J. Appl. Math. Comput. (2018) 58:151–181
https://doi.org/10.1007/s12190-017-1138-8

ORIGINAL RESEARCH

# Improved proximal ADMM with partially parallel splitting for multi-block separable convex programming

**Min Sun**[1,2] · **Hongchun Sun**[3]

Received: 2 June 2017 / Published online: 20 October 2017


**Abstract** For a type of multi-block separable convex programming raised in machine learning and statistical inference, we propose a proximal alternating direction method of multiplier with partially parallel splitting, which has the following nice properties: (1) to alleviate the weight of the proximal terms, the restrictions imposed on the proximal parameters are relaxed substantively; (2) to maintain the inherent structure of the primal variables $x_i (i = 1, 2, \ldots, m)$, the relaxation parameter $\gamma$ is only attached to the update formula of the dual variable $\lambda$. For the resulted method, we establish its global convergence and worst-case $\mathcal{O}(1/t)$ convergence rate in an ergodic sense, where $t$ is the iteration counter. Finally, three numerical examples are given to illustrate the theoretical results obtained.

**Keywords** Alternating direction method of multipliers · Multi-block separable convex programming · Global convergence

**Mathematics Subject Classification** 90C25 · 90C30

This work is supported by the National Natural Science Foundation of China (Nos. 11671228, 11601475), the foundation of First Class Discipline of Zhejiang-A (Zhejiang University of Finance and Economics-Statistics), the foundation of National Natural Science Foundation of Shandong Province (No. ZR2016AL05), Scientific Research Project of Shandong Universities (No. J15LI11), and the Educational Reform Project of Zaozhuang University (No. 1021402).

✉ Min Sun
ziyouxiaodou@163.com

1 School of Mathematics and Statistics, Zaozhuang University, Zaozhuang 277160, Shandong, China

2 School of Management Science, Qufu Normal University, Rizhao Campus, Qufu 276800, Shandong, China

3 School of Science, Linyi University, Linyi 276005, Shandong, China

## 1 Introduction

In this paper, we consider the linearly constrained multi-block separable convex programming, whose objective function is the sum of $m$ functions with decoupled variables:

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

where $\theta_i : \mathcal{R}^{n_i} \to \mathcal{R}(i = 1, 2, \ldots, m)$ are closed convex functions (not necessarily smooth); $A_i \in \mathcal{R}^{l \times n_i}(i = 1, 2, \ldots, m)$; $b \in \mathcal{R}^l$ and $\mathcal{X}_i \subseteq \mathcal{R}^{n_i}(i = 1, 2, \ldots, m)$ are nonempty closed convex sets. Throughout this paper, the solution set of (1.1) is assumed to be nonempty. Many problems encountered in machine learning and statistical inference can be posed in the model (1.1) [2,20–22,27].

In the last decades, splitting methods for solving large scale (1.1) have been investigated in depth, e.g., the split-Bregman iteration method [18,30], the fixed-point proximity methods [6,19], the alternating direction method of multipliers [13,14,26,28]. The relationship and numerical comparison of the above methods can be found in [19]. In this paper, we are going to study the alternating direction method of multipliers (ADMM), which is originally developed by [10], and has been revisited recently due to its success in the applications of separable convex programming [2,22,23,27]. The $k$th iterative scheme of ADMM for (1.1) with $m = 2$ reads as

$$\begin{cases} x_1^{k+1} = \underset{x_1 \in \mathcal{X}_1}{\operatorname{argmin}} \left\{ \mathcal{L}_\beta \left( x_1, x_2^k; \lambda^k \right) \right\}, \\ x_2^{k+1} = \underset{x_2 \in \mathcal{X}_2}{\operatorname{argmin}} \left\{ \mathcal{L}_\beta \left( x_1^{k+1}, x_2; \lambda^k \right) \right\}, \\ \lambda^{k+1} = \lambda^k - \beta \left( \sum_{i=1}^{2} A_i x_i^{k+1} - b \right), \end{cases} \tag{1.2}$$

where $\beta > 0$, and

$$\mathcal{L}_\beta(x_1, x_2; \lambda) = \sum_{i=1}^{2} \theta_i(x_i) - \left\langle \lambda, \sum_{i=1}^{2} A_i x_i - b \right\rangle + \frac{\beta}{2} \left\| \sum_{i=1}^{2} A_i x_i - b \right\|^2$$

is the augmented Lagrangian function of the model (1.1) with $m = 2$; $\lambda \in \mathcal{R}^l$ is the Lagrangian multiplier. Based on the other's most recent version, the iterative scheme (1.2) updates the primal variables $x_i(i = 1, 2)$ in a sequential way. However, the ADMM is for two-block case and the global convergence cannot be guaranteed if it is directly extended $m$ block case ($m \geq 3$) [5]. Furthermore, it is not suitable for distributed computing, which is particularly welcome for the large-scale (1.1) with big data.

To address the above two issues, in recent years many researchers incorporate the regularization technique and the parallel iterative technique into the iterative scheme (1.2), and propose some proximal ADMM-type methods with fully parallel splitting [7,20] and some ADMM-type methods with partially parallel splitting [15,28], which not only have global convergence under mild conditions, but also are suitable for distributed computing. In [7], Deng, Lai, Peng and Yin proposed a proximal Jacobian ADMM, which regularizes each subproblem by a proximal term $\frac{1}{2}\|x - x_i^k\|_{P_i}^2$, where $P_i$ is the proximal matrix and is often required to be positive semi-definite. Many choices of $P_i$ are presented in [8], e.g., $P_i = \tau_i I_{n_i} - \beta A_i^\top A_i$, which can linearize the third term of the augmented Lagrangian function $\mathcal{L}_\beta(x_1, \ldots, x_m; \lambda) := \sum_{i=1}^m \theta_i(x_i) - \left\langle \lambda, \sum_{i=1}^m A_i x_i - b \right\rangle + \frac{\beta}{2} \left\| \sum_{i=1}^m A_i x_i - b \right\|^2$ and thus the corresponding subproblem in (1.2) often admits closed-form solution in practice [30]. To ensure the global convergence of the method in [7], the proximal parameter $\tau_i$ in $P_i$ need to satisfy the condition $\tau_i > \frac{m\beta}{2-\gamma}\|A_i\|^2$, where $\gamma \in (0, 2)$. Similarly, Lin, Liu and Li [20] presented a linearized alternating direction method with parallel splitting for model (1.1), and to ensure the global convergence of this method, the proximal parameter $\tau_i$ in $P_i$ need to satisfy the condition $\tau_i > m\beta\|A_i\|^2$. Wang and Song [28] developed a twisted proximal ADMM (denoted by TPADMM) with partially parallel splitting, which updates $x_1$ and $x_i (i = 2, 3, \ldots, m)$ in a sequential way, but updates the variables $x_i (i = 2, 3, \ldots, m)$ in a parallel way. Furthermore, if we set $P_i = \tau_i I_{n_i} - \beta A_i^\top A_i$ in [28], then the proximal parameter $\tau_i$ must satisfy the condition $\tau_i > (m-1)\beta\|A_i\|^2$, and the lower bound $(m-1)\beta\|A_i\|^2$ is obviously smaller than $m\beta\|A_i\|^2$ in [20].

Fazel et al. [9] have pointed out that the choice of the proximal matrix $P_i$ is very much problem dependent and the general principle is that $P_i$ should be small as possible while the subproblem related to the variable $x_i$ is still relatively easy to compute. Therefore, the feasible region of the proximal parameter $\tau_i$ deserves researching. In this paper, we are going to study the twisted proximal ADMM (TPADMM) in [28], and show that the greatest lower bound of its proximal parameter can be substantially reduced. Specifically, a sharper lower bound of the proximal parameter $\tau_i$ is $\frac{4+\max\{1-\gamma, \gamma^2-\gamma\}}{5}(m-1)\beta\|A_i\|^2$, where $\gamma \in (0, \frac{1+\sqrt{5}}{2})$. If $\gamma = 1$, this lower bound becomes $\frac{4(m-1)}{5}\beta\|A_i\|^2$, which is obviously smaller than $m\beta\|A_i\|^2$ in [20] and $(m-1)\beta\|A_i\|^2$ in [28]. Furthermore, it is worth noting that the relaxation parameter $\gamma$ is attached to the update formulas of all variables of TPADMM [28]. However, this relaxation strategy may destroy the inherent structure of the primal variables which are stemmed from the subproblems. Then, in our new method, we use the following relaxing technique: only relax the dual variable $\lambda$. That is we only attach the relaxation factor $\gamma$ to the dual variable $\lambda$, while the updating formula of the primal variables $x_i (i = 1, 2, \ldots, m)$ are irrelevant to $\gamma$.

The rest of this paper is organized as follows. Section 2 offers some notations and basic results that will be used in the subsequent discussions. In Sect. 3, we present the proximal ADMM with smaller proximal parameter for the model (1.1) and establish its global convergence and convergence rate. In Sect. 4, we report some numerical results to demonstrate the numerical advantage of smaller proximal parameter and relaxing technique. Finally, a brief conclusion ends this paper in Sect. 5.

## 2 Preliminaries

In this section, we define some notations and present the necessary assumptions on the model (1.1) under which our convergence analysis can be conducted.

Throughout this paper, we set

$$y = [x_2; \ldots; x_m], \quad x = [x_1; y], \quad u = [x; \lambda], \quad v = [x_2; x_3; \ldots; x_m; \lambda],$$

and

$$\bar{\theta}(y) = \sum_{i=2}^{m} \theta_i(x_i), \quad \theta(x) = \theta_1(x_1) + \bar{\theta}(y), \quad \mathcal{B} = [A_2, A_3, \ldots, A_m], \quad \mathcal{A} = [A_1, \mathcal{B}]$$
$$\bar{\mathcal{X}} = \mathcal{X}_2 \times \cdots \times \mathcal{X}_m, \quad \mathcal{X} = \mathcal{X}_1 \times \bar{\mathcal{X}}.$$

Furthermore, The domain of a function $f(\cdot) : \tilde{\mathcal{X}} \to (-\infty, +\infty]$ is defined as $\mathrm{dom}(f) = \{x \in \tilde{\mathcal{X}} | f(x) < +\infty\}$; The set of all relative interior points of a given nonempty convex set $\Omega$ is denoted by $\mathrm{ri}(\Omega)$; $G \succ 0$ (or $G \succeq 0$) denotes that the symmetric matrix $G$ is positive definite (or positive semi-definite); For any vector $x$ and a symmetric matrix $G$ with compatible dimensionality, we denote by $\|x\|_G^2 = x^\top G x$ though $G$ may be not positive definite; $\mathrm{Diag}\{A_1, A_2, \ldots, A_m\}$ has $A_i$ as its $i-$th block on the diagonal.

Throughout, we make the following standard assumptions.

**Assumption 2.1** The functions $\theta_i(\cdot)(i = 1, 2, \ldots, m)$ are all convex.

**Assumption 2.2** There is a point $(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_m) \in \mathrm{ri}(\mathrm{dom}\theta_1 \times \mathrm{dom}\theta_2 \times \cdots \times \mathrm{dom}\theta_m)$ such that $A_1\hat{x}_1 + A_2\hat{x}_2 + \cdots + A_m\hat{x}_m = b$, and $\hat{x}_i \in \mathcal{X}_i, i = 1, 2, \ldots, m$.

**Assumption 2.3** The matrices $A_i(i = 2, 3, \ldots, m)$ are full column rank.

Using the first-order optimality condition for convex minimization and Assumptions 2.1,2.2, the model (1.1) can be recast as the following mixed variational inequality problem, denoted by $\mathrm{VI}(\mathcal{U}, F, \theta)$: Finding a vector $u^* \in \mathcal{U}$ such that

$$\theta(x) - \theta(x^*) + (u - u^*)^\top F(u^*) \geq 0, \quad \forall u \in \mathcal{U}, \tag{2.1}$$

where $\mathcal{U} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_m \times \mathcal{R}^l$, and

$$F(u) := \begin{pmatrix} -A_1^\top \lambda \\ -A_2^\top \lambda \\ \vdots \\ -A_m^\top \lambda \\ \mathcal{A}x - b \end{pmatrix}. \tag{2.2}$$

We denote by $\mathcal{U}^*$ the solution set of (2.1), which is nonempty under Assumtion 2.2 and the fact that the solution set of the model (1.1) is assumed to be nonempty. The mapping $F(u)$ defined in (2.2) is affine and satisfies the following simple property

$$(u'-u)^\top (F(u')-F(u))=0, \quad \forall u', u \in \mathcal{U}. \tag{2.3}$$

The following proposition provides a criterion to measure the convergence rate of the iterative methods for the model (1.1).

