorithm with a Jacobian structure can achieve parallel computing but has a lower efficiency; algorithm with Gauss-Seidel structure has a more efficient numerical performance but cannot perform parallel computing. In this paper, inspired by a mixture of two types of algorithms, we propose a novel hybrid algorithm based on a prediction-correction way. The novel method needs only a simple back substitution procedure to the original parallel splitting step. Furthermore, we establish its global convergence and a worst-case \( O(1/t) \) convergence rate. Finally, taking a sparse matrix minimization problem arising in statistical learning as an example, we show the efficient performance of the purposed method.

Keywords: augmented Lagrange method, convex optimization, convergence rate, operator splitting method, parallel computing

1 Introduction

The augmented Lagrange method (ALM) is a benchmark of convex optimization, and it was originally developed as a fundamental tool to solve the convex optimization problem with linear equality constrains in [17, 36]. In this paper, we consider the following separable convex minimization problem:

\[
\min \left\{ \sum_{i=1}^{m} \theta_i(x_i) \mid \sum_{i=1}^{m} A_i x_i = b; \ x_i \in \mathcal{X}_i, \ i = 1, 2, \ldots, m \right\}, \tag{1.1}
\]

where \( A_i \in \mathbb{R}^{l \times n_i} (i = 1, \ldots, m) \) and \( b \in \mathbb{R}^l; \mathcal{X}_i \subset \mathbb{R}^{n_i} (i = 1, \ldots, m) \) are closed convex sets; and \( \theta_i(x_i) : \mathbb{R}^{n_i} \to \mathbb{R} (i = 1, \ldots, m) \) are closed convex but not necessarily smooth functions. Such a model (1.1) captures a lot of applications in different areas, we refer the readers to [1, 2, 3, 41, 47]. Throughout, we assume the solution set of (1.1) is nonempty and each \( A_i (i = 1, \ldots, m) \) is full column rank. For problem (1.1), it is well known that its corresponding Lagrange function is defined as

\[
L(x_1, x_2, \ldots, x_m, \lambda) = \sum_{i=1}^{m} \theta_i(x_i) - \lambda^T \left( \sum_{i=1}^{m} A_i x_i - b \right), \tag{1.2}
\]

where \( \lambda \in \mathbb{R}^l \) is the associate Lagrangian multiplier.
1.1 Two fundamental operator splitting ways of ALM

Suppose $\beta > 0$, by adding the quadratic penalty term $\frac{\beta}{2} \| \sum_{i=1}^{m} A_i x_i - b \|_2^2$ for the linear equality constraint $\sum_{i=1}^{m} A_i x_i - b = 0$, the augmented Lagrange function w.r.t. (1.2) then reads as

$$\mathcal{L}_\beta(x_1, x_2, \cdots, x_m, \lambda) = \sum_{i=1}^{m} \theta_i(x_i) - \lambda^T (\sum_{i=1}^{m} A_i x_i - b) + \frac{\beta}{2} \| \sum_{i=1}^{m} A_i x_i - b \|_2^2. \quad (1.3)$$

Furthermore, for the given initial point $(x^k_1, \cdots, x^k_m, \lambda^k)$, the original ALM enjoys an iterative form:

$$(\text{ALM}) \begin{cases} (x^{k+1}_1, \cdots, x^{k+1}_m) &= \arg \min \{ \mathcal{L}_\beta(x_1, x_2, \cdots, x_m, \lambda^k) \mid x_i \in \mathcal{X}_i, \ i = 1, \cdots, m \}, \\ \lambda^{k+1} &= \lambda^k - \beta (\sum_{i=1}^{m} A_i x^{k+1}_i - b). \end{cases}$$

However, noting there are couple terms in penalty term $\| \sum_{i=1}^{m} A_i x_i - b \|_2^2$, the $x$-subproblem generally cannot be solved separately, thus it arises the following two common-used operator splitting strategy: Jacobian ALM-based splitting way

$$(\text{Jacobian}) \begin{cases} x^{k+1}_1 &= \arg \min \{ \mathcal{L}_\beta(x_1, x^k_2, \cdots, x^k_{m-1}, x^k_m, \lambda^k) \mid x_1 \in \mathcal{X}_1 \}, \\ x^{k+1}_2 &= \arg \min \{ \mathcal{L}_\beta(x^k_1, x_2, \cdots, x^k_{m-1}, x^k_m, \lambda^k) \mid x_2 \in \mathcal{X}_2 \}, \\ \vdots \\ x^{k+1}_m &= \arg \min \{ \mathcal{L}_\beta(x^k_1, x^k_2, \cdots, x^k_{m-1}, x_m, \lambda^k) \mid x_m \in \mathcal{X}_m \}, \end{cases}$$

and Gaussian ALM-based splitting way

$$(\text{Gaussian}) \begin{cases} x^{k+1}_1 &= \arg \min \{ \mathcal{L}_\beta(x_1, x^k_2, \cdots, x^k_{m-1}, x^k_m, \lambda^k) \mid x_1 \in \mathcal{X}_1 \}, \\ x^{k+1}_2 &= \arg \min \{ \mathcal{L}_\beta(x^k_1, x_2, \cdots, x^k_{m-1}, x^k_m, \lambda^k) \mid x_2 \in \mathcal{X}_2 \}, \\ \vdots \\ x^{k+1}_m &= \arg \min \{ \mathcal{L}_\beta(x^k_1, x^k_2, \cdots, x^k_{m-1}, x_m, \lambda^k) \mid x_m \in \mathcal{X}_m \}. \end{cases}$$

1.2 A brief review of ALM with the Gaussian-structure based methods

If we adopt the Gaussian structure of $x$-subproblem in the original ALM, the method will reduce to the direct extension of ADMM (denoted by DADMM), which has an iterative form:

$$(\text{DADMM}) \begin{cases} x^{k+1}_1 &= \arg \min \{ \mathcal{L}_\beta(x_1, x^k_2, \cdots, x^k_{m-1}, x^k_m, \lambda^k) \mid x_1 \in \mathcal{X}_1 \}, \\ x^{k+1}_2 &= \arg \min \{ \mathcal{L}_\beta(x^k_1, x_2, \cdots, x^k_{m-1}, x^k_m, \lambda^k) \mid x_2 \in \mathcal{X}_2 \}, \\ \vdots \\ x^{k+1}_m &= \arg \min \{ \mathcal{L}_\beta(x^k_1, x^k_2, \cdots, x^k_{m-1}, x_m, \lambda^k) \mid x_m \in \mathcal{X}_m \}, \\ \lambda^{k+1} &= \lambda^k - \beta (\sum_{i=1}^{m} A_i x^{k+1}_i - b). \end{cases}$$

Especially, if $m = 2$, it will be the classical alternating direction method of multipliers (abbreviated as ADMM), which was proposed originally in [14, 15] and is indeed an application of the Douglas-Rachford splitting method [12]. ADMM has been common used in different application areas (see details in, e.g., [2, 5, 10]), and in the literature, its convergence analysis has been well studied, see
related work in e.g., [18, 23, 25]. Meanwhile, there are many variants of ADMM, such as symmetric ADMM in [22, 31], linearized ADMM in [30, 33, 43, 46] and the generalized ADMM in [11]. For the case \( m \geq 3 \), DADMM has an efficient performance in real computations, we refer to e.g., [37, 42] and references cited therein. However, noting that there is a counterexample in [6], DADMM is not necessarily convergent.

Therefore, it arises some convergent methods for some specific applications, we can see related work in, e.g., [24, 26, 27, 28, 32, 40, 41, 42]. Especially, He et al. proposed a ADMM with Gaussian back substitution method (GADMM) in [26, 27, 29], it needs only an additional correction step based on the direct extension of ADMM, while has an efficient numerical performance.

1.3 A brief review of ALM with the Jacobian-structure based methods

If we take the Jacobian structure of \( x \)-subproblem in the classical ALM, it would reduce to the direct extension of ALM (DALM), which has the following iterative scheme:

\[
\begin{align*}
(x^{k+1}_1) &= \arg \min \{ \mathcal{L}_\beta(x_1, x_2^k, \cdots, x_{m-1}^k, x_m^k, \lambda^k) \mid x_1 \in X_1 \}, \\
(x^{k+1}_2) &= \arg \min \{ \mathcal{L}_\beta(x_1^k, x_2, \cdots, x_{m-1}^k, x_m^k, \lambda^k) \mid x_2 \in X_2 \}, \\
& \vdots \\
(x^{k+1}_m) &= \arg \min \{ \mathcal{L}_\beta(x_1^k, x_2^k, \cdots, x_{m-1}^k, x_m, \lambda^k) \mid x_m \in X_m \}, \\
\lambda^{k+1} &= \lambda^k - \beta(\sum_{i=1}^m A_i x_i^{k+1} - b).
\end{align*}
\]

In fact, as described in [47], there are many applications, including Potts Model-Based image segmentation and its variants, can be solved by DALM efficiently. However, similar with the DADMM, the DALM does not necessarily converge in theory, we can see a simple counterexample in [21]. For the purpose of convergence guaranteed, there arises some novel schemes, we refer the readers to [21, 28]. It mainly includes two ways to design algorithms:

**Do correction step for DALM:** As analyzed in [21], the authors regard DALM as a prediction step, and then do the following correction step:

\[
w^{k+1} := w^k - \alpha(w^k - w^{k+1}),
\]

where \( w := (x_1, x_2, \cdots, x_m, \lambda) \) and
\[
\alpha \in \left(0, 2(1 - \sqrt{\frac{m}{m+1}})\right).
\]

After such a simple correction step, the proposed method (PALM) enjoys a worst-case \( \mathcal{O}(1/t) \) convergence rate. However, as we can see, as the increase of \( m \), the \( \alpha \) would be very small, e.g., if \( m = 3 \), then \( \alpha \in (0, 0.268) \). Therefore, it is necessary to enlarge such a step size coefficient \( \alpha \).

