J. Appl. Numer. Optim. 2 (2020), No. 2, pp. 155-169
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.2.2020.2.03

SEMI-PROXIMAL AUGMENTED LAGRANGIAN METHOD FOR SPARSE ESTIMATION OF HIGH-DIMENSIONAL INVERSE COVARIANCE MATRICES

CAN WU[1], YUNHAI XIAO[2,*], PEILI LI[3]

[1]*School of Mathematics and Statistics, Henan University, Kaifeng 475000, China*
[2]*Institute of Applied Mathematics, School of Mathematics and Statistics, Henan University, Kaifeng 475000, China*
[3]*School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China*

Abstract. Estimating a large and sparse inverse covariance matrix is a fundamental problem in modern multivariate analysis. Recently, a generalized model for a sparse estimation was proposed in which an explicit eigenvalue bounded constraint is involved. It covers a large number of existing estimation approaches as special cases. It was shown that the dual of the generalized model contains five separable blocks, which cause more challenges for minimizing. In this paper, we use an augmented Lagrangian method to solve the dual problem, but we minimize the augmented Lagrangian function with respect to each variable in a Jacobian manner, and add a proximal point term to make each subproblem easy to solve. We show that this iterative scheme is equivalent to adding a proximal point term to the augmented Lagrangian function, and its convergence can be directly followed. Finally, we give numerical simulations by using the synthetic data which show that the proposed algorithm is very effective in estimating high-dimensional sparse inverse covariance matrices.

Keywords. Nonsmooth convex minimization; Sparse inverse covariance matrix; Augmented Lagrangian method; Jacobian iteration; Proximal point term.

1. INTRODUCTION

The estimation of the large and sparse covariance matrix or its inverse is a classical problem in modern multivariate analysis. This problem has a wide range of applications in the fields of economics, finance, biology, social networks, speech recognition, gene network analysis and others. Let ξ_i be a n-dimensional random vector following a multivariate Gaussian distribution $N(\mu,\Sigma)$ with unknown mean μ and covariance matrix Σ. The goal of the covariance matrix estimation problem is to estimate covariance matrix Σ from given samples $\{\xi_i : i = 1,2,\cdots,\zeta\}$. Let $S := \frac{1}{\zeta-1}\sum_{i=1}^{\zeta}(\xi_i-\bar{\xi})(\xi_i-\bar{\xi})^\top$ be the sample covariance matrix, where $\bar{\xi} = \frac{1}{\zeta}\sum_{i=1}^{\zeta}\xi_i$ is the sample mean. The population covariance matrix Σ can be estimated better by sample covariance matrix S in a low dimensional case [1].

Let $\Theta = \Sigma^{-1}$ be an inverse covariance matrix, where every zero entry means conditional independence between a pair of variables. To estimate Σ^{-1}, it is natural to consider the maximum

*Corresponding author.
E-mail addresses: 2019010105@m.scnu.edu.cn (C. Wu), yhxiao@henu.edu.cn (Y. Xiao), lipeili@whu.edu.cn (P. Li).

Received March 27, 2020; Accepted June 2, 2020.

likelihood estimation (MLE) given by

$$\Theta = \arg\min_{X \in \mathscr{S}_+^n} \langle S, X \rangle - \log\det X, \tag{1.1}$$

where $\langle S, X \rangle = \mathrm{tr}(S^\top X)$ presents the standard trace inner product, $\log\det 0 = -\infty$ is assumed wherever it occurs, and $\mathscr{S}_+^n$ is the set of all $n \times n$ positive semi-definite matrices. Unfortunately, the MLE alone is not enough to our goal because S may not be positive definite, and then the objective function in (1.1) is unbounded below. Even if S is positive definite, the MLE may not have the desired sparsity structure determined by a prior given conditional independence.

To obtain robust estimators and promote the sparsity in the high-dimensional inverse covariance matrix, many estimators have been presented, analyzed, and implemented over the past few years. Particularly, when the conditional independence among parts of random variables is known, the following estimation model is appropriate [6]

$$\min_{X \in \mathscr{S}_+^n} \Big\{ \langle S, X \rangle - \log\det X : X_{ij} = 0 \quad \forall (ij) \in \Omega_0 \Big\}, \tag{1.2}$$

where Ω_0 is an index set of conditionally independent variables. However, model (1.2) is inapplicable when conditional independence between variables is unknown. Specifically, the following lasso penalized normal likelihood estimator is proposed, e.g., [2, 3, 18]

$$\min_{X \in \mathscr{S}_+^n} \Big\{ \langle S, X \rangle - \log\det X + \rho \|X\|_1 : \ \alpha I_n \preceq X \preceq \beta I_n \Big\}, \tag{1.3}$$

where $\|X\|_1 = \sum_{ij} |X_{ij}|$, $\rho > 0$ is a tuning parameter, and $0 \le \alpha < \beta$ are scalars. As shown in [2] that model (1.3) has the potential ability to discover the underlying distribution's structure when possible noises are contained on the sample covariance matrix S. In many applications, block-wise sparsity structure in the inverse covariance matrix is highly desirable. Because of this, many authors [10, 14, 17, 19] considered the following general estimation model based on the adaptive group lasso

$$\min_{X \in \mathscr{S}_+^n} \Big\{ \langle S, X \rangle - \log\det X + \sum_{l=1}^{r} \omega_l \|X_{\{g_l\}}\|_p : \mathscr{A}X = b \Big\}, \tag{1.4}$$

where r is the number of groups, $\omega_l > 0$ is an adaptive tuning parameter, $\|\cdot\|_p$ is the ℓ_p-norm with popular choices $p = 1, 2, \infty$; $X_{\{g_l\}}$ is a vector constructed by the components of X with indices set g_l; operator $\mathscr{A}$ is a generic linear mapping from S_+^n to $\mathbb{R}^m$ and $b \in \mathbb{R}^m$, and the equality $\mathscr{A}X = b$ contains the prior knowledge about some entries of solution in (1.2) as a special case.

By adding bound constraint on eigenvalue to model (1.4), Xiao, Li and Lu [16] considered the generalized inverse covariance matrix estimation model

$$\begin{array}{ll} \min\limits_{X \in \mathscr{S}_+^n} & \langle S, X \rangle - \log\det X + \sum_{l=1}^{r} \omega_l \|X_{\{g_l\}}\|_p \\ \text{s.t.} & \mathscr{A}X = b, \ \alpha I_n \preceq X \preceq \beta I_n. \end{array} \tag{1.5}$$

Clearly, problem (1.5) covers all the aforementioned models as special cases. Instead of directly solving (1.5), they [16] considered two types of the alternating direction method of multipliers (ADMM) to solve the corresponding five separable variables involved dual model. The first one is an extension of ADMM in which all the variables are updated in Gauss-Seidel manner using their latest values. While in the second one, all the dual variables are divided into two groups.