**Proposition 2.1** [20] *Let $\hat{x} \in \mathcal{X}$. Then $\hat{x}$ is an optimal solution of the model (1.1) if and only if there exists $r > 0$ such that*

$$\theta(\hat{x}) - \theta(x^*) + \left(\hat{x} - x^*\right)^\top \left(-\mathcal{A}^\top \lambda^*\right) + \frac{r}{2}\|\mathcal{A}\hat{x} - b\|^2 = 0, \tag{2.4}$$

*where $(x^*, \lambda^*) \in \mathcal{U}^*$.*

## 3 The algorithm and its convergence result

In this section, we first present a new version of TPADMM in [28] for the model (1.1), which is named as the proximal ADMM with smaller proximal parameter, and then prove its global convergence and convergence rate under Assumptions 2.1–2.3. As stated earlier, the augmented Lagrangian function for the model (1.1) is defined as follows:

$$\mathcal{L}_\beta(x_1, \ldots, x_m; \lambda) := \sum_{i=1}^m \theta_i(x_i) - \left\langle \lambda, \sum_{i=1}^m A_i x_i - b \right\rangle + \frac{\beta}{2}\left\| \sum_{i=1}^m A_i x_i - b \right\|^2,$$

based on which, now we describe the proximal ADMM with smaller proximal parameter for solving (1.1) as follows.

**Algorithm 1.**

**Step 0.** Let parameters $\beta > 0$, $\gamma \in (0, \frac{1+\sqrt{5}}{2})$, $\tau > \frac{4+\max\{1-\gamma, \gamma^2-\gamma\}}{5}(m-1)$, the tolerance $\varepsilon > 0$, and the matrices $\bar{G}_i \in \mathcal{R}^{n_i \times n_i}$ with $\bar{G}_i \succeq 0 (i = 2, 3, \ldots, m)$. Choose the initial point $u^0 = [x_1^0; x_2^0; \ldots; x_m^0; \lambda^0] \in \mathcal{U}$; Set $k = 0$, and denote $G_i = \bar{G}_i - (1-\tau)\beta A_i^\top A_i (i = 2, 3, \ldots, m)$.

**Step 1.** Compute the new iterate $u^{k+1} = [x_1^{k+1}; x_2^{k+1}; \ldots, x_m^{k+1}; \lambda^{k+1}]$ via

$$\begin{cases} x_1^{k+1} := \underset{x_1 \in \mathcal{X}_1}{\operatorname{argmin}} \left\{ \mathcal{L}_\beta(x_1, x_2^k, \ldots, x_m^k; \lambda^k) \right\}, \\ x_i^{k+1} := \underset{x_i \in \mathcal{X}_i}{\operatorname{argmin}} \left\{ \mathcal{L}_\beta(x_1^{k+1}, x_2^k, \ldots, x_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k; \lambda^k) + \frac{1}{2}\|x_i - x_i^k\|_{G_i}^2 \right\}, \\ i = 2, 3, \ldots, m, \\ \lambda^{k+1} := \lambda^k - \gamma\beta \left( \sum_{i=1}^m A_i x_i^{k+1} - b \right). \end{cases} \tag{3.1}$$

**Step 2.** If $\|u^k - u^{k+1}\| \leq \varepsilon$, then stop; otherwise, set $k = k+1$, and go to Step 1.

*Remark 3.1* Since the matrices $\bar{G}_i (i = 2, 3, \ldots, m)$ can be any positive semi-definite matrices, we can set $\bar{G}_i = \tau_i I_{n_i} - \tau\beta A_i^\top A_i$ with $\tau_i \geq \tau\beta \|A_i\|^2$. Then, $\bar{G}_i \succeq 0$, $G_i = \tau_i I_{n_i} - \beta A_i^\top A_i (i = 2, 3, \ldots, m)$, and the subproblem with respect to the primal variable $x_i$ in (3.1) can be rewritten as

$$x_i^{k+1} = \mathrm{argmin}_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{\tau_i}{2} \|x_i - v_i^k\|^2 \right\},$$

where $v_i^k = \frac{1}{\tau_i}[\tau_i x_i^k - \beta A_i^\top A_1 x_1^{k+1} - \beta A_i^\top (\sum_{j \neq 2} A_j x_j^k - b) + A_i^\top \lambda^k]$ is a vector. Therefore, when $\mathcal{X}_i = \mathcal{R}^{n_i}$ and $\theta_i(x_i)$ takes some special functions, such as $\|x_i\|_1$, $\|x_i\|_2$, the above subproblem has a closed-form solution [3,17]. Furthermore, from $\tau_i \geq \tau\beta\|A_i\|^2$ and $\tau > \frac{4+\max\{1-\gamma, \gamma^2-\gamma\}}{5}(m-1)$, we have

$$\tau_i > \frac{4 + \max\{1 - \gamma, \gamma^2 - \gamma\}}{5}(m-1)\beta\|A_i\|^2.$$

Therefore, the feasible region of the proximal parameter $\tau_i$ is generally larger than those in [20,28].

*Remark 3.2* Different from $\gamma = 1$ in [13], the feasible set of $\gamma$ in Algorithm 1 is the interval $(0, \frac{1+\sqrt{5}}{2})$, and larger values of $\gamma$ can often speed up the convergence of Algorithm 1 (see Sect. 4).

When Assumptions 2.1–2.3 hold and the constant $\tau$ satisfy the following condition

$$\tau > \max\{m - 1 - c_0, 4c_0 + (m-1)\max\{1 - \gamma, \gamma^2 - \gamma\}\}, \tag{3.2}$$

where $c_0$ is a positive constant that will be specified later, Algorithm 1 is globally convergent and has the worst-case $\mathcal{O}(1/t)$ convergence rate in an ergodic sense.

**Theorem 3.1** *Let $\{u^k\}$ be the sequence generated by Algorithm 1 and $\tau$ satisfies (3.2), $\gamma \in (0, \frac{1+\sqrt{5}}{2})$. Then, the whole sequence $\{u^k\}$ converges globally to some $\hat{u}$, which belongs to $\mathcal{U}^*$.*

**Theorem 3.2** *Let $\{u^k\}$ be the sequence generated by Algorithm 1, and set*

$$\hat{x}_t = \frac{1}{t}\sum_{k=1}^{t} \hat{x}^k,$$

*where $t$ is a positive integer. Then, $\hat{x}_t \in \mathcal{X}$, and*

$$\theta(\hat{x}_t) - \theta(x^*) + (\hat{x}_t - x^*)^\top(-\mathcal{A}^\top\lambda^*) + \frac{\beta}{2}\min\left\{1, \frac{1+\gamma-\gamma^2}{\gamma}\right\}\|\mathcal{A}\hat{x}_t - b\|^2 \leq \frac{D}{t}, \tag{3.3}$$

*where* $(x^*, \lambda^*) \in \mathcal{U}^*$, *and D is a constant defined by*

$$D = \frac{1}{2}\left(\|v^1 - v^*\|_H^2 + \beta \max\left\{1-\gamma, 1-\frac{1}{\gamma}\right\}\|\mathcal{A}x^1 - b\|^2 + \|y^0 - y^1\|_N^2 \right.$$
$$\left. + c_0\beta\sum_{i=2}^{m}\|A_i\left(x_i^0 - x_i^1\right)\|^2\right). \tag{3.4}$$

To prove Theorems 3.1 and 3.2, we need the inequality (3.21) below, which is established by the following seven lemmas step by step. Based on the first-order optimality conditions of the subproblems in (3.1), we present Lemmas 3.1 and 3.2, which provide a lower bound of some vector to be a solution of VI($\mathcal{U}$, $F$, $\theta$), and then Lemma 3.3 further rewrites the lower bound as some quadratic terms. Lemma 3.4 provides a lower bound of some quadratic terms appeared in Lemma 3.3, and Lemma 3.5 gives a lower bound of the parameter $\tau$. Lemma 3.6 further deals with the lower bound of some quadratic term and crossing term. Based on the conclusions established in the first six lemmas, the most important inequality (3.21) is proved in Lemma 3.7.

For convenience, we now define two matrices to make the following analysis more compact.

$$G_0 = \begin{pmatrix} G_2 & -\beta A_2^\top A_3 & \cdots & -\beta A_2^\top A_m \\ -\beta A_3^\top A_2 & G_3 & \cdots & -\beta A_3^\top A_m \\ \vdots & \vdots & \ddots & \vdots \\ -\beta A_m^\top A_2 & -\beta A_m^\top A_3 & \cdots & G_m \end{pmatrix}, \tag{3.5}$$

and

$$Q = \begin{pmatrix} \bar{G}_2 + \tau\beta A_2^\top A_2 & 0 & \cdots & 0 & 0 \\ 0 & \bar{G}_3 + \tau\beta A_3^\top A_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & \bar{G}_m + \tau\beta A_m^\top A_m & 0 \\ -A_2 & -A_3 & \cdots & -A_m & I_l/\beta \end{pmatrix}. \tag{3.6}$$

**Lemma 3.1** *Let* $\{u^k\}$ *be the sequence generated by Algorithm 1. Then, we have* $x_i^k \in \mathcal{X}_i$ *and*

$$\beta\left(\mathcal{A}x^{k+1} - b\right)^\top \mathcal{B}\left(y^k - y^{k+1}\right)$$
$$\geq (1-\gamma)\beta\left(\mathcal{A}x^k - b\right)^\top \mathcal{B}\left(y^k - y^{k+1}\right) + \left\|y^k - y^{k+1}\right\|_{G_0}^2 \tag{3.7}$$
$$- \left(y^k - y^{k+1}\right)^\top G_0\left(y^{k-1} - y^k\right).$$

*Proof* By the first-order optimality condition for $x_i$-subproblem in (3.1), we have $x_i^{k+1} \in \mathcal{X}_i (i = 1, 2, \ldots, m)$ and

$$\theta_i(x_i) - \theta_i\left(x_i^{k+1}\right) + \left(x_i - x_i^{k+1}\right)^\top \\ \left\{-A_i^\top \lambda^k + \beta A_i^\top \left(A_1 x_1^{k+1} + \sum_{j=2, j\neq i}^{m} A_j x_j^k + A_i x_i^{k+1} - b\right)\right. \\ \left. + G_i\left(x_i^{k+1} - x_i^k\right)\right\} \geq 0, \quad \forall x_i \in \mathcal{X}_i, i = 2, 3, \ldots, m,$$

i.e.,

$$\theta_i(x_i) - \theta_i\left(x_i^{k+1}\right) + \left(x_i - x_i^{k+1}\right)^\top \left\{-A_i^\top \lambda^k + \beta A_i^\top \left(\sum_{j=1}^{m} A_j x_j^{k+1} - b\right)\right. \\ \left. -\beta A_i^\top \sum_{j=2, j\neq i}^{m} A_j\left(x_j^{k+1} - x_j^k\right) + G_i\left(x_i^{k+1} - x_i^k\right)\right\} \geq 0, \quad \forall x_i \in \mathcal{X}_i, \\ i = 2, 3, \ldots, m. \tag{3.8}$$

Then, summing the above inequality over $i = 2, 3, \ldots, m$, and using the definition of $G_0$ in (3.5), we get

$$\bar{\theta}(y) - \bar{\theta}\left(y^{k+1}\right) + \left(y - y^{k+1}\right)^\top \\ \times \left[-\mathcal{B}^\top \lambda^k + \beta \mathcal{B}^\top \left(\mathcal{A} x^{k+1} - b\right) + G_0\left(y^{k+1} - y^k\right)\right] \geq 0.$$

Note that the above inequality is also true for $k := k - 1$ and thus

$$\bar{\theta}(y) - \bar{\theta}\left(y^k\right) + \left(y - y^k\right)^\top \left[-\mathcal{B}^\top \lambda^{k-1} + \beta \mathcal{B}^\top \left(\mathcal{A} x^k - b\right) + G_0\left(y^k - y^{k-1}\right)\right] \geq 0.$$

Setting $y = y^k$ and $y = y^{k+1}$ in the above inequalities, respectively, and then adding them, we get

$$\left(y^k - y^{k+1}\right)^\top \mathcal{B}^\top \left[\left(\lambda^{k-1} - \lambda^k\right) + \beta\left(\mathcal{A} x^{k+1} - b\right) - \beta\left(\mathcal{A} x^k - b\right)\right] \\ \geq \|y^k - y^{k+1}\|_{G_0}^2 - \left(y^k - y^{k+1}\right)^\top G_0\left(y^{k-1} - y^k\right).$$

By the updating formula for $\lambda$ in (3.1), we have $\lambda^{k-1} - \lambda^k = \gamma\beta(\mathcal{A} x^k - b)$. Substituting this into the left-hand side of the above inequality, we get the assertion (3.7) immediately. This completes the proof. □

Now, let us define an auxiliary sequence $\{\hat{u}^k\} = \{[\hat{x}_1^k; \hat{x}_2^k; \ldots; \hat{x}_m^k; \hat{\lambda}^k]\}$, whose components are defined by

$$\hat{x}_i^k = x_i^{k+1} (i = 1, 2, \ldots, m), \quad \hat{\lambda}^k = \lambda^k - \beta\left(A_1 x_1^{k+1} + \sum_{i=2}^{m} A_i x_i^k - b\right), \tag{3.9}$$

**Lemma 3.2** *The auxiliary sequence* $\{\hat{u}^k\}$ *satisfies* $\hat{u}^k \in \mathcal{U}$ *and*

$$\theta(x) - \theta\left(\hat{x}^k\right) + \left(u - \hat{u}^k\right)^\top F\left(\hat{u}^k\right) \geq \left(v - \hat{v}^k\right)^\top Q\left(v^k - \hat{v}^k\right), \quad \forall u \in \mathcal{U}, \tag{3.10}$$

*where the matrix* $Q$ *is defined by* (3.6).