**Add regularization term for \( x \)-subproblem:** Another algorithm-designed way is based on the regularization term. In [28], by adding the regularization term \( \frac{\tau}{2} \beta \| A_i(x_i - x_i^k) \|_2^2 \) \((i = 2, \cdots, m)\) to
(1.3), the authors propose the following parallel splitting ADMM-like $x$-subproblem solved scheme:

\[
\begin{align*}
    x_1^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta(x_1, x_2^k, \ldots, x_{m-1}^k, x_m, \lambda^k) \mid x_1 \in \mathcal{X}_1 \right\}, \\
    x_2^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta(x_1^k, x_2^k, \ldots, x_{m-1}, x_m^k, \lambda^k) + \frac{\tau}{2} \beta \|A_2(x_2 - x_2^k)\|_2^2 \mid x_2 \in \mathcal{X}_2 \right\}, \\
    &\vdots \\
    x_m^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta(x_1^k, x_2^k, \ldots, x_{m-1}, x_m^k, \lambda^k) + \frac{\tau}{2} \beta \|A_m(x_m - x_m^k)\|_2^2 \mid x_m \in \mathcal{X}_m \right\}.
\end{align*}
\]

Such a method (denoted by PSADMM) is convergent provided $\tau > m - 2$. Obviously, it leads to a small step size with the increase of $m$. In practical computations, the relaxation of regularization term allows a bigger step size to potentially reduce required convergence iterations, we refer the readers to [19, 20, 45]. Therefore, it is necessary to reduce such a regularization parameter $\tau$.

### 1.4 Contributions and outline of the paper

Our purpose in this paper is to propose a novel parallel splitting method based on ALM for solving the multi-block convex programming problem (1.1). At first, we analyze the optimality of (1.1) from a perspective of variational inequality, which is a powerful tool for convergence analysis of algorithms, we refer the readers to [9, 16, 35]. In the later sections, we can find that the pseudo variables $(A_1x_1, A_2x_2, \ldots, A_mx_m, \lambda)$ will directly determine the convergence of algorithm when each $A_i$ is full column rank. Our novel method firstly takes the following Jacobian ALM-based scheme:

\[
\begin{align*}
    x_1^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta(x_1, x_2^k, \ldots, x_{m-1}, x_m^k, \lambda^k) + \frac{\tau}{2} \beta \|A_1(x_1 - x_1^k)\|_2^2 \mid x_1 \in \mathcal{X}_1 \right\}, \\
    x_2^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta(x_1^k, x_2^k, \ldots, x_{m-1}, x_m^k, \lambda^k) + \frac{\tau}{2} \beta \|A_2(x_2 - x_2^k)\|_2^2 \mid x_2 \in \mathcal{X}_2 \right\}, \\
    &\vdots \\
    x_m^{k+1} &= \arg \min \left\{ \mathcal{L}_\beta(x_1^k, x_2^k, \ldots, x_{m-1}, x_m^k, \lambda^k) + \frac{\tau}{2} \beta \|A_m(x_m - x_m^k)\|_2^2 \mid x_m \in \mathcal{X}_m \right\}.
\end{align*}
\]

As can be seen, the term $\lambda^{k+1}$ inherits both information of $A_ix_i^k$ and $A_ix_i^{k+1}$. In addition, the novel method takes an extra Gaussian-like back substitution correction step for $x$-subproblem as

\[
\begin{align*}
    \lambda^{k+1} = \lambda^k - 2\nu/\beta \left\{ \sum_{i=1}^m A_ix_i^k + A_ix_i^{k+1} \bigg/ 2 - b \right\},
\end{align*}
\]

where $\nu \in (0, 1)$ and the method is convergent provided $\tau > (m-4)/4$. In fact, for the case $1 < m < 4$, we can take $\tau = 0$, in such a case, the prediction step (1.4) can be interpreted as a general parallel ALM method except the term $\sum_{i=1}^m (A_ix_i^k + A_ix_i^{k+1})/2 - b$, which can be regarded as an average version of the constraints $\sum_{i=1}^m A_ix_i - b$. Therefore, we also name the new method as the parallel splitting ALM-based method (denoted by PSALMM).

The novel method enjoys the following two advantages: first, compared with GADMM and DADMM, it provides a parallel splitting solution strategy and treats the objective functions separately and enables parallel computing; second, compared with PALM, the step size $\nu \in (0, 1)$ of new method can be taken a larger number. Meanwhile, it reduces the regularization term coefficient $\tau$ from $m - 2$ to 4.
to \((m - 4)/4\) compared with PSADMM, and thus provides a wider choice of the regularization term to potentially accelerate the convergence. Finally, we also establish the convergence analysis of the proposed method.

The rest of the paper is organized as follows: Sec. 2 summarizes some preliminary results and introduces some basic matrices, which will be frequently used in the following sections. In Sec. 3, we create the new parallel splitting method based on the basic line search method. Furthermore, we establish the convergence analysis of the proposed method in Sec. 4, including its global convergence and a worst-case \(O(1/t)\) convergence rate. Finally, we numerically show the efficient performance of the novel method in Sec. 5.

2 Preliminaries

In this section, we derive the optimal condition of (1.1) from a perspective of variational inequality (VI). Meanwhile, we give an equivalent representation of the proposed method in the view of prediction-correction way. We also introduce some elementary matrices will be used frequently in later discussion.

2.1 Optimal condition of (1.1) in view of VI

The analysis of this paper is based on the following lemma, its proof is elementary and thus omitted here (details can be found in [13, 29]).

**Lemma 2.1** Suppose \(X \subseteq \mathbb{R}^n\) is a closed convex set; \(\theta(x)\) and \(\varphi(x)\) are closed convex functions; \(\varphi(x)\) is differentiable. Assume the solution set of the minimization problem

\[
\min \{ \theta(x) + \varphi(x) \mid x \in X \}
\]

is nonempty, then

\[x^* \in \arg \min \{ \theta(x) + \varphi(x) \mid x \in X \}\]

if and only if

\[
x^* \in X, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla \varphi(x^*) \geq 0, \quad \forall x \in X.
\]

For problem (1.1), we denote \(\Omega := X_1 \times \cdots \times X_m \times \mathbb{R}^l\), and we say a pair \((x_1^*, x_2^*, \cdots, x_m^*, \lambda^*)\) \(\in \Omega\) is a saddle point of the Lagrange function (1.2) means

\[
L(x_1^*, x_2^*, \cdots, x_m^*, \lambda) \leq L(x_1^*, x_2^*, \cdots, x_m^*, \lambda^*) \leq L(x_1, x_2, \cdots, x_m, \lambda^*).
\]

Furthermore, we have

\[
\begin{align*}
x_1^* &= \arg \min \{ L(x_1, x_2^*, \cdots, x_m^*, \lambda) \mid x_1 \in X_1 \}, \\
x_2^* &= \arg \min \{ L(x_1^*, x_2, \cdots, x_m^*, \lambda) \mid x_2 \in X_2 \}, \\
&\vdots \\
x_m^* &= \arg \min \{ L(x_1^*, x_2^*, \cdots, x_m, \lambda) \mid x_m \in X_m \}, \\
\lambda^* &= \arg \max \{ L(x_1^*, x_2^*, \cdots, x_m^*, \lambda) \mid \lambda \in \mathbb{R}^l \}.
\end{align*}
\]
In view of the basic Lemma (2.1), we can rewrite (2.1) as following variational inequalities:

\[
\begin{align*}
  \theta_1(x_1) - \theta_1(x_1^*) + (x_1 - x_1^*)^T(-A_1^T\lambda^*) &\geq 0, & \forall x_1 \in X_1, \\
  \theta_2(x_2) - \theta_2(x_2^*) + (x_2 - x_2^*)^T(-A_2^T\lambda^*) &\geq 0, & \forall x_2 \in X_2, \\
  & \vdots \\
  \theta_m(x_m) - \theta_m(x_m^*) + (x_m - x_m^*)^T(-A_m^T\lambda^*) &\geq 0, & \forall x_m \in X_m, \\
  (\lambda - \lambda^*)^T(\sum_{i=1}^m A_i x_i^* - b) &\equiv 0, & \forall \lambda \in \mathbb{R}^l.
\end{align*}
\] (2.2)

Moreover, if we set

\[
x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \quad \theta(x) = \sum_{i=1}^m \theta_i(x_i), \quad A = \begin{pmatrix} A_1 & A_2 & \cdots & A_m \end{pmatrix},
\]

\[
w = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \quad \text{and} \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ -A_2^T\lambda \\ \vdots \\ -A_m^T\lambda \\ \sum_{i=1}^m A_i x_i - b \end{pmatrix},
\]

(2.3)

VIs (2.2) can be rewritten in the following compact form:

\[
w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega.
\] (2.4)

Throughout this paper, we denote \(\Omega^*\) the solution set of (2.4), which is assumed to be nonempty. Noting that the operator \(F(w)\) in (2.3) is affine with a skew-symmetric matrix structure, and consequently we have

\[
(w - \bar{w})^T(F(w) - F(\bar{w})) \equiv 0, \quad \forall w, \ \bar{w} \in \Omega,
\]

which shows the monotonicity of \(F\) [39]. Therefore, we also call the compact form (2.4) the mixed monotone variational inequality (abbreviated as MVI).