The symmetric Gauss-seidel (sGS) iteration [9] and the traditional Jacobian iteration are used to ensure that all the variables can be computed separately. From the latest sGS decomposition theorem, it was shown that their implemented algorithm is equivalent to the semi-proximal ADMM. In this paper, we focus on the augmented Lagrangian method to solve the dual problem of (1.5). More preciously, we treat all the dual variables as one group and adopt traditional Jacobian iteration, i.e., the newest value of the each variable is not used promptly until the whole loop is finished. We show that the algorithm implemented in this paper is equivalent to the one block semi-proximal augmented Lagrangian method, so that its convergence theorem is directly followed from the known convergence result [7]. We also present some numerical experiments on synthetic data to demonstrate attractive efficiency of the implemented algorithm for estimating large and sparse inverse covariance matrices.

In this paper, matrices and vectors are written in uppercase and lowercase letters, respectively. The symbols $\mathscr{S}^n$, $\mathscr{S}^n_+$ and $\mathscr{S}^n_{++}$ denote the sets of all $n\times n$ symmetric, symmetric positive semi-definite, and symmetric positive definite matrices, respectively. For the sake of convenience, $X\in\mathscr{S}^n_+$ (resp., $X\in\mathscr{S}^n_{++}$) is also represented by $X\succeq 0$ (resp., $X\succ 0$) occasionally. The transpose operation of a vector or matrix variable is denoted by superscript "$\top$", and the adjoint operators of $\mathscr{A}$ is represented by $\mathscr{A}^*$. The identity matrix of appropriate size is denoted by $\mathscr{I}$. The notation $\|\cdot\|_F$ represents the Frobenius norm for matrix. For matrices X,Y and vectors x, y of appropriate sizes, we define $\langle X,Y\rangle = \mathrm{tr}(X^\top Y)$ and $\langle x,y\rangle = x^\top y$, where "tr" represents the trace operation. $\log(\cdot)$ represents the natural logarithm function, and $\log 0 = -\infty$ is assumed wherever it might occur. For any closed convex set $\mathscr{C}$, the symbol $\Pi_{\mathscr{C}}(\cdot)$ represents the Euclidean projection onto $\mathscr{C}$. Moreover, $\mathrm{Diag}(d_1,d_2,\cdots,d_n)$ denotes a diagonal matrix with elements $d_1,d_2,\cdots,d_n$ in its diagonal position. Using this notation, $\mathrm{Diag}(A)$ presents a diagonal matrix possessing the diagonal components of a matrix A on its diagonal.

The remaining parts of this paper are organized as follows. In Section 2, we summarize some necessary background material. In Section 3, we present the semi-proximal augmented Lagrangian method for solving the dual model of (1.5) as well as its convergence theorem. In Section 4, we report numerical experiments to show the efficiency of the algorithm. Finally, we conclude this paper in Section 5.

2. Background Materials

In this section, we summarize some basic analytical results that will help to construct and clarify the following analysis. Most of the materials here can be readily found (often in more general form) in Rockafellar [12].

2.1. **Basic concepts.** Let $\mathscr{X}$ and $\mathscr{Y}$ be finite dimensional real Euclidean spaces endowed with an inner product $\langle\cdot,\cdot\rangle$ and its induced norm $\|\cdot\|$. A subset $\mathscr{C}$ of $\mathscr{X}$ is said to be convex if $(1-\lambda)x+\lambda y\in\mathscr{C}$ whenever $x\in\mathscr{C}$, $y\in\mathscr{C}$ and $0<\lambda<1$. The relative interior of $\mathscr{C}$, which we denote by $ri(\mathscr{C})$, is defined as the interior which results when $\mathscr{C}$ is regards as a subset of its affine hull. A subset $\mathscr{C}\subseteq\mathscr{X}$ is said to be a cone if $\lambda x\in\mathscr{C}$ for any x in $\mathscr{C}$ and $\lambda>0$. And a cone $\mathscr{C}$ is called a convex cone if it is convex. The polar cone of any cone $\mathscr{C}$ is defined by $\mathscr{C}^0 := \{y\in\mathscr{E}\mid\langle x,y\rangle\le 0,\ \forall x\in\mathscr{C}\}$. The dual cone is $\mathscr{C}^* := -\mathscr{C}^0$. Let $f:\mathscr{X}\to(-\infty,+\infty]$ be a closed proper convex function. The effective domain of f, which is denoted by $dom\ f$, is defined as $dom\ f=\{x:f(x)<+\infty\}$. A vector x^* is said to be a subgradient of f at point x if $f(z)\ge f(x)+\langle x^*,z-x\rangle$ for all $z\in\mathscr{X}$. The set of all subgradients of f at x is called the

subdifferential of f at x and is denoted by $\partial f(x)$. Obviously, $\partial f(x)$ is a closed convex set while it is not empty. It is known that ∂f is maximal monotone, i.e., for any $x,x' \in \mathscr{X}$ with $u \in \partial f(x)$, $u' \in \partial f(x')$, it holds that $\langle u-u', x-x' \rangle \geq 0$. Furthermore, it can be seen immediately from the definition that x^* is a global minimizer of f if and only if $0 \in \partial f(x^*)$. If f is differentiable at x and convex, then $\partial f(x)$ is the singleton set $\nabla f(x)$.

For any closed proper convex function $f : \mathscr{X} \to (-\infty, +\infty]$, the proximal point of x associated with f and is denoted by $P_f^\beta(x)$, i.e.,

$$P_f^\beta(x) := \operatorname{argmin}_y \Big\{ f(y) + \frac{1}{2\beta}\|y-x\|^2 \Big\},$$

where $\beta > 0$ is a given positive scalar. The Fenchel conjugate function of a convex f at x is defined as

$$f^\star(x) = \sup_y \{\langle x,y\rangle - f(y)\} = -\inf_y \{f(y) - \langle x,y\rangle\}.$$

It is well known that the conjugate function $f^\star$ is always convex and closed, proper if and only if f is proper [12]. Take the ℓ_p-norm $\|x\|_p$ as an example, its Fenchel conjugate function and proximal point mapping is expressed respectively as

$$\|x\|_p^\star = \begin{cases} 0, & \|x\|_q \leq 1, \\ +\infty, & \|x\|_q > 1, \end{cases}$$

and

$$P_f^\beta(x) = x - \Pi_{\mathscr{B}_q^1}(\beta x),$$

where $\frac{1}{p} + \frac{1}{q} = 1$, $p,q > 0$, and $\Pi_{\mathscr{B}_q^1}(\cdot)$ is the Euclidean projection onto $\mathscr{B}_q^1$ defined as $\mathscr{B}_q^1 := \{x : \|x\|_q \leq 1\}$.

To end this subsection, we list a couple of technical results which will be used in the subsequent development. The proofs of each conclusion can be found the the relevant papers and hence it is omitted here.