*Proof* From $\hat{x}^k = x^{k+1}$, it is obvious that $\hat{u}^k \in \mathcal{U}$. The first-order optimality condition for $x_1$-subproblem in (3.1) and the definition of $\hat{\lambda}^k$ in (3.9) imply

$$\theta_1(x_1) - \theta_1\left(\hat{x}_1^k\right) + \left(x_1 - \hat{x}_1^k\right)^\top \left(-A_1^\top \hat{\lambda}^k\right) \geq 0, \quad \forall x_1 \in \mathcal{X}_1. \tag{3.11}$$

By the definitions of $\hat{\lambda}^k$ and $G_i$ $(i = 2, 3, \ldots, m)$, we can get

$$\begin{aligned} &-A_i^\top \left[\lambda^k - \beta\left(\sum_{i=1}^m A_i x_i^{k+1} - b\right)\right] - \beta A_i^\top \sum_{j=2, j\neq i}^m A_j\left(x_j^{k+1} - x_j^k\right) + G_i\left(x_i^{k+1} - x_i^k\right) \\ &= -A_i^\top \left[\lambda^k - \beta\left(A_1 x_1^{k+1} + \sum_{j=2}^m A_i x_i^k - b\right)\right] + \left(G_i + \beta A_i^\top A_i\right)\left(x_i^{k+1} - x_i^k\right) \\ &= -A_i^\top \hat{\lambda}^k + \left(G_i + \beta A_i^\top A_i\right)\left(\hat{x}_i^k - x_i^k\right) \\ &= -A_i^\top \hat{\lambda}^k + \left(\bar{G}_i + \tau\beta A_i^\top A_i\right)\left(\hat{x}_i^k - x_i^k\right). \end{aligned}$$

Then, substituting the above equality into the left-hand side of (3.8), we thus derive

$$\begin{aligned} &\theta_i(x_i) - \theta_i\left(\hat{x}_i^k\right) + \left(x_i - \hat{x}_i^k\right)^\top \left\{-A_i^\top \hat{\lambda}^k + \left(\bar{G}_i + \tau\beta A_i^\top A_i\right)\left(\hat{x}_i^k - x_i^k\right)\right\} \geq 0, \\ &\forall x_i \in \mathcal{X}_i, i = 2, 3, \ldots, m. \end{aligned} \tag{3.12}$$

Furthermore, from the definition of $\hat{\lambda}^k$, we get the following inequality

$$\left(\lambda - \hat{\lambda}^k\right)^\top \left\{\left(\sum_{i=1}^m A_i \hat{x}_i^k - b\right) - \sum_{j=2}^m A_j\left(\hat{x}_j^k - x_j^k\right) + \frac{1}{\beta}\left(\hat{\lambda}^k - \lambda^k\right)\right\} \geq 0, \quad \forall \lambda \in \mathcal{R}^l. \tag{3.13}$$

Then, adding (3.11), (3.12) over $i = 2, 3, \ldots, m$, (3.13), and by some simple manipulations, we get the assertion (3.10) immediately. This completes the proof. □

*Remark 3.3* If $u^k = u^{k+1}$, then we have

$$x^k = \hat{x}^k, \lambda^k = \hat{\lambda}^k.$$

Substituting the above two equalities into the right-hand side of (3.10), we obtain

$$\theta(x) - \theta(\hat{x}^k) + (u - \hat{u}^k)^\top F(\hat{u}^k) \geq 0, \quad \forall u \in \mathcal{U},$$

i.e.,

$$\theta(x) - \theta(x^k) + (u - u^k)^\top F(u^k) \geq 0, \quad \forall u \in \mathcal{U},$$

which implies that $u^k$ is a solution of VI($\mathcal{U}$, $F, \theta$), and thus $x^k$ is a solution of the model (1.1). This indicates that the stopping criterion of Algorithm 1 is reasonable.

The updating formula for $\lambda^{k+1}$ in (3.1) can be rewritten as

$$\begin{aligned}
&\lambda^{k+1} \\
&= \lambda^k - \left( -\gamma\beta \sum_{i=2}^m A_i \left( x_i^k - x_i^{k+1} \right) \right) - \gamma\beta \left( A_1 x_1^{k+1} + \sum_{i=2}^m A_i x_i^k - b \right) \\
&= \lambda^k - \left[ -\gamma\beta \sum_{i=2}^m A_i \left( x_i^k - \hat{x}_i^k \right) + \gamma(\lambda^k - \hat{\lambda}^k) \right].
\end{aligned}$$

Combining the above equality and (3.9), we get

$$v^{k+1} = v^k - M(v^k - \hat{v}^k), \tag{3.14}$$

where

$$M = \begin{pmatrix} I_{n_2} & 0 & \cdots & 0 & 0 \\ 0 & I_{n_3} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & I_{n_m} & 0 \\ -\gamma\beta A_2 & -\gamma\beta A_3 & \cdots & -\gamma\beta A_m & \gamma I_l \end{pmatrix}.$$

Now we define two auxiliary matrices as

$$H := QM^{-1} \quad \text{and} \quad G := Q^\top + Q - M^\top H M.$$

By simple calculation, we have

$$H = \begin{pmatrix} \bar{G}_2 + \tau\beta A_2^\top A_2 & 0 & \cdots & 0 & 0 \\ 0 & \bar{G}_3 + \tau\beta A_3^\top A_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & \bar{G}_m + \tau\beta A_m^\top A_m & 0 \\ 0 & 0 & \cdots & 0 & I_l/(\gamma\beta) \end{pmatrix},$$

and

$$G = \begin{pmatrix} \tilde{G}_2 & -\gamma\beta A_2^\top A_3 & \cdots & -\gamma\beta A_2^\top A_m & (\gamma-1)A_2^\top \\ -\gamma\beta A_3^\top A_2 & \tilde{G}_3 & \cdots & -\gamma\beta A_3^\top A_m & (\gamma-1)A_3^\top \\ \vdots & \vdots & \ddots & \vdots & \\ -\gamma\beta A_m^\top A_2 & -\gamma\beta A_m^\top A_3 & \cdots & \tilde{G}_m & (\gamma-1)A_m^\top \\ (\gamma-1)A_2 & (\gamma-1)A_3 & \cdots & (\gamma-1)A_m & (2-\gamma)I_l/\beta \end{pmatrix},$$

where $\tilde{G}_i = \bar{G}_i + (\tau-\gamma)\beta A_i^\top A_i (i = 2, 3, \ldots, m)$. Obviously, under Assumption 3.3, the matrix $H$ is positive definite, and the following relationship with respect to the matrices $Q, M, H$ holds

$$Q = HM. \tag{3.15}$$

Based on (3.10) and (3.14), it seems that Algorithm 1 can be categorized into the prototype algorithm proposed in [12]. However, the convergence results of Algorithm 1 cannot be guaranteed by the results in [12] because we cannot ensure that the matrix $G$ is positive semi-definite, which is a necessary condition for the global convergence of the prototype algorithm in [12]. To show the matrix $G$ maybe indefinite, we only need to show that its sub-matrix $G(1:m-1, 1:m-1)$ maybe indefinite. Here $G(1:m-1, 1:m-1)$ denotes the corresponding sub-matrix formed from the rows and columns with the indices $(1:m-1)$ and $(1:m-1)$ in the block sense. In fact, setting $\bar{G}_i = 0\,(i = 2, 3, \ldots, m)$, one has

$$G(1:m-1, 1:m-1) = \begin{pmatrix} (\tau-\gamma)\beta A_2^\top A_2 & -\gamma\beta A_2^\top A_3 & \cdots & -\gamma\beta A_2^\top A_m \\ -\gamma\beta A_3^\top A_2 & (\tau-\gamma)\beta A_3^\top A_3 & \cdots & -\gamma\beta A_3^\top A_m \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma\beta A_m^\top A_2 & -\gamma\beta A_m^\top A_3 & \cdots & (\tau-\gamma)\beta A_m^\top A_m \end{pmatrix}$$

$$= \beta \begin{pmatrix} A_2^\top & 0 & \cdots & 0 \\ 0 & A_3^\top & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m^\top \end{pmatrix} \begin{pmatrix} (\tau-\gamma)I_l & -\gamma I_l & \cdots & -\gamma I_l \\ -\gamma I_l & (\tau-\gamma)I_l & \cdots & -\gamma I_l \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma I_l & -\gamma I_l & \cdots & (\tau-\gamma)I_l \end{pmatrix} \begin{pmatrix} A_2 & 0 & \cdots & 0 \\ 0 & A_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m \end{pmatrix}.$$

The middle matrix in the above expression can be further written as

$$\begin{pmatrix} \tau-\gamma & -\gamma & \cdots & -\gamma \\ -\gamma & \tau-\gamma & \cdots & -\gamma \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma & -\gamma & \ldots & \tau-\gamma \end{pmatrix} \otimes I_l$$

Then, we only need to show the $(m-1)$-by-$(m-1)$ matrix

$$\begin{pmatrix} \tau-\gamma & -\gamma & \cdots & -\gamma \\ -\gamma & \tau-\gamma & \cdots & -\gamma \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma & -\gamma & \cdots & \tau-\gamma \end{pmatrix}$$

may be indefinite. In fact, its eigenvalues are

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{m-2} = \tau,\ \lambda_{m-1} = \tau - (m-1)\gamma.$$

Then, the matrix $G(1:m-1, 1:m-1)$ is positive semi-definite iff $\tau \geq (m-1)\gamma$. Obviously, we cannot ensure $G(1:m-1, 1:m-1)$ is positive semi-definite for any $\tau \geq \frac{4+\max\{1-\gamma,\gamma^2-\gamma\}}{5}(m-1)$. Therefore, the matrix $G$ may be indefinite.

Now, let us rewrite the crossing term $(v-\hat{v}^k)^\top Q(v^k-\hat{v}^k)$ on the right-hand side (3.10) as some quadratic terms.