### 2.2 Prediction-correction representation of the novel method

We will establish our new algorithm by a standard prediction-correction way. Firstly, we will check the proposed method (1.4)-(1.5) is equivalent to the following prediction-correction (P-C) algorithm:

\[
\begin{align*}
  \tilde{x}_1^k & = \arg \min \left\{ L_\beta(x_1, x_2^k, \cdots, x_{m-1}^k, x_m^k, \lambda^k) + \frac{\beta}{2} \| A_1(x_1 - x_1^k) \|_2^2 \mid x_1 \in X_1 \right\}, \\
  \tilde{x}_2^k & = \arg \min \left\{ L_\beta(x_1^k, x_2, \cdots, x_{m-1}^k, x_m^k, \lambda^k) + \frac{\beta}{2} \| A_2(x_2 - x_2^k) \|_2^2 \mid x_2 \in X_2 \right\}, \\
  \vdots \\
  \tilde{x}_m^k & = \arg \min \left\{ L_\beta(x_1^k, \cdots, x_m, \lambda^k) + \frac{\beta}{2} \| A_m(x_m - x_m^k) \|_2^2 \mid x_m \in X_m \right\}, \\
  \tilde{\lambda}^k & = \lambda^k - \beta (Ax^k - b).
\end{align*}
\] (Pre) (2.5)
where $\mathcal{A}$ is defined in (2.3). Noting that $\tilde{\lambda}^k = \lambda^k - \beta(Ax^k - b)$, and plugging it into the correction step (2.6), we have

$$
\rho^{k+1} = \lambda^k - \nu \left\{ - \beta A(x^k - \tilde{x}^k) + 2(\lambda^k - \tilde{\lambda}^k) \right\},
$$

and

$$
A_i x_i^{k+1} = A_i x_i^k - \nu \left\{ \frac{2\tau + 1}{1 + \tau} A_i (x_i^k - \tilde{x}_i^k) - \frac{1}{1 + \tau} \left[ \sum_{j \neq i} A_j (x_j^k - \tilde{x}_j^k) - \frac{1}{\beta} (\lambda^k - \tilde{\lambda}^k) \right] \right\},
$$

(2.6)

This shows the equivalence of (2.5)-(2.6) and (1.4)-(1.5). Next, we will derive the VI-structure of the prediction step (2.5).

**Lemma 2.2** Let $\tilde{x}^k$ be the point generated by the prediction step (2.5) for problem (1.1), then it satisfies

$$
\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T \Phi(w^k - \tilde{w}^k), \quad \forall w \in \Omega,
$$

(2.7)

where

$$
\Phi = \begin{pmatrix}
(1 + \tau)\beta A_1^T A_1 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & (1 + \tau)\beta A_m^T A_m & 0 \\
-A_1 & \cdots & -A_m & \frac{1}{\beta} I_l
\end{pmatrix}.
$$

(2.8)

**Proof** For $x_i$-subproblem of the prediction step (2.5), by the basic Lemma 2.1, $\tilde{x}_i^k \in \mathcal{X}_i$ satisfies

$$
\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \left\{ - A_i^T (\lambda^k - \beta(A_i x_i^k + \sum_{j \neq i} A_j x_j^k - b)) + \tau \beta A_i^T A_i (\tilde{x}_i^k - x_i^k) \right\} \geq 0, \quad \forall x_i \in \mathcal{X}_i.
$$

Noting that $\tilde{\lambda}^k = \lambda^k - \beta(Ax^k - b)$, we have

$$
\begin{align*}
-A_i^T (\lambda^k - \beta(A_i x_i^k + \sum_{j \neq i} A_j x_j^k - b)) + \tau \beta A_i^T A_i (\tilde{x}_i^k - x_i^k) &= -A_i^T [\lambda^k - \beta(Ax^k - b)] + \beta A_i^T A_i (\tilde{x}_i^k - x_i^k) + \tau \beta A_i^T A_i (\tilde{x}_i^k - x_i^k) \\
&= -A_i^T \tilde{\lambda}^k + (1 + \tau) \beta A_i^T A_i (\tilde{x}_i^k - x_i^k).
\end{align*}
$$
Therefore, the $x_i$-subproblem satisfies the VI:
\[
\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \{ - A_i^T \tilde{\lambda}^k + (1 + \tau)\beta A_i^T A_i(\tilde{x}_i^k - x_i^k) \} \geq 0.
\] (2.9)

Similarly, for $\lambda$-subproblem in (2.5), due to $A\tilde{x}_i^k - b + \frac{1}{\beta}(\tilde{\lambda}_i^k - \lambda_i^k) = 0$, which is also equivalent to
\[
A\tilde{x}_i - b - A(\tilde{x}_i^k - x_i^k) + \frac{1}{\beta}(\tilde{\lambda}_i^k - \lambda_i^k) = 0.
\] (2.10)

Combining with the notations in (2.3), by adding (2.9) and (2.10) together, we can obtain the assertion directly.

\[\square\]

**Lemma 2.3** For the correction step (1.5), it satisfies $w^{k+1} = w^k - \nu M(w^k - \tilde{w}^k)$, where

\[
M = \begin{pmatrix}
\frac{2\tau + 1}{1 + \tau} I_l & -\frac{1}{1 + \tau} I_l & \cdots & -\frac{1}{1 + \tau} I_l & \frac{1}{(1 + \tau)\beta} I_l \\
-\frac{1}{1 + \tau} I_l & \frac{2\tau + 1}{1 + \tau} I_l & \cdots & -\frac{1}{1 + \tau} I_l & \frac{1}{(1 + \tau)\beta} I_l \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
-\frac{1}{1 + \tau} I_l & -\frac{1}{1 + \tau} I_l & \cdots & \frac{2\tau + 1}{1 + \tau} I_l & \frac{1}{(1 + \tau)\beta} I_l \\
-\beta I_l & -\beta I_l & \cdots & -\beta I_l & 2I_l
\end{pmatrix}.
\] (2.11)

**Proof** We directly obtain it by (2.6), and it is a compact representation of the correction step. \[\square\]

### 2.3 Pseudo variables essentially determine the convergence

According to (2.7), $\tilde{w}^k$ would be a solution if the right hand side $(w - \tilde{w}^k)^T \Phi(w^k - \tilde{w}^k)$ goes to 0. Noting that

\[
\Phi = \begin{pmatrix}
(1 + \tau)\beta A_1^T A_1 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & (1 + \tau)\beta A_m^T A_m & 0 \\
-A_1 & \cdots & -A_m & \frac{1}{\beta} I_1
\end{pmatrix}
\]

\[
= \begin{pmatrix}
A_1^T & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & A_m^T & 0 \\
0 & \cdots & 0 & I_l
\end{pmatrix}
\begin{pmatrix}
(1 + \tau)\beta I_l & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & (1 + \tau)\beta I_l & 0 \\
0 & \cdots & 0 & A_m
\end{pmatrix}

\begin{pmatrix}
A_1 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & A_m & 0 \\
0 & \cdots & 0 & I_l
\end{pmatrix},
\]

then we have

\[
(w - \tilde{w}^k)^T \Phi(w^k - \tilde{w}^k) = (w - \tilde{w}^k)^T P^T Q P (w^k - \tilde{w}^k) =: (v - \tilde{v}^k)Q(v^k - \tilde{v}^k),
\]

where

\[
P = \begin{pmatrix}
A_1 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & A_m & 0 \\
0 & \cdots & 0 & I_l
\end{pmatrix}, \quad v := Pw \quad \text{and} \quad Q = \begin{pmatrix}
(1 + \tau)\beta I_l & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & (1 + \tau)\beta I_l & 0 \\
-I_l & \cdots & -I_l & \frac{1}{\beta} I_l
\end{pmatrix}.
\] (2.12)
Refer to [27, 28], in the later sections, we will show \( v^k - \tilde{v}^k \) goes to 0, i.e., it follows from (2.7) that \( \tilde{w}^k \) would be a solution point. On the one hand, suppose \( H \) is an arbitrary symmetric and positive definite matrix, we consider a general \( H \)-norm. If \( w^* = (x_1^*, \cdots, x_m^*, \lambda^*) \) is a solution point of problem (1.1), our aim for solving problem (1.1) is equivalent to solving

\[
\min \left\{ \frac{1}{2} \| w - w^* \|_2^2 \mid w \in \Omega \right\}.
\]

Furthermore, we denote \( V := \{(A_1x_i, \cdots, A_mx_m, \lambda) \mid x_i \in \mathcal{X}_i, i = 1, \cdots, m; \lambda \in \mathbb{R}^l \} \). Since each \( A_i \) \( (i = 1, \cdots, m) \) is full column rank, so is \( P \), then (2.13) is also equivalent to the minimization problem

\[
\min \left\{ \frac{1}{2} \| v - v^* \|_H^2 \mid v \in V \right\}.
\]

On the other hand, noting that for the prediction step (2.5), its initialization only needs \( A_i x_i^k \) and \( \lambda^k \), thus we can regard \( A_i x_i^k \) as a whole. Take the above two factors into consideration, our aim in this paper is to obtain a point of the pseudo solution set \( V^* = \{(A_1x_1^1, A_2x_2^2, \cdots, A_mx_m^*, \lambda^*) \} \).