Proposition 2.1. [17] *Let $X \in \mathscr{S}^n$ and $X = QDiag(d_1,d_2,\cdots,d_n)Q^\top$ be its eigenvalue decomposition, where $d_1 \geq d_2 \geq \cdots \geq d_n$. For the scalar function $\phi_\beta(x) = (\sqrt{x^2+4\beta}+x)/2$ and $\beta > 0$, we define their matrix counterparts by*

$$\Phi_\beta(X) = QDiag\big(\phi_\beta(d_1), \phi_\beta(d_2), \cdots, \phi_\beta(d_n)\big)Q^\top. \tag{2.1}$$

Then the proximal point mapping of $f(X) = -\log\det X$ for any $X \in \mathscr{S}_{++}^n$ is given by $P_f^\beta(X) = \Phi_\beta(X)$.

Proposition 2.2. [5] *Let $X \in \mathscr{S}^n$ and $X = QDiag(d_1,d_2,\cdots,d_n)Q^\top$ be its eigenvalue decomposition, where $d_1 \geq d_2 \geq \cdots \geq d_n$. Define $\pi(d_i) = \max\{d_i, 0\}$, then the metric projection of X onto $\mathscr{S}_+^n$ is given by*

$$\Pi_{\mathscr{S}_+^n}(X) = QDiag\big(\pi(d_1), \pi(d_2), \cdots, \pi(d_n)\big)Q^\top. \tag{2.2}$$

2.2. **Review of the semi-proximal augmented Lagrangian method.** Although it can be reviewed in a separable form, the following problem formulation is sufficient for our statement of the augmented Lagrangian method

$$\begin{array}{ll} \min\limits_{y\in\mathscr{Y}} & f(y) \\ \text{s.t.} & \mathscr{G}^*y = c, \end{array} \tag{2.3}$$

where $f:\mathscr{Y}\to(-\infty,+\infty]$ is a closed proper convex function, $\mathscr{G}:\mathscr{X}\to\mathscr{Y}$ is a given linear map with adjoint $\mathscr{G}^*$, and $c\in\mathscr{X}$ is given. The Lagrangian function of problem (2.3) is defined by

$$\mathscr{L}(y;x) := f(y) - \langle x, \mathscr{G}^*y - c\rangle,\ \forall (x,y)\in\mathscr{X}\times\mathscr{Y}.$$

The dual of problem (2.3) takes the following form

$$\max_x \Big\{ \langle c,x\rangle - f^\star(\mathscr{G}x)\Big\}. \tag{2.4}$$

A vector $(\bar{x},\bar{y})\in\mathscr{X}\times\mathscr{Y}$ is said to be a saddle point to the Lagrangian function if it satisfies the following Karush-Kuhn-Tucker (KKT) system

$$0\in\partial f(y) - \mathscr{G}x, \quad \text{and} \quad 0 = \mathscr{G}^*y - c. \tag{2.5}$$

Let $\sigma>0$ be a given penalty parameter. The augmented Lagrangian function associated with (2.3) is given as follows:

$$\mathscr{L}_\sigma(y;x) := f(y) - \langle x, \mathscr{G}^*y - c\rangle + \frac{\sigma}{2}\|\mathscr{G}^*y - c\|^2.$$

Starting from an initial point $(x^0,y^0)\in\mathscr{X}\times(\mathrm{dom} f)$, the ordinary augmented Lagrangian method originally suggested by Hestenes [8] and Powell [11] takes the following from, for $k=0,1,\cdots$

$$\left\{\begin{array}{l} y^{k+1} := \mathrm{argmin}_y\ \mathscr{L}_\sigma(y;x^k), \\ x^{k+1} := x^k - \sigma(\mathscr{G}^*y - c). \end{array}\right. \tag{2.6}$$

The semi-proximal augmented Lagrangian method [7], expressed by semi-proximal alternating direction method of multipliers, is similar to (2.6) but with the following iterative form

$$\left\{\begin{array}{l} y^{k+1} := \mathrm{argmin}_y\ \mathscr{L}_\sigma(y;x^k) + \frac{1}{2}\|y^k - y\|^2_{\mathscr{T}}, \\ x^{k+1} := x^k - \tau\sigma(\mathscr{G}^*y - c), \end{array}\right. \tag{2.7}$$

where the steplength $\tau\in(0,2)$ and $\mathscr{T}$ is sef-adjoint positive semi-definite operator. If $\mathscr{T}\succ 0$ and $\tau\equiv 1$, then (2.7) is reduced to the proximal augmented Lagrangian method of Rockafellar [13] which owns the attractive feature that the objective function in (2.7) is strongly convex.

Next, we describe the convergence theory of the semi-proximal augmented Lagrangian method simply from [7]. Since ∂f is maximal monotone, there exists a self-adjoint and positive semi-definite operator Σ_f such that, for all $y,\hat{y}\in\mathrm{dom}(f)$, $u\in\partial f(y)$ and $\hat{u}\in\partial f(\hat{y})$,

$$\langle u-\hat{u}, y-\hat{y}\rangle \geq \|y-\hat{y}\|^2_{\Sigma_f}.$$

We assume that there exists $y_0\in\mathrm{ri}(\mathrm{dom} f)\cap P$, where P is the constraint set in (2.3) and $\Sigma_f + \mathscr{T} + \sigma GG^*$ is positive definite. Let $\{(y^k,x^k)\}$ be generated from the semi-proximal augmented Lagrangian method. Then it is from [7] that, for $\tau\in(0,2)$, the sequence $\{y^k\}$ converges to an optimal solution to (2.3), and $\{x^k\}$ converges to an optimal solution to the dual problem of (2.3).

3. Semi-proximal augmented Lagrangian method

3.1. **Model reformulation.** For the sake of convenience of the algorithm's development, we make some equivalent variants for model (1.5). For each group l, we define a linear mapping $\mathscr{P}_l$ such that $\mathscr{P}_l X = X_{\{g_l\}}$, and then set an auxiliary variable $y_l \in \mathbb{R}^{|g_l|}$ to take out the entries in $X_{\{g_l\}}$. Denote $y := (y_1; y_2; \ldots; y_r) \in \mathbb{R}^{|g_1|} \times \mathbb{R}^{|g_2|} \times \cdots \times \mathbb{R}^{|g_r|}$ and let $s := \sum_{l=1}^r |g_l|$ be the total number of elements in y. For convenience, we also set $\mathscr{P} := [\mathscr{P}_1; \mathscr{P}_2; \ldots; \mathscr{P}_r] : \mathbb{R}^{n\times n} \to \mathbb{R}^s$ and $\varphi(y) := \sum_{l=1}^r \omega_l \|y_l\|_p$. With these notations, the model (1.5) can be rewritten equivalently as

$$\begin{array}{ll} \min\limits_{X,y} & \langle S, X\rangle - \log\det X + \varphi(y) \\ \text{s.t.} & \mathscr{A}X = b,\ \mathscr{P}X - y = 0,\ \alpha I_n \preceq X \preceq \beta I_n. \end{array} \tag{3.1}$$