**Lemma 3.3** *Let $\{u^k\}$ be the sequence generated by Algorithm 1. Then, for any $u \in \mathcal{U}$, we have*

$$(v-\hat{v}^k)^\top Q(v^k-\hat{v}^k) = \frac{1}{2}(\|v-v^{k+1}\|_H^2 - \|v-v^k\|_H^2) + \frac{1}{2}(\|v^k-\hat{v}^k\|_H^2 - \|v^{k+1}-\hat{v}^k\|_H^2). \tag{3.16}$$

*Proof* By $H \succ 0$, (3.14) and (3.15), we have

$$\begin{aligned} &(v-\hat{v}^k)^\top Q(v^k-\hat{v}^k) \\ &= (v-\hat{v}^k)^\top HM(v^k-\hat{v}^k) \\ &= (v-\hat{v}^k)^\top H(v^k-v^{k+1}) \\ &= \frac{1}{2}(\|v-v^{k+1}\|_H^2 - \|v-v^k\|_H^2) + \frac{1}{2}(\|v^k-\hat{v}^k\|_H^2 - \|v^{k+1}-\hat{v}^k\|_H^2), \end{aligned}$$

where the last equality is obtained by setting $a=v, b=\hat{v}^k, c=v^k, d=v^{k+1}$ in the identity

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

This completes the proof. □

Substituting (3.16) into the left-hand side of (3.10), for any $u \in \mathcal{U}$, we get

$$\begin{aligned} \theta(x) - \theta(\hat{x}^k) + (u-\hat{u}^k)^\top F(\hat{u}^k) &\geq \frac{1}{2}(\|v-v^{k+1}\|_H^2 - \|v-v^k\|_H^2) \\ &+ \frac{1}{2}(\|v^k-\hat{v}^k\|_H^2 - \|v^{k+1}-\hat{v}^k\|_H^2). \end{aligned} \tag{3.17}$$

The first group-term $\frac{1}{2}(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2)$ on the right-hand side of (3.17) can be manipulated consecutively between iterates. Therefore, we only to deal with the second group-term $\frac{1}{2}(\|v^k-\hat{v}^k\|_H^2-\|v^{k+1}-\hat{v}^k\|_H^2)$.

**Lemma 3.4** *Let $\{u^k\}$ be the sequence generated by Algorithm 1. Then, we have*

$$\begin{aligned}
&\|v^k-\hat{v}^k\|_H^2-\|v^{k+1}-\hat{v}^k\|_H^2\\
&\geq \sum_{i=2}^m \|x_i^k-x_i^{k+1}\|_{\bar{G}_i}^2+\beta(\tau-(m-1)\max\{1-\gamma,\gamma^2-\gamma\})\\
&\quad \sum_{i=2}^m \|A_i(x_i^k-x_i^{k+1})\|^2+\beta\min\left\{1,\frac{1+\gamma-\gamma^2}{\gamma}\right\}\|\mathcal{A}x^{k+1}-b\|^2 \qquad (3.18)\\
&\quad +\beta\max\left\{1-\gamma,1-\frac{1}{\gamma}\right\}(\|\mathcal{A}x^{k+1}-b\|^2-\|\mathcal{A}x^k-b\|^2)\\
&\quad +2\|y^k-y^{k+1}\|_{G_0}^2-2(y^k-y^{k+1})^\top G_0(y^{k-1}-y^k).
\end{aligned}$$

*Proof* Obviously, $\hat{\lambda}^k$ defined by (3.9) can be rewritten as

$$\hat{\lambda}^k=\lambda^k-\beta\left(A_1x_1^{k+1}+\mathcal{B}y^k-b\right).$$

Then, from (3.14) and the definition of $H$, we can expand $\|v^{k+1}-\hat{v}^k\|_H^2$ as

$$\begin{aligned}
&\|v^{k+1}-\hat{v}^k\|_H^2\\
&=\|(I-M)(v^k-\hat{v}^k)\|_H^2\\
&=\frac{1}{\gamma\beta}\|(\lambda^k-\hat{\lambda}^k)-[-\gamma\beta\mathcal{B}(y^k-\hat{y}^k)+\gamma(\lambda^k-\hat{\lambda}^k)]\|^2\\
&=\frac{1}{\gamma\beta}\|(\lambda^k-\hat{\lambda}^k)-[-\gamma\beta\mathcal{B}(y^k-y^{k+1})+\gamma\beta(A_1x_1^{k+1}+\mathcal{B}y^k-b)]\|^2\\
&=\frac{1}{\gamma\beta}\|(\lambda^k-\hat{\lambda}^k)-\gamma\beta(\mathcal{A}x^{k+1}-b)\|^2\\
&=\frac{1}{\gamma\beta}\|\lambda^k-\hat{\lambda}^k\|^2-2(\lambda^k-\hat{\lambda}^k)^\top(\mathcal{A}x^{k+1}-b)+\gamma\beta\|\mathcal{A}x^{k+1}-b\|^2,
\end{aligned}$$

and thus, we have

$$\begin{aligned}
&\|v^k-\hat{v}^k\|_H^2-\|v^{k+1}-\hat{v}^k\|_H^2\\
&=\sum_{i=2}^m\|x_i^k-x_i^{k+1}\|_{\bar{G}_i}^2+\tau\beta\sum_{i=2}^m\|A_i(x_i^k-x_i^{k+1})\|^2+\frac{1}{\gamma\beta}\|\lambda^k-\hat{\lambda}^k\|^2\\
&\quad-\left(\frac{1}{\gamma\beta}\|\lambda^k-\hat{\lambda}^k\|^2-2(\lambda^k-\hat{\lambda}^k)^\top(\mathcal{A}x^{k+1}-b)+\gamma\beta\|\mathcal{A}x^{k+1}-b\|^2\right)
\end{aligned}$$

$$= \sum_{i=2}^{m} \|x_i^k - x_i^{k+1}\|_{\bar{G}_i}^2 + \tau\beta \sum_{i=2}^{m} \|A_i(x_i^k - x_i^{k+1})\|^2$$
$$+2(\lambda^k - \hat{\lambda}^k)^\top(\mathcal{A}x^{k+1} - b) - \gamma\beta\|\mathcal{A}x^{k+1} - b\|^2.$$

From (3.9) again, we have $\lambda^k - \hat{\lambda}^k = \beta\mathcal{B}(y^k - y^{k+1}) + \beta(\mathcal{A}x^{k+1} - b)$, and then $2(\lambda^k - \hat{\lambda}^k)^\top(\mathcal{A}x^{k+1} - b) = 2\beta(\mathcal{A}x^{k+1} - b)^\top\mathcal{B}(y^k - y^{k+1}) + 2\beta\|\mathcal{A}x^{k+1} - b\|^2$. Substituting this into the right-hand side of the above equality, we obtain

$$\|v^k - \hat{v}^k\|_H^2 - \|v^{k+1} - \hat{v}^k\|_H^2$$
$$= \sum_{i=2}^{m} \|x_i^k - x_i^{k+1}\|_{\bar{G}_i}^2 + \tau\beta \sum_{i=2}^{m} \|A_i(x_i^k - x_i^{k+1})\|^2$$
$$+2\beta(\mathcal{A}x^{k+1} - b)^\top\mathcal{B}(y^k - y^{k+1}) + (2-\gamma)\beta\|\mathcal{A}x^{k+1} - b\|^2$$
$$\geq \sum_{i=2}^{m} \|x_i^k - x_i^{k+1}\|_{\bar{G}_i}^2 + \tau\beta \sum_{i=2}^{m} \|A_i(x_i^k - x_i^{k+1})\|^2 + \beta(2-\gamma)\|\mathcal{A}x^{k+1} - b\|^2$$
$$+2\beta(1-\gamma)(\mathcal{A}x^k - b)^\top\mathcal{B}(y^k - y^{k+1}) + 2\|y^k - y^{k+1}\|_{G_0}^2 - 2(y^k - y^{k+1})^\top$$
$$G_0(y^{k-1} - y^k),$$

where the last inequality comes from (3.7). By applying the Cauchy–Schwartz inequality, we can get

$$\begin{cases} 2\beta(1-\gamma)(\mathcal{A}x^k - b)^\top\mathcal{B}(y^k - y^{k+1}) \geq -\beta(1-\gamma)(\|\mathcal{A}x^k - b\|^2 \\ \quad +\|\mathcal{B}(y^k - y^{k+1})\|^2), \quad \text{if } \gamma \in (0, 1], \\ 2\beta(1-\gamma)(\mathcal{A}x^k - b)^\top\mathcal{B}(y^k - y^{k+1}) \geq -\beta(\gamma - 1)(\frac{1}{\gamma}\|\mathcal{A}x^k - b\|^2 \\ \quad +\gamma\|\mathcal{B}(y^k - y^{k+1})\|^2), \quad \text{if } \gamma \in (1, +\infty). \end{cases}$$

Substituting this into the right-hand side of the above inequality, we obtain

$$\|v^k - \hat{v}^k\|_H^2 - \|v^{k+1} - \hat{v}^k\|_H^2$$
$$\geq \sum_{i=2}^{m} \|x_i^k - x_i^{k+1}\|_{\bar{G}_i}^2 + \tau\beta \sum_{i=2}^{m} \|A_i(x_i^k - x_i^{k+1})\|^2$$
$$+\beta \min\left\{1, \frac{1+\gamma-\gamma^2}{\gamma}\right\}\|\mathcal{A}x^{k+1} - b\|^2$$
$$+\beta \max\{1-\gamma, 1 - \frac{1}{\gamma}\}(\|\mathcal{A}x^{k+1} - b\|^2 - \|\mathcal{A}x^k - b\|^2)$$
$$-\beta \max\{1-\gamma, \gamma^2 - \gamma\}\|\mathcal{B}(y^k - y^{k+1})\|^2$$
$$+2\|y^k - y^{k+1}\|_{G_0}^2 - 2(y^k - y^{k+1})^\top G_0(y^{k-1} - y^k),$$

which together with $\|\mathcal{B}(y^k - y^{k+1})\|^2 \leq (m-1)\sum_{i=2}^{m}\|A_i(x_i^k - x_i^{k+1})\|^2$ implies that (3.18) holds. This completes the proof. □

Now, let us deal with the last two terms $2\|y^k - y^{k+1}\|_{G_0}^2 - 2(y^k - y^{k+1})^\top G_0(y^{k-1} - y^k)$ on the right-hand side of (3.18). For this purpose, we define

$$N = G_0 + c_0\beta \mathrm{Diag}\left\{A_2^\top A_2, A_3^\top A_3, \ldots, A_m^\top A_m\right\}, \tag{3.19}$$

where $c_0$ is a positive constant that will be specified later.

**Lemma 3.5** *The matrix $N$ is positive definite if $\tau > m - 1 - c_0$.*

*Proof* Since $G_i = \bar{G}_i - (1-\tau)\beta A_i^\top A_i (i = 2, 3, \ldots, m)$, then for any $v = [x_2; x_3; \ldots; x_m] \neq 0$, we have

$$\begin{aligned} & v^\top N v \\ & = \sum_{i=2}^m \|x_i\|_{\bar{G}_i}^2 + \beta(\tau - 1 + c_0)\sum_{i=2}^m \|A_i x_i\|^2 - \beta\sum_{i\neq j}(A_i x_i)^\top(A_j x_j) \\ & \geq \beta(\tau + c_0 - m + 1)\sum_{i=2}^m \|A_i x_i\|^2 > 0, \end{aligned}$$

where the first inequality follows from $\bar{G}_i \succeq 0 (i = 2, 3, \ldots, m)$ and the second inequality comes from $\tau > m - 1 - c_0$ and Assumption 3.3. Therefore, the matrix $N$ is positive definite. The proof is completed. □

**Lemma 3.6** *Let $\{u^k\}$ be the sequence generated by Algorithm 1. Then,*

$$\begin{aligned} & 2\|y^k - y^{k+1}\|_{G_0}^2 - 2(y^k - y^{k+1})^\top G_0(y^{k-1} - y^k) \\ & \quad \geq \|y^k - y^{k+1}\|_N^2 - \|y^{k-1} - y^k\|_N^2 - 3c_0\beta\sum_{i=2}^m \|A_i(x_i^k - x_i^{k+1})\|^2 \\ & \quad - c_0\beta\sum_{i=2}^m \|A_i(x_i^{k-1} - x_i^k)\|^2. \end{aligned} \tag{3.20}$$

*Proof* From (3.19), we have $G_0 = N - c_0\beta\mathrm{Diag}\{A_2^\top A_2, A_3^\top A_3, \ldots, A_m^\top A_m\}$. Therefore,

$$\begin{aligned} & 2\|y^k - y^{k+1}\|_{G_0}^2 - 2(y^k - y^{k+1})^\top G_0(y^{k-1} - y^k) \\ & = 2(y^k - y^{k+1})^\top G_0((y^k - y^{k+1}) - (y^{k-1} - y^k)) \\ & = 2(y^k - y^{k+1})^\top [N - c_0\beta\mathrm{Diag}\{A_2^\top A_2, A_3^\top A_3, \ldots, A_m^\top A_m\}]((y^k - y^{k+1}) \\ & \quad -(y^{k-1} - y^k)) \\ & = 2\|y^k - y^{k+1}\|_N^2 - 2(y^k - y^{k+1})^\top N(y^{k-1} - y^k) \\ & \quad -2c_0\beta(y^k - y^{k+1})^\top \mathrm{Diag}\{A_2^\top A_2, A_3^\top A_3, \ldots, A_m^\top A_m\}(y^k - y^{k+1}) \\ & \quad +2c_0\beta(y^k - y^{k+1})^\top \mathrm{Diag}\{A_2^\top A_2, A_3^\top A_3, \ldots, A_m^\top A_m\}(y^{k-1} - y^k) \end{aligned}$$

$$\geq \|y^k - y^{k+1}\|_N^2 - \|y^{k-1} - y^k\|_N^2 - 3c_0\beta \sum_{i=2}^{m} \|A_i(x_i^k - x_i^{k+1})\|^2$$
$$-c_0\beta \sum_{i=2}^{m} \|A_i(x_i^{k-1} - x_i^k)\|^2,$$

where the last inequality follows from the Cauchy–Schwartz inequality. The proof is completed. □

Based on the above three lemmas, we can deduce the recurrence relation of the sequence generated by Algorithm 1 as follows.