### 2.4 A basic assumption

Throughout this paper, in order to obtain the convergence of the proposed method, we assume the matrix \( Q^T + Q \), i.e., the matrix

\[
D := Q^T + Q = \begin{pmatrix}
2(1+\tau)\beta I_l & \cdots & 0 & -I_l \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 2(1+\tau)\beta I_l & -I_l \\
-I_l & \cdots & -I_l & \frac{2\beta}{\tau} I_l
\end{pmatrix}
\]

(2.14)
is positive definite. It is obvious that \( D \succ 0 \) provided \( \tau \geq \frac{1}{4}(m-4) \). Especially, the regularization term coefficient \( \tau \) can be negative when \( m < 4 \). For example, if \( m = 3 \), the proposed method is convergent when \( \tau > -\frac{1}{4} \).

### 3 Algorithm

#### 3.1 Parallel splitting step (2.5) provides an intrinsic ascent direction

In this subsection, we show the parallel splitting step (2.5) provides an ascent search direction to pseudo solution set \( V^* \). According to the VI-structures of (2.5), it satisfies the following VI-structure:

\[
\begin{align*}
\tilde{w}^k & \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^TF(\tilde{w}^k) \geq (v - \tilde{v}^k)^TQ(v^k - \tilde{v}^k), \quad \forall w \in \Omega.
\end{align*}
\]

(3.15)

Without loss of generality, we assume \( v^k \neq \tilde{v}^k \) throughout our discussion, otherwise, \( \tilde{w}^k \) would be an optimal solution by (2.4). Furthermore, if we set \( w = w^* \) in (3.15), combining with the monotonicity of \( F \), we have

\[
(v^k - v^*)^TQ(v^k - v^*) \geq \theta(x^*) - \theta(\tilde{x}^k) + (w^k - w^*)^TF(\tilde{w}^k) \geq \theta(x^*) - \theta(\tilde{x}^k) + (\tilde{w}^k - w^*)^TF(w^*) \geq 0.
\]

Furthermore, noting that \( \tilde{v}^k - v^* = \tilde{v}^k - v^k + v^k - v^* \), the above inequality can be rewritten as

\[
(v^k - v^*)^TQ(v^k - \tilde{v}^k) \geq \frac{1}{2} \| v^k - \tilde{v}^k \|_2^2 := \frac{1}{2}(v^k - \tilde{v}^k)^T(Q^T + Q)(v^k - \tilde{v}^k).
\]

(3.16)
Taking the nonsingularity of $Q$ into consideration, we consider a new balanced norm matrix, which is defined as 
\[ H = QD^{-1}Q^T, \quad \text{where} \quad D = (Q^T + Q). \] (3.17)
Then we can rewrite (3.16) as 
\[ (v^k - v^*)^T Q(v^k - \tilde{v}^k) = \left( \nabla \left( \frac{1}{2} \| v - v^* \|^2_H \right) \right)^T H^{-1} Q (v^k - \tilde{v}^k) \geq \frac{1}{2} \| v^k - \tilde{v}^k \|^2_D. \] (3.18)
This indicates the direction 
\[ d(v^k, \tilde{v}^k) := H^{-1} Q(v^k - \tilde{v}^k) = Q^{-T} D(v^k - \tilde{v}^k) \]
is an ascent direction of the function $\frac{1}{2} \| v - v^* \|^2_H$ at the point $v^k$. It naturally captures the basic line search method in optimization.

### 3.2 How to determine the search step size?

Similar with [44], we establish our new algorithm by a standard line search method in numerical optimization, see details in, e.g., [38]. To begin with, we do a linear search step:
\[ v^{k+1} = v^k - \alpha d(v^k, \tilde{v}^k). \]
In order to determine an optimal step size $\alpha^*$, noting that 
\[ \| v^{k+1} - v^* \|^2_H = \| v^k - \tau Q^{-T} D(v^k - \tilde{v}^k) - v^* \|^2_H \]
\[ = \| v^k - v^* \|^2_H - 2\alpha (v^k - v^*)^T Q(v^k - \tilde{v}^k) + \tau^2 \| v^k - \tilde{v}^k \|^2_D \]
\[ \leq \| v^k - v^* \|^2_H - \alpha \| v^k - \tilde{v}^k \|^2_D + \alpha^2 \| v^k - \tilde{v}^k \|^2_D. \] (3.16)
Therefore, we obtain a basic inequality 
\[ \| v^k - v^* \|^2_H - \| v^{k+1} - v^* \|^2_H \geq q(\alpha) := \alpha \| v^k - \tilde{v}^k \|^2_D - \alpha^2 \| v^k - \tilde{v}^k \|^2_D. \]
By maximizing the lower bound quadratic contraction function $q(\alpha)$, it follows 
\[ \alpha^* = \frac{\| v^k - \tilde{v}^k \|^2_D}{2 \| v^k - v^* \|^2_D} = \frac{1}{2}. \]
For the matrix $M$ defined in (2.11), we can check $M = Q^{-T} D$. Noting that $q(\alpha)$ is a lower bounded quadratic contraction function, similar with [29], we introduce a relaxation factor $\gamma \in (0, 2)$, thus $\nu := \frac{\gamma}{2} \in (0, 1)$, and the correction step then reads as 
\[ v^{k+1} = v^k - \frac{\gamma}{2} Q^{-T} D (v^k - \tilde{v}^k) = v^k - \nu M (v^k - \tilde{v}^k). \] (3.19)
At the same time, noting that (3.19) is equivalent to 
\[ Q^T (v^{k+1} - v^k) = \nu D (\tilde{v}^k - v^k), \]
and $Q$ is a lower triangular matrix, then the correction step (3.19) is similar with the Gaussian back substitution step of GADMM. Hence, we also interpret (2.6) as a Gaussian-like back substitution of
Since the scheme (1.4)-(1.5) is equivalent to (2.5)-(2.6). Therefore, in specific computations, the novel method can be achieved by the following novel algorithm:

**Algorithm 1** PSALMM

**Input:** the initial point \((x_0^1, x_0^2, \cdots, x_0^m, \lambda^0)\) and the step size \(\nu \in (0,1)\).

**Output:** \((\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_m^k, \tilde{\lambda}^k)\)

**Convergence condition:** \(\beta > 0\) and \(\tau > \frac{1}{4}(m-4)\).

1: for \(k = 1, \cdots, \text{Max-iteration} \) do

   \[
   \begin{align*}
   x_1^{k+1} &= \arg\min \{ L_\beta(x_1, x_2^k, \cdots, x_{m-1}^k, x_m^k, \lambda^k) + \frac{\tau}{2}\beta \| A_1(x_1 - x_1^k) \|_H^2 \mid x_1 \in X_1 \}, \\
   x_2^{k+1} &= \arg\min \{ L_\beta(x_1^k, x_2, \cdots, x_{m-1}^k, x_m^k, \lambda^k) + \frac{\tau}{2}\beta \| A_2(x_2 - x_2^k) \|_H^2 \mid x_2 \in X_2 \}, \\
   & \quad \vdots \\
   x_m^{k+1} &= \arg\min \{ L_\beta(x_1^k, x_2^k, \cdots, x_{m-1}^k, x_m, \lambda^k) + \frac{\tau}{2}\beta \| A_m(x_m - x_m^k) \|_H^2 \mid x_m \in X_m \}. \\
   \lambda^{k+1} &= \lambda^k - 2\nu\beta \left\{ \sum_{i=1}^m A_i x_i^{k+1} + A_i x_i^{k+1} - b \right\}.
   \end{align*}
   \]

Gaussian-like back substitution step:

\[
\begin{align*}
&\text{for } i = 1, \cdots, m, \text{ do}:

&\quad A_i x_i^{k+1} := A_i x_i^k - \nu \left\{ 2A_i(x_i^k - x_i^{k+1}) + \frac{1}{1+\tau}(\sum_{i=1}^m A_i x_i^{k+1} - b) \right\},
\end{align*}
\]

2: while the stop rule satisfied do

3: return \((x_1^{k+1}, x_2^{k+1}, \cdots, x_m^{k+1}, \lambda^{k+1}) := (\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_m^k, \tilde{\lambda}^k)\)

---

### 4 Convergence

Refer to [27] and references cited therein, we establish the convergence analysis of novel scheme (2.5)-(2.6) in this section. Furthermore, we show the novel method has a global convergence and a worst-case \(O(1/t)\) convergence rate.

#### 4.1 Global convergence

The convergence analysis of the proposed scheme is based on the following key lemma.