Using the standard Lagrangian duality, the dual of (3.1) takes the following form [10],

$$\begin{array}{ll} \max\limits_{u,v,Z,\Lambda_1,\Lambda_2} & n + \log\det Z - \varphi^\star(-v) + \langle u, b\rangle + \alpha\langle \Lambda_1, I\rangle - \beta\langle \Lambda_2, I\rangle \\ \text{s.t.} & \mathscr{A}^* u + \mathscr{P}^* v + \Lambda_1 - \Lambda_2 + Z = S, \\ & u \in \mathbb{R}^m,\ v \in \mathbb{R}^s,\ Z \in \mathscr{S}^n_{++},\ \Lambda_1 \in \mathscr{S}^n_+,\ \Lambda_2 \in \mathscr{S}^n_+, \end{array} \tag{3.2}$$

The Lagrangian function associated to the dual problem (3.2) is defined as

$$\begin{aligned} \mathscr{L}(u, v, Z, \Lambda_1, \Lambda_2; X) := {} & \log\det Z - \varphi^\star(-v) + \langle u, b\rangle + \alpha\langle \Lambda_1, I\rangle \\ & - \beta\langle \Lambda_2, I\rangle - \langle X, \mathscr{A}^* u + \mathscr{P}^* v + \Lambda_1 - \Lambda_2 + Z - S\rangle. \end{aligned}$$

If $(\bar{u}, \bar{v}, \bar{Z}, \bar{\Lambda}_1, \bar{\Lambda}_2)$ is the optimal solution of problem (3.2), then there exists a Lagrangian multiplier $\bar{X}$ such that the following KKT conditions are satisfied:

$$\left\{ \begin{array}{l} \partial\varphi^\star(-\bar{v}) - \mathscr{P}\bar{X} \ni 0, \\ b - \mathscr{A}\bar{X} = 0,\ \bar{X}\bar{Z} = \mathscr{I}_n,\ \bar{Z} \in \mathscr{S}^n_{++}, \\ \alpha\mathscr{I}_n \preceq \bar{X} \preceq \beta\mathscr{I}_n,\ \bar{\Lambda}_1 \in \mathscr{S}^n_+,\ \bar{\Lambda}_2 \in \mathscr{S}^n_+, \\ \mathscr{A}^* u + \mathscr{P}^* v + \Lambda_1 - \Lambda_2 + Z - S = 0, \\ \langle \bar{\Lambda}_1,\ \bar{X} - \alpha\mathscr{I}_n\rangle = 0,\ \langle \bar{\Lambda}_2,\ \beta\mathscr{I}_n - \bar{X}\rangle = 0. \end{array} \right. \tag{3.3}$$

Denote $B_q := \{v \in \mathbb{R}^s : \|v_l\| \le w_l, l = 1, 2, \cdots, r\}$, where $1/q + 1/p = 1$ and $q \ge 0$. Therefore, problem (3.2) can be further reformulated equivalently as

$$\begin{array}{ll} \min\limits_{u,v,Z,\Lambda_1,\Lambda_2} & -\log\det Z - \langle u, b\rangle - \alpha\langle \Lambda_1, I\rangle + \beta\langle \Lambda_2, I\rangle \\ \text{s.t.} & \mathscr{A}^* u + \mathscr{P}^* v + \Lambda_1 - \Lambda_2 + Z = S, \\ & u \in \mathbb{R}^m,\ v \in \mathscr{B}_q,\ Z \in \mathscr{S}^n_{++},\ \Lambda_1 \in \mathscr{S}^n_+,\ \Lambda_2 \in \mathscr{S}^n_+. \end{array} \tag{3.4}$$

The dual formulation (3.4) seems to be much more complicated because more variables are involved. In the subsequent, we show that it can be solved efficiently by using the semi-proximal augmented Lagrangian method based on a Jacobian iteration.

3.2. **Using semi-proximal augmented Lagrangian method to problem (3.4).** The augmented Lagrangian function of (3.4) is

$$\begin{aligned}\mathscr{L}_\sigma(u,v,Z,\Lambda_1,\Lambda_2;X) = &-\log\det Z - \langle u,b\rangle - \alpha\langle\Lambda_1,I\rangle + \beta\langle\Lambda_2,I\rangle\\ &+\langle X,\mathscr{A}^*u+\mathscr{P}^*v+\Lambda_1-\Lambda_2+Z-S\rangle\\ &+\frac{\sigma}{2}\|\mathscr{A}^*u+\mathscr{P}^*v+\Lambda_1-\Lambda_2+Z-S\|_F^2,\end{aligned}$$

where $\sigma>0$ is a penalty parameter. Considering that the multiplier of dual problem (3.2) is the argument of primal problem (1.5), it is enough to take into account the case of $X\in\mathscr{S}^n_{++}$. Give an initial points $(u^0,v^0,Z^0,\Lambda_1^0,\Lambda_2^0;X^0)\in\mathbb{R}^m\times\mathscr{B}_q\times\mathscr{S}^n_{++}\times\mathscr{S}^n_+\times\mathscr{S}^n_+\times\mathscr{S}^n_{++}$. The directly extended ADMM implemented in [16] updated each variable in a single Gauss-Seidel manner, i.e., using the order $u\to v\to\Lambda_1\to\Lambda_2\to Z$. However, it was shown that this direct extension does not necessarily convergent [4] although it performs numerically well. The second variant of ADMM in [16] updated each variable by using the combination of Gauss-Seidel and Jacobian iteration, i.e., using the order $u\to v\rightrightarrows\Lambda_1\rightrightarrows\Lambda_2\to u\to Z$. It is known that this update approach is equivalent to the two-block semi-proximal ADMM and hence its convergence is followed directly.