**Lemma 3.7** *Let $\{u^k\}$ be the sequence generated by Algorithm 1. Then,*

$$\begin{aligned}
&\|v^{k+1} - v^*\|_H^2 + \beta \max\left\{1-\gamma, 1-\frac{1}{\gamma}\right\} \|\mathcal{A}x^{k+1} - b\|^2 + \|y^k - y^{k+1}\|_N^2 \\
&\quad + c_0\beta \sum_{i=2}^{m} \|A_i(x_i^k - x_i^{k+1})\|^2 \\
&\leq \|v^k - v^*\|_H^2 + \beta \max\left\{1-\gamma, 1-\frac{1}{\gamma}\right\} \|\mathcal{A}x^k - b\|^2 + \|y^{k-1} - y^k\|_N^2 \\
&\quad + c_0\beta \sum_{i=2}^{m} \|A_i(x_i^{k-1} - x_i^k)\|^2 \\
&\quad - \sum_{i=2}^{m} \|x_i^k - x_i^{k+1}\|_{\bar{G}_i}^2 - \beta[\tau - 4c_0 - (m-1)\max\{1-\gamma, \gamma^2-\gamma\}] \\
&\quad \sum_{i=2}^{m} \|A_i(x_i^k - x_i^{k+1})\|^2 - \beta \min\left\{1, \frac{1+\gamma-\gamma^2}{\gamma}\right\} \|\mathcal{A}x^{k+1} - b\|^2.
\end{aligned} \tag{3.21}$$

*Proof* Substituting (3.20) into (3.18), one has

$$\begin{aligned}
&\|v^k - \hat{v}^k\|_H^2 - \|v^{k+1} - \hat{v}^k\|_H^2 \\
&\quad \geq \sum_{i=2}^{m} \|x_i^k - x_i^{k+1}\|_{\bar{G}_i}^2 + \beta[\tau - 4c_0 - (m-1)\max\{1-\gamma, \gamma^2-\gamma\}] \\
&\qquad \sum_{i=2}^{m} \|A_i(x_i^k - x_i^{k+1})\|^2 + \beta \min\left\{1, \frac{1+\gamma-\gamma^2}{\gamma}\right\} \|\mathcal{A}x^{k+1} - b\|^2 \\
&\quad +\beta \max\left\{1-\gamma, 1-\frac{1}{\gamma}\right\} (\|\mathcal{A}x^{k+1} - b\|^2 - \|\mathcal{A}x^k - b\|^2)
\end{aligned}$$

$$+(\|y^k - y^{k+1}\|_N^2 - \|y^{k-1} - y^k\|_N^2) + c_0\beta\left(\sum_{i=2}^m \|A_i(x_i^k - x_i^{k+1})\|^2 - \sum_{i=2}^m \|A_i(x_i^{k-1} - x_i^k)\|^2\right).$$

This together with (3.17) implies

$$\begin{aligned}
&2\theta(x) - 2\theta(\hat{x}^k) + 2(u - \hat{u}^k)^\top F(\hat{u}^k) \\
&\geq (\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \sum_{i=2}^m \|x_i^k - x_i^{k+1}\|_{\bar{G}_i}^2 \\
&\quad + \beta[\tau - 4c_0 - (m-1)\max\{1-\gamma, \gamma^2 - \gamma\}]\sum_{i=2}^m \|A_i(x_i^k - x_i^{k+1})\|^2 \\
&\quad + \beta\min\left\{1, \frac{1+\gamma-\gamma^2}{\gamma}\right\}\|\mathcal{A}x^{k+1} - b\|^2 + \beta\max\left\{1-\gamma, 1-\frac{1}{\gamma}\right\} \\
&\quad (\|\mathcal{A}x^{k+1} - b\|^2 - \|\mathcal{A}x^k - b\|^2) + (\|y^k - y^{k+1}\|_N^2 - \|y^{k-1} - y^k\|_N^2) \\
&\quad + c_0\beta\left(\sum_{i=2}^m \|A_i(x_i^k - x_i^{k+1})\|^2 - \sum_{i=2}^m \|A_i(x_i^{k-1} - x_i^k)\|^2\right).
\end{aligned} \tag{3.22}$$

Setting $u = u^* \in \mathcal{U}^*$ in the above inequality and using (2.3), we get (3.21). This completes the proof. □

With the above conclusions in hand, we can prove the Theorem 3.1 and Theorem 3.2 as follows.

*Proof of Theorem 3.1.* By (3.21), (3.2) and $\gamma \in \left(0, \frac{1+\sqrt{5}}{2}\right)$, one has

$$\begin{aligned}
&\sum_{k=1}^\infty \left(\sum_{i=2}^m \|x_i^k - x_i^{k+1}\|_{\bar{G}_i}^2 + c_1 \sum_{i=2}^m \|A_i(x_i^k - x_i^{k+1})\|^2 + c_2\|\mathcal{A}x^{k+1} - b\|^2\right) \\
&\quad \leq \|v^1 - v^*\|_H^2 + c_3\|\mathcal{A}x^1 - b\|^2 + \|y^0 - y^1\|_N^2 + c_0\beta\sum_{i=2}^m \|A_i(x_i^0 - x_i^1)\|^2 \\
&\quad < +\infty,
\end{aligned} \tag{3.23}$$

where $c_1 = \beta[\tau - 4c_0 - (m-1)\max\{1-\gamma, \gamma^2-\gamma\}]$, $c_2 = \beta\min\{1, \frac{1+\gamma-\gamma^2}{\gamma}\}$, $c_3 = \beta\max\{1-\gamma, 1-\frac{1}{\gamma}\}$. This and Assumption 2.3, $\bar{G}_i \succeq 0 (i = 2, 3, \ldots, m)$ indicate that

$$\lim_{k\to\infty} \|y^k - y^{k+1}\| = \lim_{k\to\infty} \|\mathcal{A}x^{k+1} - b\| = 0. \tag{3.24}$$

Furthermore, by the iterative scheme (3.1) and (3.24), one has

$$\begin{aligned}&\|A_1(x_1^k - x_1^{k+1})\|\\&= \|\frac{1}{\gamma\beta}(\lambda^{k-1} - \lambda^k) - \frac{1}{\gamma\beta}(\lambda^k - \lambda^{k+1}) - \sum_{i=2}^m A_i(x_i^k - x_i^{k+1})\|\\&\le \frac{1}{\gamma\beta}\|\lambda^{k-1} - \lambda^k\| + \frac{1}{\gamma\beta}\|\lambda^k - \lambda^{k+1}\| + \sum_{i=2}^m \|A_i(x_i^k - x_i^{k+1})\| \to 0(\text{as } k \to \infty),\end{aligned}$$

which together with Assumption 2.3 implies that

$$\lim_{k\to\infty} \|x_1^k - x_1^{k+1}\| = 0. \tag{3.25}$$

On the other hand, by $H \succ 0$ and (3.21) again, we have the sequences $\{\|v^{k+1} - v^*\|\}$, $\{\|\mathcal{A}x^{k+1} - b\|\}$ are both bounded. Then, by

$$\begin{aligned}&\|A_1(x_1^{k+1} - x_1^*)\|\\&= \|\mathcal{A}x^{k+1} - b - \sum_{i=2}^m A_i(x_i^{k+1} - x_i^*)\|\\&\le \|\mathcal{A}x^{k+1} - b\| + \sum_{i=2}^m A_i\|(x_i^{k+1} - x_i^*)\|,\end{aligned}$$

and Assumption 2.3, we have that the sequence $\{\|x_1^{k+1} - x_1^*\|\}$ is also bounded. In conclusion, the sequence $\{u^k\}$ generated by Algorithm 1 is bounded. Then, it has at least a cluster point, saying $u^\infty$, and suppose the subsequence $\{u^{k_i}\}$ converges to $u^\infty$. Then, taking limits on both sides of (3.10) along the subsequence $\{u^{k_i}\}$ and using (3.24), (3.25), one has

$$\theta(x) - \theta(x^\infty) + (u - u^\infty)^\top F(u^\infty) \ge 0, \quad \forall u \in \mathcal{U}.$$

Therefore, $u^\infty \in \mathcal{U}^*$.

Hence, replacing $u^*$ by $u^\infty$ in (3.21), we get

$$\begin{aligned}&\|v^{k+1} - v^\infty\|_H^2 + c_3\|\mathcal{A}x^{k+1} - b\|^2 + \|y^k - y^{k+1}\|_N^2 + c_0\beta\sum_{i=2}^m \|A_i(x_i^k - x_i^{k+1})\|^2\\&\le \|v^k - v^\infty\|_H^2 + c_3\|\mathcal{A}x^k - b\|^2 + \|y^{k-1} - y^k\|_N^2 + c_0\beta\sum_{i=2}^m \|A_i(x_i^{k-1} - x_i^k)\|^2.\end{aligned}$$

From (3.24), (3.25), we have that for any given $\varepsilon > 0$, there exists $l_0 > 0$, such that

$$c_3\|\mathcal{A}x^k - b\|^2 + \|y^{k-1} - y^k\|_N^2 + c_0\beta\sum_{i=2}^m \|A_i(x_i^{k-1} - x_i^k)\|^2 < \frac{\varepsilon}{2}, \quad \forall k \ge l_0.$$

Since $v^{k_i} \to v^\infty$ for $i \to \infty$, there exists $k_l > l_0$, such that

$$\|v^{k_l} - v^\infty\|_H^2 < \frac{\varepsilon}{2}.$$

Then, the above three inequalities lead to that, for any $k > k_l$, we have

$$\begin{aligned}
&\|u^k - u^\infty\|_H^2 \\
&\le \|u^{k_l} - u^\infty\|_H^2 + c_3\|\mathcal{A}x^{k_l} - b\|^2 + \|y^{k_l-1} - y^{k_l}\|_N^2 + c_0\beta\sum_{i=2}^{m}\|A_i(x_i^{k_l-1} - x_i^{k_l})\|^2 \\
&< \varepsilon.
\end{aligned}$$

Therefore the whole sequence $\{u^k\}$ converges to the solution $u^\infty$. The proof is completed. □

Obviously, to get a large feasible region of $\tau$, from the condition (3.2), we can set $m-1-c_0 = 4c_0+(m-1)\max\{1-\gamma, \gamma^2-\gamma\}$, and solve this linear equation, we get $c_0 = \frac{m-1-(m-1)\max\{1-\gamma,\gamma^2-\gamma\}}{5}$, which is obvious greater than zero if $\gamma \in (0, \frac{1+\sqrt{5}}{2})$. Thus, the constant $\tau$ only need to satisfy the following conditon

$$\tau > \frac{4+\max\{1-\gamma, \gamma^2-\gamma\}}{5}(m-1).$$

*Remark 3.4* In the recent work [14], He and Yuan proposed an ADMM-like splitting method for the model (1.1) with $m = 3$, which belongs to the mixed Gauss–Seidel and Jacobian ADMMs. More specifically, the iterative scheme of the method proposed in [14] is given by

$$\begin{cases}
x_1^{k+1} := \underset{x_1\in\mathcal{X}_1}{\operatorname{argmin}}\left\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k; \lambda^k)\right\}, \\
\begin{cases}
x_2^{k+1} := \underset{x_2\in\mathcal{X}_2}{\operatorname{argmin}}\left\{\mathcal{L}_\beta(x_1^{k+1}, x_2, x_3^k; \lambda^k) + \frac{\tau}{2}\|A_2(x_2 - x_2^k)\|^2\right\}, \\
x_3^{k+1} := \underset{x_3\in\mathcal{X}_3}{\operatorname{argmin}}\left\{\mathcal{L}_\beta(x_1^{k+1}, x_2^k, x_3; \lambda^k) + \frac{\tau}{2}\|A_3(x_3 - x_3^k)\|^2\right\},
\end{cases} \\
\lambda^{k+1} := \lambda^k - \beta\left(\sum_{i=1}^3 A_i x_i^{k+1} - b\right),
\end{cases} \tag{3.26}$$

where $\tau \ge 0.6$. Though Algorithm 1 and (3.26) both belong to the mixed Gauss–Seidel and Jacobian ADMMs, Algorithm 1 has an important advantage over (3.26). That is the matrices $\bar{G}_i (i = 2, 3, \ldots, m)$ can be any positive semi-definite matrices, and when they are taken some special cases, the subproblems in Algorithm 1 often admit closed-form solutions; see Remark 3.1. However, when $A_i^\top A_i \ne I_{n_i} (i = 2, 3)$, the subproblems with respect to $x_i (i = 2, 3)$ in (3.26) don't have closed-form solutions even if $\theta_i(x_i) = \|x_i\|_1$ or $\theta_i(x_i) = \|x_i\|_2 (i = 2, 3)$.