**Lemma 4.1** Let \(\{\tilde{u}^k\}\) and \(\{v^k\}\) be the sequence generated by the proposed scheme (2.5)-(2.6) for problem (1.1). Then for any step size \(\nu \in (0,1)\), it satisfies the following Féjer monotone property:

\[
\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \geq \nu(1-\nu)\|v^k - \tilde{v}^k\|_D^2.
\] (4.1)
Furthermore, according to (4.3), we have
\[
\|v^{k+1} - v^*\|_H^2 \\
= \|v^k - v^*\|_H^2 - 2\nu(v^k - v^*)^T Q(v^k - \tilde{v}^k) + \nu^2 \|v^k - \tilde{v}^k\|_D^2 \\
\geq \|v^k - v^*\|_H^2 - \nu\|v^k - \tilde{v}^k\|_D^2 + \nu^2 \|v^k - \tilde{v}^k\|_D^2 \\
= \|v^k - v^*\|_H^2 - \nu(1 - \nu)\|v^k - \tilde{v}^k\|_D^2.
\]

This indicates the sequence
\[
\{v^k\} \rightarrow w^* \in \mathcal{V}^* \quad \text{and} \quad w^k \rightarrow w^\infty \in \Omega^*.
\]

**Theorem 4.1** Let \{\tilde{w}^k\} and \{v^k\} be the sequence generated by the proposed scheme (2.5)-(2.6) for problem (1.1). Then for any step size \(\nu \in (0, 1)\), we have
\[
v^k \rightarrow v^\infty \in \mathcal{V}^* \quad \text{and} \quad w^k \rightarrow w^\infty \in \Omega^*.
\]

**Proof** According to inequality (4.1), the generated sequence \{v^k\} satisfies
\[
\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \nu(1 - \nu)\|v^k - \tilde{v}^k\|_D^2 \\
\leq \|v^k - v^*\|_H^2 \leq \cdots \leq \|v^0 - v^*\|_H^2.
\]

Consequently, the generated sequence \{v^k\} is bounded. By adding (4.1) from \(k = 0, 1, \cdots, \infty\), it follows that
\[
\sum_{k=0}^{\infty} \nu(1 - \nu)\|v^k - \tilde{v}^k\|_D^2 \leq \sum_{k=0}^{\infty} (\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2) \leq \|v^0 - v^*\|_H^2,
\]

and hence we obtain
\[
\lim_{k \rightarrow \infty} \|v^k - \tilde{v}^k\|_D^2 = 0. \tag{4.2}
\]

This indicates the sequence \{\tilde{v}^k\} is also bounded. Due to \(P\) is full column rank, it follows from the principle of norm equivalence that \(\tilde{w}^k\) is also bounded. Using the compactness of the closed bounded set in Euclidean space, there exists a subsequence \{\tilde{w}^{kj}\} converges to \(w^\infty\). Noting \(D\) is assumed positive definite, and combining with (4.2), we have
\[
\lim_{j \rightarrow \infty} \|v^{kj} - v^\infty\|_D = 0. \tag{4.3}
\]

Due to (4.2) and the matrix \(Q\) is not singular, it follows from the continuity of \(\theta(x)\) and \(F(w)\) that
\[
w^\infty \in \Omega, \quad \theta(x) - \theta(x^\infty) + (w - w^\infty)^T F(w^\infty) \geq 0, \quad \forall w \in \Omega,
\]

which means \(w^\infty \in \Omega^*\) is a solution point and thus \(v^\infty \in \mathcal{V}^*\). Then by (4.1), we also have
\[
\|v^{k+1} - v^\infty\|_H^2 \leq \|v^k - v^\infty\|_H^2. \tag{4.4}
\]

Furthermore, according to (4.3), we have \(\lim_{k \rightarrow \infty} \|v^k - v^\infty\|_D = 0\). Therefore, we obtain \(\lim_{k \rightarrow \infty} v^k = v^\infty\), and \(w^k \rightarrow w^\infty\) because the matrix \(P\) in (2.12) is full column rank. This completes the proof. \(\square\)
4.2 Convergence rate

4.2.1 $O(1/t)$ convergence rate in an ergodic sense

Similar with [7, 8], at first, we need to define a suitable approximate-solution of the variational inequality (2.4). Given $\epsilon > 0$, as described in [12], $\tilde{w}$ is a $\epsilon$-approximate solution of (2.4) means

$$\tilde{w} \in \Omega, \quad \theta(x) - \theta(\tilde{x}) + (w - \tilde{w})^T F(w) \geq -\epsilon, \quad \forall w \in D_{\tilde{w}}, \tag{4.5}$$

in which

$$D_{\tilde{w}} = \{ w \in \Omega \mid \| w - \tilde{w} \| \leq 1 \}.$$ 

Multiplying both sides by $-1$ in (4.5), we have

$$\tilde{w} \in \Omega, \quad \theta(\tilde{x}) - \theta(x) + (\tilde{w} - w)^T F(w) \leq \epsilon, \quad \forall w \in D_{\tilde{w}}. \tag{4.6}$$

Next, our goal is showing that after $t$ iteration times, we can find an approximate solution $\tilde{w}_t \in \Omega$ satisfies

$$\sup_{w \in D_{\tilde{w}_t}} \left\{ \theta(\tilde{x}_t) - \theta(x) + (\tilde{w}_t - w)^T F(w) \right\} \leq O(1/t). \tag{4.7}$$

**Lemma 4.2** Let $\{\tilde{w}^k\}$ and $\{v^k\}$ be the sequence generated by the proposed scheme (2.5)-(2.6) for problem (1.1). Then for any step size $\nu \in (0, 1)$, we have

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq \frac{1}{2\nu} \left\{ \| v - v^{k+1} \|^2_H - \| v - v^k \|^2_H \right\} + \frac{1}{2} (1 - \nu) \| v^k - \tilde{v}^k \|^2_D. \tag{4.8}$$

**Proof** To begin with, according to (3.15) and (3.19), we have

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq \frac{1}{\nu} (v - \tilde{v}^k)^T QD^{-1}Q^T (v^k - v^{k+1}). \tag{4.9}$$

Noting that $H = QD^{-1}Q^T$ in (3.17) is symmetric and positive definite and combining the identity

$$(a - b)H(c - d) = \frac{1}{2} \left( \| a - d \|^2_H - \| a - c \|^2_H \right) + \frac{1}{2} \left( \| c - b \|^2_H - \| d - b \|^2_H \right), \tag{4.10}$$

We obtain

$$\frac{1}{2} \left\{ \| v^k - \tilde{v}^k \|^2_H - \| v^k - v^{k+1} \|^2_H \right\} = \frac{1}{2} \left\{ \| v^k - \tilde{v}^k \|^2_H - \| v^k - v^{k+1} - \tilde{v}^k \|^2_H \right\}. \tag{4.11}$$

if we set $a = v$, $b = \tilde{v}^k$, $c = v^k$ and $d = v^{k+1}$ in (4.8). For the second term of (4.9), we have

$$\frac{1}{2} \left\{ \| v^k - \tilde{v}^k \|^2_H - \| v^k - \tilde{v}^k \|^2_D \right\} = \frac{1}{2} \left\{ \nu \| v^k - \tilde{v}^k \|^2_D - \nu \| v^k - \tilde{v}^k \|^2_D \right\} = \frac{1}{2} \nu (1 - \nu) \| v^k - \tilde{v}^k \|^2_D, \tag{4.12}$$

and thus we obtain

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq \frac{1}{2\nu} \left( \| v - v^{k+1} \|^2_H - \| v - v^k \|^2_H \right) + \frac{1}{2} \left( \| v^k - \tilde{v}^k \|^2_H - \| v^{k+1} - \tilde{v}^k \|^2_H \right)$$

$$= \frac{1}{2\nu} \left( \| v - v^{k+1} \|^2_H - \| v - v^k \|^2_H \right) + \frac{1}{2} \nu (1 - \nu) \| v^k - \tilde{v}^k \|^2_D.$$

$$= \frac{1}{2\nu} \left( \| v - v^{k+1} \|^2_H - \| v - v^k \|^2_H \right) + \frac{1}{2} (1 - \nu) \| v^k - \tilde{v}^k \|^2_D.$$
This completes the proof of Lemma 4.2.

According to the inequality (4.7) and the monotonicity of the operator $F$, and hence we obtain

$$2
\nu\left\{ \theta(x) - \theta(\bar{x}^k) + (w - \bar{w}^k)^TF(w) \right\} \geq \left\{ \|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2 \right\}.$$  \hspace{1cm} (4.11)

Based on the inequality (4.11), we have the following theorem.

**Theorem 4.2** Let $\{\bar{w}^k\}$ and $\{v^k\}$ be the sequence generated by the prediction step of proposed scheme (2.5)-(2.6) for problem (1.1) and we define $\bar{w}_t = \frac{1}{t+1} \{ \sum_{k=0}^{t} \bar{w}^k \}$.

Then for any positive integer $t$ and step size $\nu \in (0, 1)$, it follows that

$$\bar{w}_t \in \Omega, \quad \theta(\bar{x}_t) - \theta(x) + (\bar{w}_t - w)^TF(w) \leq \frac{1}{2\nu(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega.$$  \hspace{1cm} (4.12)

**Proof** At first, by summing (4.11) over $k = 0, \cdots, t$, it follows that

$$\sum_{k=0}^{t} 2\nu\left\{ \theta(\bar{x}^k) - \theta(x) + (\bar{w}^k - w)^TF(w) \right\} \leq \|v - v^0\|_H^2.$$  \hspace{1cm} (4.13)

By dividing $2\nu(t+1)$ both sides, we have

$$\left\{ \frac{1}{t+1} \sum_{k=0}^{t} \theta(\bar{x}^k) \right\} - \theta(x) + (\bar{w}^k - w)^TF(w) \leq \frac{1}{2\nu(t+1)}\|v - v^0\|_H^2.$$  \hspace{1cm} (4.14)

Meanwhile, according to the definition of $\bar{w}_t$ and the convexity of $\theta(x)$, we know

$$\theta(\bar{x}_t) \leq \frac{1}{t+1} \left\{ \sum_{k=0}^{t} \theta(\bar{x}^k) \right\}.$$  \hspace{1cm} (4.14)

By substituting it in (4.13), the assertion of this theorem follows directly. \hspace{1cm} □

With the help of the preceding lemmas and theorems, now we can show the ergodic-sense convergence rate of the proposed method. By (4.12), it is easy to see that

$$\theta(\bar{x}_t) - \theta(x) + (\bar{w}_t - w)^TF(w) \leq \frac{d}{2\nu(t+1)} = O\left(\frac{1}{t}\right)$$

if we set $d = \sup_{w \in D_{\bar{w}_t}}\{ \|v - v^0\|_H^2 \}$. Consequently, a $O(1/t)$ convergence rate in an ergodic sense is established. Hence, the novel methods enjoy a global convergence and a sub-linear convergence rate.