Unlike both mentioned approaches, we view all the variable together as one group, and then update each variable is a parallel way. In addition, a semi-proximal point term is added to each subproblem to ensure that each subproblem admits closed-form solution. The iterative scheme is summarized as follows

$$\begin{cases}u^{k+1}=\underset{u\in\mathbb{R}^m}{\arg\min}\left\{\mathscr{L}_\sigma(u,v^k,Z^k,\Lambda_1^k,\Lambda_2^k;X^k)+\frac{1}{2}\|u-u^k\|^2_{T_1}\right\},\\ v^{k+1}=\underset{v\in\mathscr{B}_q}{\arg\min}\left\{\mathscr{L}_\sigma(u^k,v,Z^k,\Lambda_1^k,\Lambda_2^k;X^k)+\frac{1}{2}\|v-v^k\|^2_{T_2}\right\},\\ Z^{k+1}=\underset{Z\in\mathscr{S}^n_{++}}{\arg\min}\left\{\mathscr{L}_\sigma(u^k,v^k,Z,\Lambda_1^k,\Lambda_2^k;X^k)+\frac{1}{2}\|Z-Z^k\|^2_{T_3}\right\},\\ \Lambda_1^{k+1}=\underset{\Lambda_1\in\mathscr{S}^n_+}{\arg\min}\left\{\mathscr{L}_\sigma(u^k,v^k,Z^k,\Lambda_1,\Lambda_2^k;X^k)+\frac{1}{2}\|\Lambda_1-\Lambda_1^k\|^2_{T_4}\right\},\\ \Lambda_2^{k+1}=\underset{\Lambda_2\in\mathscr{S}^n_+}{\arg\min}\left\{\mathscr{L}_\sigma(u^k,v^k,Z^k,\Lambda_1^k,\Lambda_2;X^k)+\frac{1}{2}\|\Lambda_2-\Lambda_2^k\|^2_{T_5}\right\},\\ X^{k+1}=X^k+\tau\sigma(\mathscr{A}^*u^{k+1}+\mathscr{P}^*v^{k+1}+\Lambda_1^{k+1}-\Lambda_2^{k+1}+Z^{k+1}-S),\end{cases}\tag{3.5}$$

where $\mu\geq 4$, $\tau\in(0,2)$, and $T_1=\mu\sigma\mathscr{A}\mathscr{A}^*$, $T_2=\mu\sigma\mathscr{P}\mathscr{P}^*$, $T_3=T_4=T_5=\mu\sigma I$ are self-adjoint positive semi-definite operators. Clearly, the computational burden mainly lies in the solving of the subproblems. Similar to the derivations of [16], we show that each subproblem enjoys closed-form solution, which makes the algorithm easy to implement.

First, we get, for every $k=0,1,\cdots$, that

$$\begin{aligned}u^{k+1}=\underset{u\in\mathbb{R}^m}{\arg\min}\Big\{&\langle u,\mathscr{A}X^k-b\rangle+\frac{\sigma(1+\mu)}{2}\Big\|\mathscr{A}^*u\\ &+\frac{1}{1+\mu}(-\mu\mathscr{A}^*u^k+\mathscr{P}^*v^k+\Lambda_1^k-\Lambda_2^k+Z^k-S)\Big\|_F^2\Big\}.\end{aligned}$$

Therefore, the solution u^{k+1} is given explicitly by

$$u^{k+1}=\frac{1}{\sigma(1+\mu)}(\mathscr{A}\mathscr{A}^*)^{-1}\Big(b-\mathscr{A}X^k-\sigma\mathscr{A}(-\mu\mathscr{A}^*u^k+\mathscr{P}^*v^k+\Lambda_1^k-\Lambda_2^k+Z^k-S)\Big).$$

Second, for the v-subproblem in (3.5), we choose $g_i \cap g_j = \emptyset$ for all $1 \le i < j \le r$. Then from the definition of $\mathscr{P}$, we have $\mathscr{P}\mathscr{P}^* = I$. In view of $T_2 = \mu\sigma\mathscr{P}\mathscr{P}^*$, we have that the v-subproblem can be transformed into

$$\begin{aligned} v^{k+1} = \arg\min_{v\in\mathscr{B}_q} \Big\{ \langle \mathscr{P}X^k, v\rangle + \frac{\sigma(1+\mu)}{2}\Big\| \mathscr{P}^* v \\ + \frac{1}{1+\mu}(\mathscr{A}^* u^k - \mu\mathscr{P}^* v^k + \Lambda_1^k - \Lambda_2^k + Z^k - S)\Big\|_F^2 \Big\}, \end{aligned}$$

which can be written explicitly as

$$v^{k+1} = \Pi_{\mathscr{B}_q}\Big(-\frac{1}{\sigma(1+\mu)}\mathscr{P}X^k - \frac{1}{1+\mu}\mathscr{P}(\mathscr{A}^* u^k - \mu\mathscr{P}^* v^k + \Lambda_1^k - \Lambda_2^k + Z^k - S)\Big).$$

Third, with $T_3 = \mu\sigma I$, then the Z-subproblem which based on Proposition 2.1 can be given by

$$\begin{aligned} Z^{k+1} &= \arg\min_{Z\in\mathscr{S}_{++}^n} \Big\{ -\log\det Z + \frac{\sigma(1+\mu)}{2}\Big\| Z + \frac{1}{1+\mu}(\mathscr{A}^* u^k + \mathscr{P}^* v^k + \Lambda_1^k - \Lambda_2^k - S - \mu Z^k) \\ &\quad + \frac{1}{\sigma(1+\mu)}X^k\Big\|_F^2 \Big\} \\ &= \Phi_{\frac{1}{\sigma(1+\mu)}}\Big(-\frac{1}{1+\mu}(\mathscr{A}^* u^k + \mathscr{P}^* v^k + \Lambda_1^k - \Lambda_2^k - S - \mu Z^k) - \frac{1}{\sigma(1+\mu)}X^k\Big). \end{aligned}$$

Last, with given $T_4 = T_5 = \mu\sigma I$, the Λ_1- and Λ_2-subproblems can be reformulated, respectively, as

$$\begin{aligned} \Lambda_1^{k+1} &= \arg\min_{\Lambda_1\in\mathscr{S}_+^n} \Big\{ \frac{\sigma(1+\mu)}{2}\Big\| \Lambda_1 + \frac{1}{1+\mu}(\mathscr{A}^* u^k + \mathscr{P}^* v^k - \mu\Lambda_1^k - \Lambda_2^k + Z^k - S) \\ &\quad + \frac{1}{\sigma(1+\mu)}(X^k - \alpha I)\Big\|_F^2 \Big\} \\ &= \Pi_{\mathscr{S}_+^n}\Big(\frac{1}{\sigma(1+\mu)}(\alpha I - X^k) - \frac{1}{1+\mu}(\mathscr{A}^* u^k + \mathscr{P}^* v^k - \mu\Lambda_1^k - \Lambda_2^k + Z^k - S)\Big), \end{aligned}$$

and

$$\begin{aligned} \Lambda_2^{k+1} &= \arg\min_{\Lambda_2\in\mathscr{S}_+^n} \Big\{ \frac{\sigma(1+\mu)}{2}\Big\| \Lambda_2 - \frac{1}{1+\mu}(\mathscr{A}^* u^k + \mathscr{P}^* v^k + \Lambda_1^k + \mu\Lambda_2^k + Z^k - S) \\ &\quad + \frac{1}{\sigma(1+\mu)}(\beta I - X^k)\Big\|_F^2 \Big\} \\ &= \Pi_{\mathscr{S}_+^n}\Big(\frac{1}{1+\mu}(\mathscr{A}^* u^k + \mathscr{P}^* v^k + \Lambda_1^k + \mu\Lambda_2^k + Z^k - S) + \frac{1}{\sigma(1+\mu)}(X^k - \beta I)\Big). \end{aligned}$$

From the above analysis, we see that the above iterative framework is easily performed in sense that each subproblem enjoys closed-form solution. So, we are ready to present the iterative framework of the semi-proximal augmented Lagrangian method (abbr. SP-ALM) for solving problem (3.4).