*Remark 3.5* Algorithm 1 contains the methods in [13,14] as special cases. More precisely:

1. For Algorithm 1, setting $m = 2$, $\gamma = 1$, and $\bar{G}_2 = \tau_2 I_{n_2} - \tau\beta A_2^\top A_2$ with $\tau_2 \geq \tau\beta\|A_2\|^2$, $\tau > 0.8$. Then, $\bar{G}_2 \succeq 0$, and $G_2 = \tau_2 I_{n_2} - \beta A_2^\top A_2$. It is easy to check that Algorithm 1 reduces to the method in [13].
2. For Algorithm 1, setting $m = 3$, $\gamma = 1$, and $\bar{G}_i = 0 (i = 2, 3)$, then $G_i = (\tau - 1)\beta A_i^\top A_i (i = 2, 3)$ with $\tau > 1.6$. It is easy to check that Algorithm 1 reduces to the method (3.26) in [14].

At the end of this section, let us prove the worst-case $\mathcal{O}(1/t)$ convergence rate in an ergodic sense of Algorithm 1 according to the criterion established in Proposition 2.1.

*Proof of Theorem 3.2.* From $\hat{x}^k \in \mathcal{X}$ and the convexity of $\mathcal{X}$, we have $x_t \in \mathcal{X}$. Setting $x = x^*$, $\lambda = \lambda^*$ in (3.22), and summing the resulted inequality over $k = 1, 2, \ldots, t$, we have

$$\begin{aligned}
&\sum_{k=1}^{t}\left[\theta(\hat{x}^k) - \theta(x^*) + (\hat{u}^k - u^*)^\top F(\hat{u}^k) + \frac{\beta}{2}\min\left\{1, \frac{1+\gamma-\gamma^2}{\gamma}\right\}\|\mathcal{A}\hat{x}^k - b\|^2\right] \\
&\quad \leq \frac{1}{2}(\|v^1 - v^*\|_H^2 + \beta\max\left\{1-\gamma, 1-\frac{1}{\gamma}\right\}\|\mathcal{A}x^1 - b\|^2 + \|y^0 - y^1\|_N^2 \\
&\qquad + c_0\beta\sum_{i=2}^{m}\|A_i(x_i^0 - x_i^1)\|^2)
\end{aligned} \tag{3.27}$$

Furthermore, the crossing term $(\hat{u}^k - u^*)^\top F(\hat{u}^k)$ on the left-hand side of (3.27) can be rewritten as

$$\begin{aligned}
&(\hat{u}^k - u^*)^\top F(\hat{u}^k) \\
&\quad = (\hat{u}^k - u^*)^\top F(u^*) \\
&\quad = (\hat{x}^k - x^*)^\top(-\mathcal{A}^\top\hat{\lambda}^k) + (\hat{\lambda}^k - \lambda^*)^\top(\mathcal{A}\hat{x}^k - b) \\
&\quad = (b - \mathcal{A}\hat{x}^k)^\top\hat{\lambda}^k + (\hat{\lambda}^k - \lambda^*)^\top(\mathcal{A}\hat{x}^k - b) \\
&\quad = (-\lambda^*)^\top(\mathcal{A}\hat{x}^k - b) \\
&\quad = (\hat{x}^k - x^*)^\top(-\mathcal{A}^\top\lambda^*),
\end{aligned} \tag{3.28}$$

where the first equality comes from (2.3). Substituting (3.28) into the left-hand side of (3.27), we get

$$\begin{aligned}
&\sum_{k=1}^{t}\left[\theta(\hat{x}^k) - \theta(x^*) + (\hat{x}^k - x^*)^\top(-\mathcal{A}^\top\lambda^*) + \frac{\beta}{2}\min\left\{1, \frac{1+\gamma-\gamma^2}{\gamma}\right\}\|\mathcal{A}\hat{x}^k - b\|^2\right] \\
&\quad \leq \frac{1}{2}(\|v^1 - v^*\|_H^2 + \beta\max\left\{1-\gamma, 1-\frac{1}{\gamma}\right\}\|\mathcal{A}x^1 - b\|^2 + \|y^0 - y^1\|_N^2 \\
&\qquad + c_0\beta\sum_{i=2}^{m}\|A_i(x_i^0 - x_i^1)\|^2)
\end{aligned} \tag{3.29}$$

Then, dividing (3.29) by $t$ and using the convexity of $\theta(\cdot)$ and $\|\cdot\|^2$ lead to

$$\theta(\hat{x}_t) - \theta(x^*) + (\hat{x}_t - x^*)^\top(-\mathcal{A}^\top\lambda^*) + \frac{\beta}{2}\min\left\{1, \frac{1+\gamma-\gamma^2}{\gamma}\right\}\|\mathcal{A}\hat{x}_t - b\|^2 \le \frac{D}{t},$$

where the constant $D$ is defined by (3.4). Then, we get the result (3.3). This completes the proof. □

## 4 Numerical results

In this section, we use Algorithm 1 to solve three concrete models of (1.1), including the Lasso model ($m = 2$), the latent variable Gaussian graphical model selection on some synthetic data sets ($m = 3$), the linear homogeneous equations ($m = 4$). We mainly compare the performance of Algorithm 1 with the method in [24], denoted by SPADMM, and the method in [28], denoted by TPADMM. All codes were written in Matlab, and run in MatlabR2010a on a laptop with Pentium(R) Dual-Core CPU T4400@2.2GHz, 4GB of memory.

**Problem 1** The Lasso model

The Lasso model is

$$\min_{y\in\mathcal{R}^n} \mu\|y\|_1 + \frac{1}{2}\|Ay - b\|^2, \tag{4.1}$$

where $A \in \mathcal{R}^{m\times n}$, $b \in \mathcal{R}^m$, $\mu > 0$ is a regularization parameter. By introducing a new variable $x$, we can rewrite (4.1) as

$$\begin{aligned} &\min \frac{1}{2}\|x - b\|^2 + \mu\|y\|_1 \\ &\text{s.t. } x - Ay = 0, \\ &\qquad x \in \mathcal{R}^m, y \in \mathcal{R}^n, \end{aligned} \tag{4.2}$$

which is a special case of problem (1.1) with the following specifications:

$$\theta_1(x_1) = \frac{1}{2}\|x - b\|^2,\ \theta_2(x_2) = \mu\|y\|_1,\ A_1 = I_n,\ A_2 = -A,\ b = 0.$$

Then, Algorithm 1 with $\bar{G}_1 = 0$, $\bar{G}_2 = \tau_2 I_n - \beta\tau A^\top A$ can be used to solve (4.2), where $\tau_2 = \tau\beta\|A^\top A\|$, and the closed-form solutions of the subproblems resulted by Algorithm 1 are similar to those in [2,26], which are not listed here for succinctness. In this experiment, the matrix $A$ is generated by `A=randn(m,n); A = A*spdiags(1./sqrt(sum(A).^2))',0,n,n)`. The sparse vector $x^*$ is generated by `p=100/n; x*= sprandn(n,1,p)`, and the observed vector $b$ is generated via `b=A*x*+sqrt(0.001)*randn(m,1)`. We set the regularization parameter $\mu = 0.1\|\mathtt{A}^\top \mathtt{b}\|_\infty$. For Algorithm 1, we set $\beta = 1$, $\tau = (4 + \max\{1-\gamma, \gamma^2-\gamma\})/5$. For SPADMM, we set $\beta = 1$ and $\gamma$ in Table 1 is $r$ in SPADMM. For TPADMM, we set $\beta = 1$ and utilize the same technique to linear its second subproblem. The initial points are $y = 0$, $\lambda = 0$, and the stopping criterion is [2]:

$$\|x^{k+1} - Ay^{k+1}\| \le \epsilon^{\mathrm{pri}} \quad \text{and} \quad \|\beta A(y^{k+1} - y^k)\| \le \epsilon^{\mathrm{dual}},$$

where $\epsilon^{\mathrm{pri}} = \sqrt{n}\epsilon^{\mathrm{abs}} + \epsilon^{\mathrm{rel}} \max\{\|x^{k+1}\|, \|Ay^{k+1}\|\}$, and $\epsilon^{\mathrm{dual}} = \sqrt{n}\epsilon^{\mathrm{abs}} + \epsilon^{\mathrm{rel}}\|y^{k+1}\|$ with $\epsilon^{\mathrm{abs}} = 10^{-4}$ and $\epsilon^{\mathrm{rel}} = 10^{-3}$. The numerical results generated by the tested methods are reported in Table 1, in which we only report the number of iterations, because the tested methods has the same structure.

Numerical results in Table 1 indicate that the performance of Algorithm 1 is obviously better than the other two tested methods in the sense that the number of iterations taken by Algorithm 1 is at most 91% of those taken by SPADMM and at most 82% of those taken by TPADMM. We believe that the improvement should be contributed to the smaller proximal parameter of Algorithm 1. Then, the advantage of smaller proximal parameter is verified. Furthermore, though bigger relaxation factor the $\gamma$ can often speed up the corresponding iteration method, we observe that Algorithm 1 with $\gamma = 1.0$ almost always performs better than Algorithm 1 with $\gamma = 0.8, 1.2$. The reason maybe that $\tau$ gives more influence to the performance of Algorithm than $\gamma$ for this problem.