4.3 $O(1/t)$ convergence rate in a point-wise sense

Similar with [27], our aim in this subsection is to show the $O(1/t)$ convergence rate of the proposed method in a point-wise sense. In view of (3.15), we want to show the tail $\|Q(v^t - \tilde{v}^t)\|_H^2 \leq \frac{c_0}{t+1} = O\left(\frac{1}{t}\right)$, where $c > 0$ is a constant. Next, we only need to show

$$\|M(v^k - \tilde{v}^k)\|_H^2 \leq \frac{c_0}{t+1} = O\left(\frac{1}{t}\right),$$

by using the norm equivalence principle in Euclidean space, in which $c_0 > 0$ is a constant.
Lemma 4.3 Let \( \{\tilde{w}^k\} \) and \( \{v^k\} \) be the sequence generated by the proposed scheme (2.5)-(2.6) for problem (1.1). Then for any step size \( \nu \in (0,1) \), we have

\[
(v^k - \tilde{v}^k)^T M^T H M \{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \frac{1}{2\nu} \| (v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \|^2_D. \tag{4.14}
\]

Moreover, we have

\[
\| M(v^k - \tilde{v}^k) \|^2_H \leq \| M(v^{k+1} - \tilde{v}^{k+1}) \|^2_H, \quad \forall \ k > 0. \tag{4.15}
\]

Proof We consider VI structure of \( k \)-th and \( (k+1) \)-th prediction steps:

\[
\begin{align*}
&\theta(u) - \theta(\tilde{a}^k) + (w - \tilde{w}^k)^T F(\tilde{a}^k) \geq (v - \tilde{v}^k)^T Q(v - \tilde{v}^k), \\
&\theta(u) - \theta(\tilde{a}^{k+1}) + (w - \tilde{w}^{k+1})^T F(\tilde{a}^{k+1}) \geq (v - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}).
\end{align*}
\]

Plug \( \tilde{w}^{k+1} \) and \( \tilde{w}^k \) into above VIs respectively, by adding them together and using the monotonicity of \( F \), we have

\[
(\tilde{v}^k - \tilde{v}^{k+1})^T (v - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \geq 0.
\]

Adding \( \{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}^T Q \{ (v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \} \) both sides and by \( Q = HM \), we have

\[
(\nu M(v^k - \tilde{v}^k))^T Q \{ (v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \} = \nu (v^k - \tilde{v}^k)^T M^T H M \{ (v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \}
\]

\[
\geq \frac{1}{2} \| (v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \|^2_D.
\]

It directly follows the first assertion by dividing \( \nu \) both sides. Next, using the identity

\[
\|a\|^2_H - \|b\|^2_H = 2a^T H (a - b) - \|a - b\|^2_H
\]

for \( a = M(v^k - \tilde{v}^k) \) and \( b = M(v^{k+1} - \tilde{v}^{k+1}) \), we obtain

\[
\| M(v^k - \tilde{v}^k) \|^2_H - \| M(v^{k+1} - \tilde{v}^{k+1}) \|^2_H
\]

\[
= 2(v^k - \tilde{v}^k)^T M^T H M \{ (v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \} - \| M \{ (v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \} \|^2_H
\]

\[
\geq \frac{1}{2} \| (v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \|^2_D - \| M \{ (v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \} \|^2_H
\]

\[
= (\nu - 1) \| (v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \|^2_D \geq 0.
\]

This is the second assertion and thus we complete the proof.

\[\square\]

\[\textbf{Theorem 4.3} \]
Let \( \{\tilde{w}^k\} \) and \( \{v^k\} \) be the sequence generated by the proposed scheme (2.5)-(2.6) for problem (1.1). Then for any positive integer \( t \) and step size \( \nu \in (0,1) \), we have

\[
\| v^t - v^{t+1} \|^2_H \leq \frac{c_0}{t + 1} \| v^0 - v^* \|^2_H.
\]

Proof According to (4.1), we have \( \nu (1 - \nu) \| v^k - \tilde{v}^k \|^2_D \leq \| v^k - v^* \|^2_H - \| v^{k+1} - v^* \|^2_H \). By the principle of norm equivalence in Euclidean space, there exists \( c_0 > 0 \), such that

\[
\frac{1}{c_0} \|M(v^k - \tilde{v}^k)\|_H^2 \leq \| v^k - v^* \|_H^2 - \| v^{k+1} - v^* \|_H^2 \tag{4.16}
\]

\[\text{15}\]
Summing (4.16) over \( k = 0, \cdots , t \), it follows that
\[
\sum_{k=0}^{t} \frac{1}{c_0} \|M(v^k - \tilde{v}^k)\|_H^2 \leq \|v^0 - v^*\|_H^2 - \|v^{t+1} - v^*\|_H^2 \leq \|v^0 - v^*\|_H^2.
\]
Furthermore, noting the monotonicity of \( \{\|M(v^k - \tilde{v}^k)\|_H^2\} \), we have
\[
\frac{1}{c_0} (t + 1) \|M(v^t - \tilde{v}^t)\|_H^2 \leq \sum_{k=0}^{t} \frac{1}{c_0} \|M(v^k - \tilde{v}^k)\|_H^2 \leq \|v^0 - v^*\|_H^2.
\]
Therefore,
\[
\|v^t - v^{t+1}\|_H^2 = \|M(v^t - \tilde{v}^t)\|_H^2 \leq \frac{c_0}{t + 1} \|v^0 - v^*\|_H^2.
\]
This completes the proof of theorem. \( \square \)

With the help of the above theorem, by defining \( d := \sup \{c_0\|v^0 - v^*\|_H^2 \mid v^* \in \mathcal{V}^*\} \), we obtain
\[
\|M(v^t - \tilde{v}^t)\|_H^2 \leq \frac{d}{t + 1},
\]
and consequently a point-wise \( \mathcal{O}(1/t) \) convergence rate is established.

5 Numerical experiments

In this section, we conduct some extensive numerical experiments to check the performance of the proposed method. All codes are written in a MATLAB R2018b version and implemented in a computer with 3.40 GHz Intel Core i7-6700 CPU and 24 GB memory. Considering the parallel efficiency of MATLAB for matrix computation, we still use the serial way to compare different parallel algorithms.

5.1 Test model

The latent variable Gaussian graphical model selection problem (abbreviated as LVGGMS) arising in statistical learning is a typical three-block convex minimization problem [3, 4, 34], which reads as
\[
\min \quad \langle X, C \rangle - \log \det(X) + \nu \|Y\|_1 + \mu \text{tr}(Z)
\]
\[
\quad \text{s.t.} \quad X - Y + Z = 0, \quad Z \succeq 0.
\]
(5.1)

In which \( X, Y, Z \in \mathbb{R}^{n \times n} ; C \in \mathbb{R}^{n \times n} \) denotes the covariance matrix obtained from the observation; \( \nu \) and \( \mu \) represent two given positive weight parameters; \( \|\cdot\|_1 \) is the \( l_1 \)-norm of a matrix and \( \text{tr}(\cdot) \) denotes the trace of a matrix. For problem (5.1), its corresponding augmented Lagrange function is defined as
\[
\mathcal{L}_\beta(X, Y, Z, \Lambda) = \langle X, C \rangle - \log \det(X) + \nu \|Y\|_1 + \mu \text{tr}(Z) - \langle \Lambda, X - Y + Z \rangle + \frac{\beta}{2} \|X - Y + Z\|_F^2.
\]