Algorithm: (SP-ALM)

Initialization: Choose $\tau \in (0,2)$, $\mu \geq 4$ and $\sigma > 0$. let $(u^0, v^0, Z^0, \Lambda_1^0, \Lambda_2^0; X^0) \in \mathbb{R}^m \times \mathscr{B}_q \times \mathscr{S}^n_{++} \times \mathscr{S}^n_+ \times \mathscr{S}^n_+ \times \mathscr{S}^n_{++}$ be the initial point. For $k = 0, 1, \cdots$, perform the following steps iteratively:

Step 1. Compute

$$\begin{cases} u^{k+1} = \frac{1}{\sigma(1+\mu)}(\mathscr{A}\mathscr{A}^*)^{-1}\big(b - \mathscr{A}X^k - \sigma\mathscr{A}(-\mu\mathscr{A}^*u^k + \mathscr{P}^*v^k + \Lambda_1^k - \Lambda_2^k + Z^k - S)\big), \\ v^{k+1} = \Pi_{\mathscr{B}_q}\Big(-\frac{1}{\sigma(1+\mu)}\mathscr{P}X^k - \frac{1}{1+\mu}\mathscr{P}(\mathscr{A}^*u^k - \mu\mathscr{P}^*v^k + \Lambda_1^k - \Lambda_2^k + Z^k - S)\Big), \\ Z^{k+1} = \Phi_{\frac{1}{\sigma(1+\mu)}}\Big(-\frac{1}{1+\mu}(\mathscr{A}^*u^k + \mathscr{P}^*v^k + \Lambda_1^k - \Lambda_2^k - S - \mu Z^k) - \frac{1}{\sigma(1+\mu)}X^k\Big), \\ \Lambda_1^{k+1} = \Pi_{\mathscr{S}^n_+}\Big(\frac{1}{\sigma(1+\mu)}(\alpha I - X^k) - \frac{1}{1+\mu}(\mathscr{A}^*u^k + \mathscr{P}^*v^k - \mu\Lambda_1^k - \Lambda_2^k + Z^k - S)\Big), \\ \Lambda_2^{k+1} = \Pi_{\mathscr{S}^n_+}\Big(\frac{1}{1+\mu}(\mathscr{A}^*u^k + \mathscr{P}^*v^k + \Lambda_1^k + \mu\Lambda_2^k + Z^k - S) + \frac{1}{\sigma(1+\mu)}(X^k - \beta I)\Big). \end{cases}$$

Step 2. Compute

$$X^{k+1} = X^k + \tau\sigma(\mathscr{A}^*u^{k+1} + \mathscr{P}^*v^{k+1} + \Lambda_1^{k+1} - \Lambda_2^{k+1} + Z^{k+1} - S).$$

3.3. **Convergence analysis.** This section targets to establish the global convergence of the SP-ALM to solve (3.4) by linking it to the basic semi-proximal augmented Lagrangian (2.7). For this purpose, we report a lemma, which is essential for the convergence theorem. The proof of this lemma is very simple and insightful from Jenson's inequality.

Lemma 3.1. *Let $m \geq 1$ be a positive integer, and $\mathscr{X} = \mathscr{X}_1 \otimes \mathscr{X}_2 \otimes \cdots \otimes \mathscr{X}_m$ and $\mathscr{Z}$ be real finite dimensional Euclidean spaces. Let $\mathscr{A}_i : Z \to \mathscr{X}_i$ be a linear operator with its adjoint $\mathscr{A}_i^* : \mathscr{X}_i \to Z$ for $i = 1, \cdots, m$. Define $\mathscr{A}^*x = \mathscr{A}_1^*x_1 + \mathscr{A}_2^*x_2 + \cdots + \mathscr{A}_m^*x_m$ and set $\mathscr{T} := (1+\mu)Diag(\mathscr{A}\mathscr{A}^*) - \mathscr{A}\mathscr{A}^*$. Then $\mathscr{T} \succeq 0$ for any $\mu \geq m-1$.*

Proof. Using the Jenson's inequality, we get

$$\langle \mathscr{A}^*x, \mathscr{A}^*x\rangle = \|\mathscr{A}_1^*x_1 + \mathscr{A}_2^*x_2 + \cdots + \mathscr{A}_m^*x_m\|^2 \leq m\sum_{i=1}^m \|\mathscr{A}_i^*x_i\|^2.$$

For any $x \in \mathscr{X}$ and $\mu \geq m-1$, we have

$$\langle x, \mathscr{T}x\rangle = (1+\mu)\sum_{i=1}^m \langle x_i, \mathscr{A}_i\mathscr{A}_i^*x_i\rangle - \langle x, \mathscr{A}\mathscr{A}^*x\rangle \geq m\sum_{i=1}^m \|\mathscr{A}_i^*x_i\|^2 - \langle \mathscr{A}^*x, \mathscr{A}^*x\rangle \geq 0,$$

which indicates $\mathscr{T}$ is positive semi-definite. □

Next, we give the following lemma, which is fundamental for the global convergence of the SP-ALM.

Lemma 3.2. *Define $\mathscr{M} := [\mathscr{A}; \mathscr{P}; \mathscr{I}; \mathscr{I}; -\mathscr{I}]$, $Y := [u; v; Z; \Lambda_1; \Lambda_2]$, $\mathscr{Y} := \mathbb{R}^m \times \mathscr{B}_q \times \mathscr{S}^n_{++} \times \mathscr{S}^n_+ \times \mathscr{S}^n_+ \times \mathscr{S}^n_{++}$, and $\mathscr{M}^* = [\mathscr{A}^*; \mathscr{P}^*; \mathscr{I}; \mathscr{I}; -\mathscr{I}]$. Then the $[u; v; Z; \Lambda_1; \Lambda_2]$-subproblem in (3.5) can be reformulated as*

$$Y^{k+1} = \arg\min_{Y \in \mathscr{Y}} \Big\{ f(Y) + \langle X^k, \mathscr{M}^*Y - S\rangle + \frac{\sigma}{2}\|\mathscr{M}^*Y - S\|_F^2 + \frac{1}{2}\|Y - Y^k\|_{\mathscr{T}}^2 \Big\}, \tag{3.6}$$

where $f(Y) = -\log\det Z - \langle u,b\rangle - \alpha\langle\Lambda_1, I\rangle + \beta\langle\Lambda_2, I\rangle$, *and* $\mathscr{T} = \sigma\big((1+\mu)Diag(\mathscr{M}\mathscr{M}^*) - \mathscr{M}\mathscr{M}^*\big)$, $\mu \geq 4$, *or, equivalently,*

$$\begin{aligned}&[u^{k+1}; v^{k+1}; Z^{k+1}; \Lambda_1^{k+1}; \Lambda_2^{k+1}]\\ &= \underset{u,v,Z,\Lambda_1,\Lambda_2}{\arg\min}\ \mathscr{L}_\sigma(u,v,Z,\Lambda_1,\Lambda_2;X^k) + \frac{1}{2}\Big\|[u;v;Z;\Lambda_1;\Lambda_2] - [u^k;v^k;Z^k;\Lambda_1^k;\Lambda_2^k]\Big\|_{\mathscr{T}}^2.\end{aligned}$$