**Problem 2** The latent variable Gaussian graphical model selection

First, let us review latent variable Gaussian graphical model selection (LVGGMS) [4,21]. Let $X_{p\times 1}$ be the observed variables, and $Y_{r\times 1} (r \ll p)$ be the hidden variables such that $(X, Y)$ jointly follow a multivariate normal distribution in $\mathcal{R}^{p+r}$, where the covariance matrix $\Sigma_{(X,Y)} = [\Sigma_X, \Sigma_{XY}; \Sigma_{YX}, \Sigma_Y]$ and the precision matrix $\Theta_{(X,Y)} = [\Theta_X, \Theta_{XY}; \Theta_{YX}, \Theta_Y]$ are unknown. Under the prior assumption that the conditional precision matrix of observed variables $\Theta_X$ is sparse, the marginal precision matrix of observed variables, $\Sigma_X^{-1} = \Theta_X - \Theta_{XY}\Theta_Y^{-1}\Theta_{YX}$ is a difference between the sparse term $\Theta_X$ and the low-rank term $\Theta_{XY}\Theta_Y^{-1}\Theta_{YX}$. Therefore, the problem of interest is to recover the sparse conditional matrix $\Theta_X$ and the low-rank term $\Theta_{XY}\Theta_Y^{-1}\Theta_{YX}$ based on the observed variables $X$. Setting $\Sigma_X^{-1} = S - L$, Chandrasekaran et al. [4] introduced the following latent variable graphical model selection

$$\begin{aligned} &\min_{S,L} \langle S - L, \hat{\Sigma}_X\rangle - \operatorname{logdet}(S - L) + \alpha_1\|S\|_1 + \alpha_2 \mathbf{Tr}(L), \\ &\text{s.t. } S - L \succ 0, L \succeq 0, \end{aligned} \tag{4.3}$$

where $\hat{\Sigma}_X$ is the sample covariance matrix of $X$; $\|S\|_1$ is the $\ell_1$-norm of the matrix $S$ defined by $\sum_{i,j=1}^p |S_{ij}|$; and $\mathbf{Tr}(L)$ denotes the trace of the matrix $L$; $\alpha_1 > 0$ and $\alpha_2 > 0$ are given scalars controlling the sparsity and the low-rankness of the solution. Obviously, the model (4.3) can be rewritten as

$$\begin{aligned} &\min \langle R, \hat{\Sigma}_X\rangle - \operatorname{logdet} R + \alpha_1\|S\|_1 + \alpha_2 \mathbf{Tr}(L) \\ &\text{s.t. } R = S - L, R \succ 0, L \succeq 0, \end{aligned}$$

**Table 1** Comparison between the number of iterations taken by the tested methods for Lasso model

| $m$ | $n$ | SPADMM ($\gamma = 0.8$) | TPADMM ($\gamma = 0.8$) | Algorithm 1 ($\gamma = 0.8$) | SPADMM ($\gamma = 1.0$) | TPADMM ($\gamma = 1.0$) | Algorithm 1 ($\gamma = 1.0$) | SPADMM ($\gamma = 1.2$) | TPADMM ($\gamma = 1.2$) | Algorithm 1 ($\gamma = 1.2$) |
|---|---|---|---|---|---|---|---|---|---|---|
| 900 | 3000 | 51 | 57 | 45 | 51 | 55 | 44 | 51 | 56 | 46 |
| 1050 | 3500 | 53 | 58 | 46 | 53 | 57 | 45 | 53 | 56 | 47 |
| 1200 | 4000 | 45 | 51 | 40 | 45 | 50 | 39 | 45 | 49 | 41 |
| 1350 | 4500 | 48 | 53 | 42 | 46 | 52 | 41 | 48 | 51 | 43 |
| 1500 | 5000 | 41 | 46 | 36 | 41 | 45 | 35 | 41 | 45 | 37 |

which can be furthered casted as

$$\begin{aligned} &\min \langle R, \hat{\Sigma}_X \rangle - \mathrm{logdet} R + \alpha_1 \|S\|_1 + \alpha_2 \mathbf{Tr}(L) + \mathcal{I}(L \succeq 0) \\ &\text{s.t. } R - S + L = 0, \end{aligned} \tag{4.4}$$

where the indicator function $\mathcal{I}(L \succeq 0)$ is defined as

$$\mathcal{I}(L \succeq 0) := \begin{cases} 0, & \text{if} \quad L \succeq 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

The constraint $R \succ 0$ is removed since it is already imposed by the logdet$R$ function. The convex minimization (4.4) is a special case of the model (1.1) where $x_1 = R, x_2 = S, x_3 = L, \theta_1(x_1) = \langle R, \hat{\Sigma}_X \rangle - \mathrm{logdet} R, \theta_2(x_2) = \alpha_1 \|S\|_1, \theta_3(x_3) = \alpha_2 \mathbf{Tr}(L) + \mathcal{I}(L \succeq 0), A_1 = I_p, A_2 = -I_p, A_3 = I_p, b = 0$. Then, Algorithm 1 with $\bar{G}_i = 0 (i = 2, 3)$ can be used to solve (4.4), and its iterative scheme is listed as follows.

$$\begin{cases} R^{k+1} := \underset{R}{\mathrm{argmin}} \left\{ \mathcal{L}_\beta(R, S^k, L^k; \Lambda^k) \right\}, \\ S^{k+1} := \underset{S}{\mathrm{argmin}} \left\{ \mathcal{L}_\beta(R^{k+1}, S, L^k; \Lambda^k) + \frac{\beta(\tau-1)}{2} \|S - S^k\|_F^2 \right\}, \\ L^{k+1} := \underset{L}{\mathrm{argmin}} \left\{ \mathcal{L}_\beta(R^{k+1}, S^k, L; \Lambda^k) + \frac{\beta(\tau-1)}{2} \|L - L^k\|_F^2 \right\}, \\ \Lambda^{k+1} := \Lambda^k - \gamma\beta \left( R^{k+1} - S^{k+1} + L^{k+1} \right), \end{cases} \tag{4.5}$$

where the augmented Lagrangian function $\mathcal{L}_\beta(R, S, L; \Lambda)$ is defined as

$$\begin{aligned} \mathcal{L}_\beta(R, S, L; \Lambda) = \langle R, \hat{\Sigma}_X \rangle - \mathrm{logdet} R + \alpha_1 \|S\|_1 + \alpha_2 \mathbf{Tr}(L) + \mathcal{I}(L \succeq 0) \\ - \langle \Lambda, R - S - L \rangle + \frac{\beta}{2} \|R - S + L\|_F^2. \end{aligned}$$

In [21], the authors have elaborated on the similar subproblems as those in (4.5). Therefore, based on the discussion of [21], we give the closed-form solutions of the subproblems in (4.5).

- For the given $R^k, S^k, L^k$ and $\Lambda^k$, the $R$ subproblem in (4.5) admits a closed-form solution as
$$R^{k+1} = U \hat{\wedge} U^\top, \tag{4.6}$$
where $U$ is obtained by the eigenvalue decomposition: $U \mathrm{Diag}(\sigma) U^\top = (\hat{\Sigma}_X - \Lambda^k)/(\beta\tau) - [(\tau - 1)R^k + S^k - L^k]/\tau$, and $\hat{\wedge} = \mathrm{Diag}(\hat{\sigma})$ is obtained by:
$$\hat{\sigma}_j = \frac{-\sigma_j + \sqrt{\sigma_j^2 + 4/(\tau\beta)}}{2}, \quad j = 1, 2, \ldots, p.$$
- For the given $R^{k+1}, S^k, L^k$ and $\Lambda^k$, the $S$ subproblem in (4.5) admits a closed-form solution as
$$S^{k+1} = \mathrm{Shrink}(Z^k, \alpha_1/(\tau\beta)), \tag{4.7}$$

where $Z^k = -\Lambda^k/(\beta\tau) + [(\tau - 1)S^k + R^{k+1} + L^k]/\tau$ and $\text{Shrink}(Z, \tau) = \text{sign}(Z_{ij}) \cdot \max\{0, |Z_{ij}| - \tau\}$.

- For the given $R^{k+1}$, $S^k$, $L^k$ and $\Lambda^k$, the $L$ subproblem in (4.5) admits a closed-form solution as

$$L^{k+1} = U\tilde{\wedge}U^\top, \tag{4.8}$$

where $U$ is obtained by: $U\text{Diag}(\sigma)U^\top$ is the eigenvalue decomposition of the matrix $\Lambda^k/(\beta\tau) + [(\tau - 1)L^k + S^k - R^{k+1}]/\tau$, and $\hat{\wedge} = \text{Diag}(\hat{\sigma})$ is obtained by:

$$\hat{\sigma}_j = \max\{\sigma_j - \alpha_2/(\tau\beta), 0\}, \; j = 1, 2, \ldots, p.$$

The synthetic data of our experiment is generated by the following procedures [21]. Given the number of the observed variables $p$ and the number of latent variables $r$, we created a sparse matrix $\mathcal{W} \in \mathcal{R}^{(p+r)\times(p+r)}$ with sparsity around 10%, in which the nonzero entries were set to $-1$ or 1 with equality probability. From $\mathcal{W}$, we can get the true precision matrix $K = (\mathcal{W} * \mathcal{W}^\top)^{-1}$ and then obtain two submatrices of $K$, $\hat{S} = K(1:p, 1:p) \in \mathcal{R}^{p\times p}$ and $\hat{L} = K(1:p, p+1:p+r)K(p+1:p+r, p+1:p+r)^{-1}K(p+1:p+r, 1:p) \in \mathcal{R}^{p\times p}$, which are the ground truth matrices of the sparse matrix $S$ and the low-rank matrix $L$, respectively. The sample covariance matrix of the observed variables is defined by $\Sigma_X = \frac{1}{N}\sum_{i=1}^N Y_i^\top Y_i^\top$, where $N = 5p$ and the i.i.d. vectors $Y_1, Y_2, \ldots, Y_N$ are drew from the gaussian distribution $\mathcal{N}(0, (\hat{S} - \hat{L})^{-1})$. Throughout this experiment, we use the following stopping criterion:

$$\max\left\{\frac{\|R^{k+1} - R^k\|_F}{1 + \|R^k\|_F}, \frac{\|S^{k+1} - S^k\|_F}{1 + \|S^k\|_F}, \frac{\|L^k - L^k\|_F}{1 + \|L^k\|_F}\right\} < 10^{-5}.$$

In the following, we are going to compare Algorithm 1 with TPADMM [28] by solving (4.4). For Algorithm 1, we set $\beta = 0.5$, $\tau = 2.02(4 + \max\{1 - \gamma, \gamma^2 - \gamma\})/5$. For TPADMM, we set $\beta = \upsilon = 0.5, M_2 = M_3 = \upsilon I_p$. The initial points are $R^0 = \texttt{eye(p, p)}$, $S^0 = R^0$, $L^0 = \texttt{zeros(p, p)}$, $\Lambda^0 = \texttt{zeros(p, p)}$. The numerical results generated by Algorithm 1 and TPADMM are reported in Tables 2, 3 and 4.

From Tables 2, 3 and 4, several conclusions can be drawn here: (i) Both methods successfully solved all the tested problems and can deal with medium scale LVGGMS; (ii) Numerical results in the three tables show that Algorithm 1 performs better than TPADMM because Algorithm 1 always takes less number of iterations and less CPU time. In fact, Algorithm 1 can achieve an improvement of at least 20% (16%) reduction in the number of iterations (CPU time) over TPADMM; (iii) When the parameters $\alpha_1$ and $\alpha_2$ decrease, the advantage of Algorithm 1 over TPADMM becomes more clearly; (iv) Different to Problem 1, numerical results in the three tables also show that, for a fixed $(\alpha_1, \alpha_2)$, the performance of Algorithm 1 and TPADMM become better as the parameter $\gamma$ increases.

**Problem 3** Linear homogeneous equations

Consider a system of linear homogeneous equations in four variables, which is a special case of (1.1) with a null objective function and takes the form of

$$A_1x_1 + A_2x_2 + A_3x_3 + A_4x_4 = 0,$$

**Table 2** Comparison between the number of iterations (time in seconds) taken by Algorithm 1 and TPADMM for LVGGMS model with $p = 200, r = 10$

| $\alpha_1$ | $\alpha_2$ | Algorithm 1 ($\gamma = 0.8$) | TPADMM ($\gamma = 0.8$) | Ratio (%) | Algorithm 1 ($\gamma = 1.0$) | TPADMM ($\gamma = 1.0$) | Ratio (%) | Algorithm 1 ($\gamma = 1.2$) | TPADMM ($\gamma = 1.2$) | Ratio (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.01 | 0.1 | 223 (39.2) | 375 (64.4) | 0.59 (0.61) | 174 (22.1) | 305 (45.0) | 0.57 (0.49) | 154 (24.5) | 257 (41.8) | 0.60 (0.59) |
| 0.025 | 0.15 | 207 (25.3) | 263 (36.2) | 0.79 (0.70) | 154 (20.5) | 209 (27.2) | 0.73 (0.75) | 135 (16.0) | 174 (24.0) | 0.78 (0.67) |
| 0.05 | 0.1 | 101 (12.5) | 174 (23.7) | 0.58 (0.53) | 79 (9.5) | 135 (19.0) | 0.58 (0.50) | 69 (8.0) | 109 (13.3) | 0.63 (0.60) |
| 0.075 | 0.15 | 104 (12.7) | 157 (20.4) | 0.66 (0.62) | 80 (9.5) | 118 (15.4) | 0.67 (0.61) | 70 (9.0) | 92 (12.7) | 0.76 (0.70) |
| 0.1 | 0.1 | 107 (17.3) | 150 (24.0) | 0.71 (0.73) | 80 (9.4) | 112 (14.5) | 0.71 (0.84) | 69 (11.5) | 86 (14.8) | 0.80 (0.78) |