In the following, taking the Algorithm 1 as an example, we give the specific solution of \( x \)-subproblem:
\[
\begin{aligned}
\hat{X}^k &= \arg \min \left\{ \mathcal{L}_\beta(X^k, Y^k, Z^k, \Lambda^k) + \frac{\tau}{2} \|X - X^k\|_F^2 \mid X \in \mathbb{R}^{n \times n} \right\}, \\
\hat{Y}^k &= \arg \min \left\{ \mathcal{L}_\beta(X^k, Y^k, Z^k, \Lambda^k) + \frac{\tau}{2} \|Y - Y^k\|_F^2 \mid Y \in \mathbb{R}^{n \times n} \right\}, \\
\hat{Z}^k &= \arg \min \left\{ \mathcal{L}_\beta(X^k, Y^k, Z, \Lambda^k) + \frac{\tau}{2} \|Z - Z^k\|_2^2 \mid Z \succeq 0, Z \in \mathbb{R}^{n \times n} \right\}.
\end{aligned}
\]
For $X$-subproblem in (5.2), ignoring the constant term, we need to solve
\[
\min \left\{ (X,C) - \log \det(X) + \frac{\beta}{2} \|X - Y^k + Z^k - \frac{1}{\beta} \Lambda^k\|_F^2 + \frac{\tau}{2} \beta \|X - X^k\|_F^2 \mid X \in \mathbb{R}^{n \times n} \right\}.
\]
According to the first-order optimal condition, we have
\[
C - X^{-1} + \beta (X - Y^k + Z^k - \frac{1}{\beta} \Lambda^k) + \tau \beta (X - X^k) = 0.
\]
Multiplying both sides by $X$, we obtain
\[
(1 + \tau) \beta X^2 + (C - \tau \beta X^k - \beta Y^k + \beta Z^k - \Lambda^k) X - I = 0. \tag{5.3}
\]
If we do an eigenvalue decomposition for the matrix $C - \tau \beta X^k - \beta Y^k + \beta Z^k - \Lambda^k$, in other words, there exists a diagonal matrix $D$ and an unitary matrix $U$ satisfying
\[
UDU^T = C - \tau \beta X^k - \beta Y^k + \beta Z^k - \Lambda^k. \tag{5.4}
\]
If we plug (5.4) back into (5.3) and set $P = U^T X U$, it follows that $(1 + \tau) \beta PP + DP - I = 0$. Therefore, we have
\[
P_{ii} = \frac{1}{2 \beta (1 + \tau)} \left( -D_{ii} + \sqrt{D_{ii}^2 + 4(1 + \tau) \beta} \right)
\]
is a solution of (5.3). For $Y$-subproblem, due to
\[
\tilde{Y}^k = \arg \min \left\{ \nu \|Y\|_1 + \frac{\beta}{2} \|X^k - Y + Z^k - \frac{1}{\beta} \Lambda^k\|_F^2 + \frac{\tau}{2} \beta \|Y - Y^k\|_F^2 \mid Y \in \mathbb{R}^{n \times n} \right\}
\]
\[
= \arg \min \left\{ \|Y\|_1 + \frac{1}{2 \nu} \beta (1 + \tau) \|Y - \frac{1}{1 + \tau} (X^k + \tau Y^k + Z^k - \frac{1}{\beta} \Lambda^k)\|_F^2 \mid Y \in \mathbb{R}^{n \times n} \right\}.
\]
Similar with [42], we can solve it by the soft shrinkage operator. Finally, for $Z$-subproblem, since
\[
\tilde{Z}^k = \arg \min \left\{ \mu \text{tr}(Z^2) + \frac{\beta}{2} \|X^k - Y^k + Z - \frac{1}{\beta} \Lambda^k\|_F^2 + \frac{\tau}{2} \beta \|Z - Z^k\|_F^2 \mid Z \succeq 0 \right\}
\]
\[
= \arg \min \left\{ \frac{1}{2} \beta (1 + \tau) \|Z - \frac{1}{1 + \tau} (-X^k + Y^k + \tau Z^k + \frac{1}{\beta} \Lambda^k - \frac{\mu}{\beta} I)\|_2^2 \mid Z \succeq 0 \right\}.
\]
Suppose $VD_1 V^T$ is the eigenvalue decomposition of
\[
\frac{1}{1 + \tau} (-X^k + Y^k + \tau Z^k + \frac{1}{\beta} \Lambda^k - \frac{\mu}{\beta} I),
\]
where $D_1$ is a diagonal matrix and $V$ is a unitary matrix. Then $\tilde{Z}^k = V \cdot \max(D_1, 0) \cdot V^T$ would be a solution of $Z$-subproblem.

In the following, we list the numerical performance of the proposed method (1.4)-(1.5). Meanwhile, we also give a comparison to other parallel splitting algorithms, including a parallel augmented Lagrange method (denoted by PALM) proposed in [21], a splitting ADMM-like method (denoted by PSADMM) in [28]. As a reference, we also list the performance of the Gaussian ALM-based methods, including the ADMM with Gaussian back substitution method (denoted by GADMM) in [26] and the direct extension of ADMM (denoted by DADMM).
5.2 Numerical results

In our experiment, as described in [3], we take $\nu = 0.005$, $\mu = 0.05$, $n = 100$ and the initial values are set to $(X^0, Y^0, Z^0, \Lambda^0) = (I, 2I, I, 0)$. Moreover, the covariance matrix $C$ is randomly generated by the following code:

```matlab
1 randn(seed,0); rand(seed,0); n=m; N=10*n;
2 Sinv=diag(abs(ones(n,1))); idx=randsample(n^2,0.001*n^2);
3 Sinv(idx)=ones(numel(idx),1); Sinv=Sinv+Sinv;
4 if min(eig(Sinv))<0
5 Sinv=Sinv+1.1*abs(min(eig(Sinv)))*eye(n);
6 end
7 S=inv(Sinv);
8 DD=mvnrnd(zeros(1,n),S,N); C=cov(DD);
```

Furthermore, we take three stop rules in our experiment, including the iterative error (Itr-E), the absolute error (Abs-E) and linear constraint error (CER), which are defined by

$$\text{Itr-E}(k) := \max \{ \| X^k - X^{k+1} \|_\infty, \| Y^k - Y^{k+1} \|_\infty, \| Z^k - Z^{k+1} \|_\infty \},$$

$$\text{Abs-E}(k) := \sqrt{\| X^k - X^* \|_F^2 + \| Y^k - Y^* \|_F^2 + \| Z^k - Z^* \|_F^2},$$

$$\text{CER}(k) := \| X^k - Y^k + Z^k \|_F.$$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>DADMM</th>
<th>GADMM</th>
<th>PSADMM</th>
<th>PALM</th>
<th>PSALMM($\tau = 0$)</th>
<th>PSALMM($\tau = 1/3$)</th>
<th>PSALMM($\tau = 1.001$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>151</td>
<td>151</td>
<td>160</td>
<td>523</td>
<td>344</td>
<td>206</td>
<td>156</td>
</tr>
<tr>
<td>0.06</td>
<td>131</td>
<td>131</td>
<td>138</td>
<td>465</td>
<td>293</td>
<td>178</td>
<td>133</td>
</tr>
<tr>
<td>0.07</td>
<td>117</td>
<td>116</td>
<td>121</td>
<td>415</td>
<td>257</td>
<td>157</td>
<td>116</td>
</tr>
<tr>
<td>0.08</td>
<td>105</td>
<td>104</td>
<td>121</td>
<td>376</td>
<td>232</td>
<td>140</td>
<td>103</td>
</tr>
<tr>
<td>0.09</td>
<td>96</td>
<td>95</td>
<td>155</td>
<td>346</td>
<td>210</td>
<td>126</td>
<td>94</td>
</tr>
<tr>
<td>0.10</td>
<td>89</td>
<td>88</td>
<td>181</td>
<td>323</td>
<td>193</td>
<td>115</td>
<td>111</td>
</tr>
<tr>
<td>0.11</td>
<td>83</td>
<td>82</td>
<td>205</td>
<td>349</td>
<td>178</td>
<td>106</td>
<td>132</td>
</tr>
<tr>
<td>0.12</td>
<td>77</td>
<td>76</td>
<td>227</td>
<td>409</td>
<td>165</td>
<td>98</td>
<td>150</td>
</tr>
<tr>
<td>0.13</td>
<td>73</td>
<td>72</td>
<td>249</td>
<td>458</td>
<td>154</td>
<td>92</td>
<td>166</td>
</tr>
<tr>
<td>0.14</td>
<td>69</td>
<td>69</td>
<td>270</td>
<td>504</td>
<td>144</td>
<td>97</td>
<td>182</td>
</tr>
<tr>
<td>0.15</td>
<td>67</td>
<td>79</td>
<td>290</td>
<td>547</td>
<td>135</td>
<td>114</td>
<td>197</td>
</tr>
<tr>
<td>0.16</td>
<td>78</td>
<td>88</td>
<td>310</td>
<td>588</td>
<td>128</td>
<td>128</td>
<td>212</td>
</tr>
<tr>
<td>0.17</td>
<td>87</td>
<td>96</td>
<td>330</td>
<td>628</td>
<td>121</td>
<td>140</td>
<td>227</td>
</tr>
<tr>
<td>0.18</td>
<td>94</td>
<td>103</td>
<td>350</td>
<td>668</td>
<td>117</td>
<td>151</td>
<td>242</td>
</tr>
<tr>
<td>0.19</td>
<td>101</td>
<td>111</td>
<td>369</td>
<td>706</td>
<td>113</td>
<td>162</td>
<td>256</td>
</tr>
<tr>
<td>0.20</td>
<td>108</td>
<td>118</td>
<td>388</td>
<td>745</td>
<td>109</td>
<td>173</td>
<td>270</td>
</tr>
<tr>
<td>0.25</td>
<td>139</td>
<td>150</td>
<td>483</td>
<td>929</td>
<td>150</td>
<td>223</td>
<td>339</td>
</tr>
</tbody>
</table>

Take the correction step into consideration, we also give the performance of stop rules with CPU time. Due to LVGGMS is a typical three-block convex optimization programming, we need to give
the convergence condition of the tested methods:

Table 2: The first iteration of different methods for LVGGMS with $\text{Abs-E}(k) < 10^{-9}$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>DADMM</th>
<th>GADMM</th>
<th>PSADMM</th>
<th>PALM</th>
<th>PSALMM($\tau = 0$)</th>
<th>PSALMM($\tau = 1/3$)</th>
<th>PSALMM($\tau = 1.001$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>207</td>
<td>207</td>
<td>216</td>
<td>826</td>
<td>529</td>
<td>307</td>
<td>220</td>
</tr>
<tr>
<td>0.06</td>
<td>176</td>
<td>176</td>
<td>183</td>
<td>704</td>
<td>441</td>
<td>259</td>
<td>184</td>
</tr>
<tr>
<td>0.07</td>
<td>155</td>
<td>154</td>
<td>159</td>
<td>619</td>
<td>381</td>
<td>224</td>
<td>158</td>
</tr>
<tr>
<td>0.08</td>
<td>137</td>
<td>136</td>
<td>153</td>
<td>551</td>
<td>337</td>
<td>196</td>
<td>139</td>
</tr>
<tr>
<td>0.09</td>
<td>124</td>
<td>122</td>
<td>198</td>
<td>501</td>
<td>302</td>
<td>175</td>
<td>125</td>
</tr>
<tr>
<td>0.10</td>
<td>113</td>
<td>112</td>
<td>234</td>
<td>463</td>
<td>274</td>
<td>158</td>
<td>140</td>
</tr>
<tr>
<td>0.11</td>
<td>105</td>
<td>104</td>
<td>266</td>
<td>477</td>
<td>250</td>
<td>144</td>
<td>167</td>
</tr>
<tr>
<td>0.12</td>
<td>97</td>
<td>96</td>
<td>297</td>
<td>564</td>
<td>229</td>
<td>132</td>
<td>191</td>
</tr>
<tr>
<td>0.13</td>
<td>91</td>
<td>90</td>
<td>328</td>
<td>637</td>
<td>212</td>
<td>123</td>
<td>214</td>
</tr>
<tr>
<td>0.14</td>
<td>86</td>
<td>86</td>
<td>357</td>
<td>704</td>
<td>197</td>
<td>122</td>
<td>236</td>
</tr>
<tr>
<td>0.15</td>
<td>84</td>
<td>97</td>
<td>386</td>
<td>769</td>
<td>184</td>
<td>144</td>
<td>257</td>
</tr>
<tr>
<td>0.16</td>
<td>95</td>
<td>108</td>
<td>415</td>
<td>831</td>
<td>173</td>
<td>161</td>
<td>278</td>
</tr>
<tr>
<td>0.17</td>
<td>106</td>
<td>119</td>
<td>443</td>
<td>892</td>
<td>163</td>
<td>178</td>
<td>299</td>
</tr>
<tr>
<td>0.18</td>
<td>115</td>
<td>129</td>
<td>472</td>
<td>952</td>
<td>154</td>
<td>193</td>
<td>319</td>
</tr>
<tr>
<td>0.19</td>
<td>125</td>
<td>138</td>
<td>500</td>
<td>1011</td>
<td>146</td>
<td>208</td>
<td>339</td>
</tr>
<tr>
<td>0.20</td>
<td>133</td>
<td>147</td>
<td>528</td>
<td>1070</td>
<td>140</td>
<td>223</td>
<td>359</td>
</tr>
<tr>
<td>0.25</td>
<td>174</td>
<td>191</td>
<td>666</td>
<td>1356</td>
<td>192</td>
<td>293</td>
<td>457</td>
</tr>
</tbody>
</table>

- DADMM: $\beta > 0$ and the convergence analysis is missing.
- GADMM: It needs $\beta > 0$, and as suggested in [26], a larger $\alpha \in (0, 1)$ is more efficient, thus we take $\alpha = 0.95$ in our experiment.
- PSADMM: It needs $\beta > 0$, meanwhile, it needs $\tau > 3 - 2 = 1$ and thus we take $\tau = 1.0001$.
- PALM: It needs $\beta > 0$, and as suggested in [28], it needs a larger $\gamma \in (0, 2)$, thus we take $\gamma = 1.95$.
- PSALMM: It needs $\beta > 0$ and $\tau > \frac{3 - 4}{4} = -\frac{1}{4}$. Since $m = 3$, we take $\tau = \frac{1}{4}$ and $\nu = 0.75$.

To begin with, we need to find the most efficient $\beta$s for different methods. Based on three different stop rules, we list some efficient candidates $\beta \in [0.05, 0.25]$ in Table 1, 2 and 3. As can be seen, the DADMM with $\beta = 0.15$ has an more efficient performance for all three stop rules, and the GADMM has a same efficient effect with DADMM. Moreover, these three stop rules have a almost same effect with $\beta$, therefore, in order to comparing different methods, we only list the most efficient parameters of the mentioned algorithms:

- DADMM: $\beta = 0.15$.
- GADMM: $\beta = 0.14$ and Gaussian back substitution factor $\nu = 0.95$.
- PSADMM: $\beta = 0.08$ and the regularization term $\tau = 1.001$.
- PALM: $\beta = 0.10$ and the relaxation factor $\gamma = 1.95$.
- PSALMM: $\beta = 0.13$ and the step size $\nu = 0.75$. 

19
Figure 1: Three types of error curves with the number of iterations
Figure 2: Three types of error curves with CPU time.
Meanwhile, noting that the proposed method is convergent provided \( \tau > -\frac{1}{2} \), we also list the performance of the case \( \tau = 0 \) and \( \tau = 1.001 \) in Table 1, 2 and 3. As can be seen, such two cases is also more efficient than PSADMM and PALM for such a problem. Thus the novel method provides more efficient parameters for other applications is possible.

Table 3: The first iteration of different methods for LVGGMS with CER\((k) < 10^{-9}\).

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>DADMM</th>
<th>GADMM</th>
<th>PSADMM</th>
<th>PALM</th>
<th>PSALMM((\tau = 0))</th>
<th>PSALMM((\tau = 1/3))</th>
<th>PSALMM((\tau = 1.001))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>199</td>
<td>200</td>
<td>207</td>
<td>794</td>
<td>487</td>
<td>257</td>
<td>193</td>
</tr>
<tr>
<td>0.06</td>
<td>170</td>
<td>169</td>
<td>175</td>
<td>676</td>
<td>407</td>
<td>217</td>
<td>162</td>
</tr>
<tr>
<td>0.07</td>
<td>149</td>
<td>148</td>
<td>152</td>
<td>591</td>
<td>351</td>
<td>188</td>
<td>139</td>
</tr>
<tr>
<td>0.08</td>
<td>132</td>
<td>132</td>
<td>142</td>
<td>527</td>
<td>311</td>
<td>165</td>
<td>122</td>
</tr>
<tr>
<td>0.09</td>
<td>119</td>
<td>118</td>
<td>175</td>
<td>477</td>
<td>279</td>
<td>147</td>
<td>108</td>
</tr>
<tr>
<td>0.10</td>
<td>109</td>
<td>108</td>
<td>203</td>
<td>437</td>
<td>252</td>
<td>133</td>
<td>117</td>
</tr>
<tr>
<td>0.11</td>
<td>100</td>
<td>100</td>
<td>227</td>
<td>436</td>
<td>230</td>
<td>121</td>
<td>136</td>
</tr>
<tr>
<td>0.12</td>
<td>93</td>
<td>93</td>
<td>250</td>
<td>499</td>
<td>211</td>
<td>111</td>
<td>153</td>
</tr>
<tr>
<td>0.13</td>
<td>87</td>
<td>86</td>
<td>273</td>
<td>554</td>
<td>195</td>
<td>103</td>
<td>168</td>
</tr>
<tr>
<td>0.14</td>
<td>82</td>
<td>82</td>
<td>294</td>
<td>605</td>
<td>182</td>
<td>101</td>
<td>183</td>
</tr>
<tr>
<td>0.15</td>
<td>79</td>
<td>90</td>
<td>315</td>
<td>653</td>
<td>169</td>
<td>114</td>
<td>198</td>
</tr>
<tr>
<td>0.16</td>
<td>86</td>
<td>100</td>
<td>336</td>
<td>699</td>
<td>159</td>
<td>126</td>
<td>212</td>
</tr>
<tr>
<td>0.17</td>
<td>94</td>
<td>109</td>
<td>356</td>
<td>744</td>
<td>149</td>
<td>137</td>
<td>225</td>
</tr>
<tr>
<td>0.18</td>
<td>101</td>
<td>117</td>
<td>376</td>
<td>787</td>
<td>141</td>
<td>147</td>
<td>239</td>
</tr>
<tr>
<td>0.19</td>
<td>108</td>
<td>126</td>
<td>395</td>
<td>830</td>
<td>134</td>
<td>157</td>
<td>252</td>
</tr>
<tr>
<td>0.20</td>
<td>115</td>
<td>133</td>
<td>415</td>
<td>872</td>
<td>127</td>
<td>167</td>
<td>265</td>
</tr>
</tbody>
</table>

Table 1, 2 and 3 show a comparison between the proposed algorithm and some known methods. As can be seen, the proposed algorithm has a faster convergence rate than some known parallel methods with theoretically guaranteed convergence. We can also find that PSALMM enjoys the efficiency of GADMM and is faster than PSADMM. Meanwhile, the numerical linear convergence rate of the proposed method is intuitively given in Figure 1 and 2 both iterations and CPU time.

### 6 Conclusions

We present a class of new parallel splitting ALM-based method for solving the separable convex minimization problem. By adding a Gaussian-like back substitution step (1.5) to the parallel splitting prediction step (1.4), the novel method need only a reduced parameter of the regularization terms. The new method can be regarded as a hybrid case of GADMM and PSADMM, and it enjoys both advantages of them numerically. Furthermore, the new algorithm has a faster convergent rate compared with some known parallel splitting algorithms for LVGGMS.

### Acknowledgements

The author is greatly indebted to Professor Bingsheng He for many useful discussions and suggestions.
References


[31] He, B.S., Ma, F., Yuan, X.M.: Convergence study on the symmetric version of ADMM with larger step sizes, SIAM. J. Imaging Science 9, 1467-1501 (2016)


[47] Yuan, J., Aaron Fenster.: Modern convex optimization to medical image analysis.