Proof. By the first order optimal condition of unconstrained optimization, the iterative scheme (3.5) without constrains can be expressed by

$$\begin{cases} 0 = \mathscr{A}X^k - b + \sigma\mathscr{A}(\mathscr{A}^*u^{k+1} + \mathscr{P}^*v^k + \Lambda_1^k - \Lambda_2^k + Z^k - S) + T_1(u^{k+1} - u^k),\\ 0 = \mathscr{P}X^k + \sigma\mathscr{P}(\mathscr{A}^*u^k + \mathscr{P}^*v^{k+1} + \Lambda_1^k - \Lambda_2^k + Z^k - S) + T_2(v^{k+1} - v^k),\\ 0 = -(Z^{k+1})^{-1} + X^k + \sigma(\mathscr{A}^*u^k + \mathscr{P}^*v^k + \Lambda_1^k - \Lambda_2^k + Z^{k+1} - S) + T_3(Z^{k+1} - Z^k),\\ 0 = X^k - \alpha I + \sigma(\mathscr{A}^*u^k + \mathscr{P}^*v^k + \Lambda_1^{k+1} - \Lambda_2^k + Z^k - S) + T_4(\Lambda_1^{k+1} - \Lambda_1^k),\\ 0 = \beta I - X^k - \sigma(\mathscr{A}^*u^k + \mathscr{P}^*v^k + \Lambda_1^k - \Lambda_2^{k+1} + Z^k - S) + T_5(\Lambda_2^{k+1} - \Lambda_2^k),\end{cases} \tag{3.7}$$

which can be rewritten equivalently as

$$\begin{aligned}0 = &\begin{pmatrix} -b\\ 0\\ -(Z^{k+1})^{-1}\\ -\alpha I\\ \beta I\end{pmatrix} + \begin{pmatrix}\mathscr{A}\\ \mathscr{P}\\ I\\ I\\ -I\end{pmatrix}X^k + \sigma\begin{pmatrix}\mathscr{A}\\ \mathscr{P}\\ I\\ I\\ -I\end{pmatrix}\begin{pmatrix}(\mathscr{A}^*, \mathscr{P}^*, I, I, -I)\begin{pmatrix}u^{k+1}\\ v^{k+1}\\ Z^{k+1}\\ \Lambda_1^{k+1}\\ \Lambda_2^{k+1}\end{pmatrix} - S\end{pmatrix}\\ &- \sigma\begin{pmatrix}\mathscr{A}\big(\mathscr{P}^*(v^{k+1}-v^k) + (\Lambda_1^{k+1}-\Lambda_1^k) - (\Lambda_2^{k+1}-\Lambda_2^k) + (Z^{k+1}-Z^k)\big)\\ \mathscr{P}\big(\mathscr{A}^*(u^{k+1}-u^k) + (\Lambda_1^{k+1}-\Lambda_1^k) - (\Lambda_2^{k+1}-\Lambda_2^k) + (Z^{k+1}-Z^k)\big)\\ \mathscr{A}^*(u^{k+1}-u^k) + \mathscr{P}^*(v^{k+1}-v^k) + (\Lambda_1^{k+1}-\Lambda_1^k) - (\Lambda_2^{k+1}-\Lambda_2^k)\\ \mathscr{A}^*(u^{k+1}-u^k) + \mathscr{P}^*(v^{k+1}-v^k) - (\Lambda_2^{k+1}-\Lambda_2^k) + (Z^{k+1}-Z^k)\\ -\big(\mathscr{A}^*(u^{k+1}-u^k) + \mathscr{P}^*(v^{k+1}-v^k) + (\Lambda_1^{k+1}-\Lambda_1^k) + (Z^{k+1}-Z^k)\big)\end{pmatrix}\\ &+ \begin{pmatrix}T_1(u^{k+1}-u^k)\\ T_2(v^{k+1}-v^k)\\ T_3(Z^{k+1}-Z^k)\\ T_4(\Lambda_1^{k+1}-\Lambda_1^k)\\ T_5(\Lambda_2^{k+1}-\Lambda_2^k)\end{pmatrix}.\end{aligned} \tag{3.8}$$

(3.9)

By the definitions of $\mathscr{M}$, Y and $\mathscr{M}^*$, (3.8) can be further reformulated as

$$\begin{aligned} 0 = &\begin{pmatrix} -b \\ 0 \\ -(Z^{k+1})^{-1} \\ -\alpha I \\ \beta I \end{pmatrix} + \mathscr{M}X^k + \sigma\mathscr{M}(\mathscr{M}^*Y^{k+1} - S) \\ &+ \begin{pmatrix} T_1 & -\sigma\mathscr{A}\mathscr{P}^* & -\sigma\mathscr{A} & -\sigma\mathscr{A} & \sigma\mathscr{A} \\ -\sigma\mathscr{P}\mathscr{A}^* & T_2 & -\sigma\mathscr{P} & -\sigma\mathscr{P} & \sigma\mathscr{P} \\ -\sigma\mathscr{A}^* & -\sigma\mathscr{P}^* & T_3 & -\sigma\mathscr{I} & \sigma\mathscr{I} \\ -\sigma\mathscr{A}^* & -\sigma\mathscr{P}^* & -\sigma\mathscr{I} & T_4 & \sigma\mathscr{I} \\ \sigma\mathscr{A}^* & \sigma\mathscr{P}^* & \sigma\mathscr{I} & \sigma\mathscr{I} & T_5 \end{pmatrix}(Y^{k+1} - Y^k). \end{aligned} \tag{3.10}$$

Define

$$\mathscr{T} := \begin{pmatrix} T_1 & -\sigma\mathscr{A}\mathscr{P}^* & -\sigma\mathscr{A} & -\sigma\mathscr{A} & \sigma\mathscr{A} \\ -\sigma\mathscr{P}\mathscr{A}^* & T_2 & -\sigma\mathscr{P} & -\sigma\mathscr{P} & \sigma\mathscr{P} \\ -\sigma\mathscr{A}^* & -\sigma\mathscr{P}^* & T_3 & -\sigma\mathscr{I} & \sigma\mathscr{I} \\ -\sigma\mathscr{A}^* & -\sigma\mathscr{P}^* & -\sigma\mathscr{I} & T_4 & \sigma\mathscr{I} \\ \sigma\mathscr{A}^* & \sigma\mathscr{P}^* & \sigma\mathscr{I} & \sigma\mathscr{I} & T_5 \end{pmatrix}.$$