**Table 3** Comparison between the number of iterations (time in seconds) taken by Algorithm 1 and TPADMM for LVGGMS model with $p = 500, r = 25$

| $\alpha_1$ | $\alpha_2$ | Algorithm 1 ($\gamma$=0.8) | TPADMM ($\gamma$=0.8) | Ratio (%) | Algorithm 1 ($\gamma$=1.0) | TPADMM ($\gamma$=1.0) | Ratio (%) | Algorithm 1 ($\gamma$=1.2) | TPADMM ($\gamma$=1.2) | Ratio (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.025 | 0.15 | 56 (89.8) | 211 (356.8) | 0.27 (0.25) | 46 (74.0) | 169 (278.9) | 0.27 (0.26) | 42 (67.5) | 142 (241.8) | 0.30 (0.28) |
| 0.05 | 0.1 | 45 (71.4) | 121 (203.8) | 0.37 (0.35) | 39 (59.9) | 96 (164.5) | 0.41 (0.36) | 36 (57.6) | 79 (134.8) | 0.46 (0.43) |
| 0.075 | 0.15 | 54 (91.5) | 102 (170.2) | 0.53 (0.54) | 42 (68.4) | 80 (145.0) | 0.53 (0.47) | 39 (69.6) | 65 (116.2) | 0.60 (0.60) |
| 0.1 | 0.1 | 53 (85.5) | 96 (167.8) | 0.55 (0.51) | 42 (69.9) | 74 (120.1) | 0.57 (0.58) | 38 (61.9) | 59 (100.0) | 0.64 (0.62) |

**Table 4** Comparison between the number of iterations (time in seconds) taken by Algorithm 1 and TPADMM for LVGGMS model with $p = 700$, $r = 35$

| $\alpha_1$ | $\alpha_2$ | Algorithm 1 ($\gamma = 0.8$) | TPADMM ($\gamma = 0.8$) | Ratio (%) | Algorithm 1 ($\gamma = 1.0$) | TPADMM ($\gamma = 1.0$) | Ratio (%) | Algorithm 1 ($\gamma = 1.2$) | TPADMM ($\gamma = 1.2$) | Ratio (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.025 | 0.15 | 45 (226.8) | 190 (1098.2) | 0.24 (0.21) | 43 (172.5) | 153 (694.3) | 0.28 (0.25) | 43 (183.8) | 128 (598.4) | 0.34 (0.31) |
| 0.05 | 0.1 | 40 (168.9) | 117 (507.0) | 0.34 (0.33) | 39 (159.5) | 92 (382.4) | 0.42 (0.42) | 39 (167.9) | 76 (346.6) | 0.51 (0.48) |
| 0.075 | 0.15 | 46 (210.7) | 105 (460.1) | 0.44 (0.46) | 37 (149.4) | 82 (339.7) | 0.42 (0.44) | 34 (144.3) | 67 (298.7) | 0.51 (0.48) |
| 0.1 | 0.1 | 45 (189.8) | 98 (421.1) | 0.46 (0.45) | 38 (155.1) | 76 (333.3) | 0.50 (0.47) | 35 (148.1) | 62 (272.2) | 0.56 (0.54) |

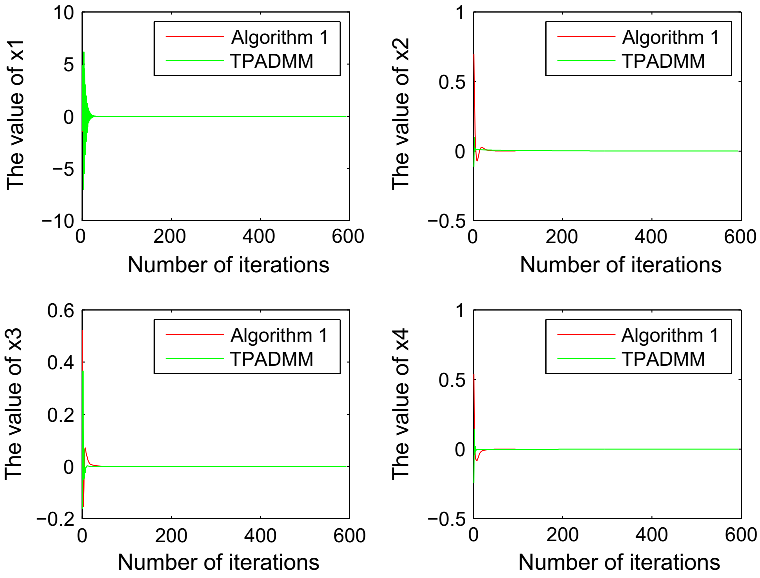


**Fig. 1** Evolution of the values of primal variables with respect to the number of iterations

where

$$A_1 = \begin{pmatrix} 1 \\ 3 \\ 8 \\ 2 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 2 \\ 5 \\ 7 \\ 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 5 \\ 3 \\ 3 \\ 9 \end{pmatrix}, \quad A_4 = \begin{pmatrix} 2 \\ 9 \\ 10 \\ 8 \end{pmatrix}.$$

We once again compare Algorithm 1 with TPADMM, where the various parameters for the two methods are set as follows: We take $\beta = 0.1, \gamma = 1.6, \tau = 2.98$, and $\bar{G}_i = 0 (i = 2, 3)$ for Algorithm 1, $\beta = 0.1, v = 5, M_2 = M_3 = vI_4$ for TPADMM. The initial points are $x_1^0 = x_2^0 = x_3^0 = x_4^0 = \lambda^0 = 1$. The stopping criterion is set as

$$\text{RelErr} := \log\left( \max \left\{ \max_{i=1,2,3,4} \{\|A_i(x_i^k - x_i^{k+1}\|_\infty\}, \|\lambda^k - \lambda^{k+1}\|_\infty\} \right\} \right) < 10^{-5}.$$

The maximum number of iteration is set 1000. In Figs. 1 and 2, the evolution of the primal variables and relative error are displayed, respecitvely.

Simulation result of the simple problem shows faster convergence rate can be obtained by the two methods for $k \le 58$. After $k = 58$, relative error of TPADMM declined at a slower rate, and the descent rate of relative error of Algorithm 1 is still quite stable. The numbers of iterations of Algorithm 1 and TPADMM are 94 and 593, respectively. In fact, we observe that the performance of TPADMM is quite sensitive to the parameter $\nu$, and for $\nu = 1, 2, 10, 11, 12$, TPADMM doesn't satisfy the above stopping criterion even for $k = 1000$.

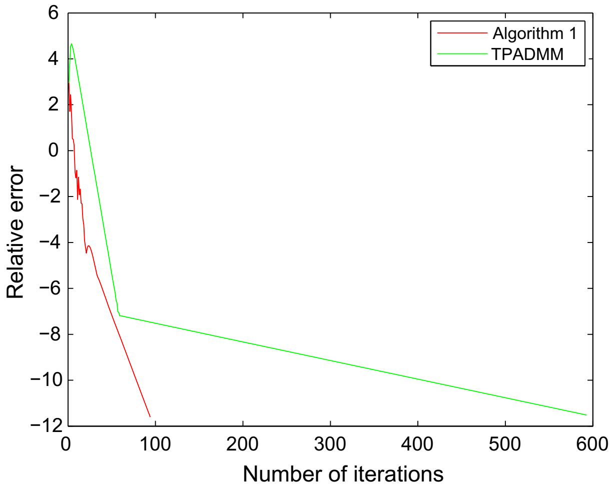


**Fig. 2** Evolution of the values of relative error with respect to the number of iterations

## 5 Conclusion

This paper provides a sharper lower bound of the proximal parameter in the proximal ADMM-type method for multi-block separable convex minimization, which can often alleviate the over-regularization effectiveness for the corresponding subproblem, and thus may speed up the convergence of corresponding method. The numerical results have verified the advantage of the smaller proximal parameter.

**Acknowledgements** The authors gratefully acknowledge the valuable comments of the anonymous reviewers.

## References

1. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, Berlin (2011)
2. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**, 1–122 (2011)
3. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim. **20**, 1956–1982 (2010)
4. Chandrasekaran, V., Parrilo, P., Willsky, A.: Latent variable graphical model selection via convex optimization. Ann. Stat. **40**(4), 1935–1967 (2012)
5. Chen, C.H., He, B.S., Ye, Y.Y., Yuan, X.M.: The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. Math. Program. **155**(1), 57–79 (2016)
6. Chen, P.J., Huang, J.G., Zhang, X.Q.: A primal-dual fixed point algorithm for minimization of the sum of three convex separable functions. Fixed Point Theory Appl. **2016**(1), 1–18 (2016)
7. Deng, W., Lai, M.J., Peng, Z.M., Yin, W.T.: Parallel multi-block ADMM with o$(1/k)$ convergence. J. Sci. Comput. **71**(2), 712–736 (2017)
8. Deng, W., Yin, W.T.: On the global and linear convergence of the generalized alternating direction method of multipliers. Rice University CAAM Technical Report TR12-14 (2012)
9. Fazel, M., Pong, T.K., Sun, D.F., Tseng, P.: Hankel matrix rank minimization with applications to system identification and realization. SIAM J. Matrix Anal. Appl. **34**, 946–977 (2013)

10. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. Comput. Math. Appl. **2**, 17–40 (1976)
11. Gao, B., Ma, F.: Symmetric alternating direction method with indefinite proximal regularization for linearly constrained convex optimization. Submitted to J. Optim. Theory Appl. (in revision) (2017)
12. He, B.S., Yuan, X.M.: On the direct extension of ADMM for multi-block separable convex programming and beyond: from variational inequality perspective, optimization-online (2014)
13. He, B.S., Ma, F., Yuan, X.M.: Linearized alternating direction method of multipliers via positive-indefinite proximal regularization for convex programming, optimization-online (2016)
14. He, B.S., Ma, F., Yuan, X.M., Positive-indefinite proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems, optimization-online (2016)
15. Hou, L.S., He, H.J., Yang, J.F.: A partially parallel splitting method for multiple-block separable convex programming with applications to robust PCA. Comput. Optim. Appl. **63**(1), 273–303 (2016)
16. Li, M., Sun, D.F., Toh, K.C.: A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization. SIAM J. Optim. **26**(2), 922–950 (2016)
17. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964–979 (1979)
18. Li, Q., Shen, L.X., Yang, L.H.: Split-Bregman iteration for framelet based image inpainting. Appl. Comput. Harmon. Anal. **32**, 145–154 (2012)
19. Li, Q., Xu, Y.S., Zhang, N.: Two-step fixed-point proximity algorithms for multi-block separable convex problems. Journal of Scientific Computing **70**(3), 1204–1228 (2017)
20. Lin, Z.C., Liu, R.S., Li, H.: Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning. Mach. Learn. **95**(2), 287–325 (2015)
21. Ma, S.Q., Xu, D.Z., Zou, H.: Alternating direction methods for latent variable Gaussian graphical model selection. Neural Comput. **25**, 2172–2198 (2013)
22. Ma, S.Q.: Alternating proximal gradient method for convex minimization. J. Sci. Comput. **68**(2), 546–572 (2016)
23. Sun, M., Liu, J.: A proximal Peaceman–Rachford splitting method for compressive sensing. J. Appl. Math. Comput. **50**(1–2), 349–363 (2016)
24. Sun, M., Liu, J.: Generalized Peaceman–Rachford splitting method for separable convex programming with applications to image processing. J. Appl. Math. Comput. **51**(1–2), 605–622 (2016)
25. Sun, M., Liu, J.: The convergence rate of the proximal alternating direction method of multipliers with indefinite proximal regularization. J. Inequal. Appl. **2017**, 19 (2017)
26. Sun, M., Wang, Y.J., Liu, J.: Generalized Peaceman–Rachford splitting method for multiple-block separable convex programming with applications to robust PCA. Caocolo **54**(1), 77–94 (2017)
27. Tao, M., Yuan, X.M.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM J. Optim. **21**, 57–81 (2011)
28. Wang, J.J., Song, W.: An algorithm twisted from generalized ADMM for multi-block separable convex minimization models. J. Comput. Appl. Math. **309**, 342–358 (2017)
29. Xu, M.H., Wu, T.: A class of linearized proximal alternating direction methods. J. Optim. Theory Appl. **151**(2), 321–337 (2011)
30. Zhang, X.Q., Burger, M., Osher, S.: A unified primal-dual algorithm framework based on Bregman iteration. J. Sci. Comput. **6**, 20–46 (2010)