From the definition of $f(Y)$, we obtain $\nabla f(Y) = (-b; 0; -(Z)^{-1}; -\alpha I; \beta I)^\top$. Furthermore, (3.10) is reformulated as

$$0 = \nabla f(Y^{k+1}) + \mathscr{M}X^k + \sigma\mathscr{M}(\mathscr{M}^*Y^{k+1} - S) + \mathscr{T}(Y^{k+1} - Y^k). \tag{3.11}$$

In addition, substituting $T_1 = \mu\sigma\mathscr{A}\mathscr{A}^*$, $T_2 = \mu\sigma\mathscr{P}\mathscr{P}^*$, $T_3 = T_4 = T_5 = \mu\sigma I$ into $\mathscr{T}$, we can easily get $\mathscr{T} := \sigma\big((1+\mu)\mathrm{Diag}(\mathscr{M}\mathscr{M}^*) - \mathscr{M}\mathscr{M}^*\big)$. And according to Lemma 3.1, $\mathscr{T} \succeq 0$ for $\mu \geq 4$. Notice that the optimal condition of (3.5) is exactly the form of (3.11). This completes the proof. □

The formulation of (3.6) falls into the framework of the semi-proximal augmented Lagrangian method of (2.7). So, its convergence result is followed directly from [7]. To end this section, we are ready to state the convergence results for Algorithm SP-ALM for solving (3.4) without proof.

Theorem 3.1. *Let the sequence* $\{(u^k, v^k, Z^k, \Lambda_1^k, \Lambda_2^k; X^k)\}$ *be generated by Algorithm SP-ALM from an initial point* $(u^0, v^0, Z^0, \Lambda_1^0, \Lambda_2^0; X^0) \in \mathbb{R}^m \times \mathscr{B}_q \times \mathscr{S}^n_{++} \times \mathscr{S}^n_+ \times \mathscr{S}^n_+ \times \mathscr{S}^n_{++}$. *Assume that* $\tau \in (0,2)$ *and* $\mu \geq 4$, *then the sequence* $\{(u^k, v^k, Z^k, \Lambda_1^k, \Lambda_2^k; X^k)\}$ *converges to a cluster point* $(\bar{u}, \bar{v}, \bar{Z}, \bar{\Lambda}_1, \bar{\Lambda}_2; \bar{X})$ *where* $\bar{X}$ *is an optimal solution of (1.5), and* $(\bar{u}, \bar{v}, \bar{Z}, \bar{\Lambda}_1, \bar{\Lambda}_2)$ *is an optimal solution of dual problem (3.4).*

4. Numerical Experiments

In this section, we present some limited numerical experiments to evaluate the feasibility and efficiency of SP-ALM by using some random synthetic data. All the experiments are performed with Microsoft Windows 7 and MATLAB R2016a, and run on a PC with an Intel Core i5-4200U CPU at 1.60 GHz and 4 GB of memory. In this test, we use "*pobj*" and "*dobj*" to denote the

original and dual objective function value, respectively,

$$pobj := \langle S, X\rangle - \operatorname{logdet} X + \sum_{l=1}^{r} \omega_l \|X_{\{g_l\}}\|_p,$$

and

$$dobj := n + \operatorname{logdet} Z + \langle u, b\rangle + \alpha\langle \Lambda_1, \mathscr{I}_n\rangle - \beta\langle \Lambda_2, \mathscr{I}_n\rangle.$$

According to (3.3), we use the KKT residuals to measure the quality of the derived solution, i.e.,

$$Res := \max\{R_P, R_D, R_G\} < 5e-3,$$

where R_P represents the primal infeasibility of (1.5), R_D presents the dual infeasibility of (3.6), and R_G presents the relative duality gap, which are defined as follows, respectively,

$$R_P := \max\Big\{\eta_{P1}, \eta_{P2}, \eta_{P3}\Big\},$$

$$R_D := \max\Big\{\eta_{D1}, \eta_{D2}, \eta_{D3}, \eta_{D4}, \eta_{D5}, \eta_{D6}, \eta_{D7}, \eta_{D8}\Big\},$$

and

$$R_G := \frac{|pobj - dobj|}{1 + |pobj| + |dobj|},$$

where

$$\eta_{P1} := \frac{\|\mathscr{A}X - b\|}{1 + \|b\|},$$

$$\eta_{P2} := \frac{\|X - \alpha\mathscr{I}_n - \Pi_{\mathscr{S}_+^n}(X - \alpha\mathscr{I}_n)\|}{1 + \|X - \alpha\mathscr{I}_n\|_F},$$

$$\eta_{P3} := \frac{\|\beta\mathscr{I}_n - X - \Pi_{\mathscr{S}_+^n}(\beta\mathscr{I}_n - X)\|_F}{1 + \|\beta\mathscr{I}_n - X\|_F},$$

and

$$\eta_{D1} := \frac{\|\mathscr{A}^* u + \mathscr{P}^* v + \Lambda_1 - \Lambda_2 + Z - S\|_F}{1 + \|S\|_F},$$

$$\eta_{D2} := \frac{\|\Lambda_1 - \Pi_{S_+^n}(\Lambda_1)\|_F}{1 + \|\Lambda_1\|_F},$$

$$\eta_{D3} := \frac{\|\Lambda_2 - \Pi_{S_+^n}(\Lambda_2)\|_F}{1 + \|\Lambda_2\|_F},\ \eta_{D4} := \frac{\|Z - \Pi_{S_{++}^n}(Z)\|_F}{1 + \|Z\|_F},$$

$$\eta_{D5} := \frac{\|XZ - \mathscr{I}_n\|_F}{1 + n},$$

$$\eta_{D6} := \frac{|\langle \Lambda_1, X - \alpha\mathscr{I}_n\rangle|}{1 + \|\Lambda_1\|_F + \|X - \alpha\mathscr{I}_n\|_F},$$

and

$$\eta_{D7} := \frac{|\langle \Lambda_2, \beta\mathscr{I}_n - X\rangle|}{1 + \|\Lambda_2\|_F + \|\beta\mathscr{I}_n - X\|_F},\ \eta_{D8} := \frac{\|v + \Pi_{B_q}(-v + \mathscr{P}X)}{1 + \|v\| + \|\mathscr{P}X\|}.$$

In addition, we also terminate the iterative process when the number of iterations exceeds 2000 without achieving convergence. The penalty parameter is initialized $\sigma_0 = 1$ and updated occasionally according to the similar adjustment in [17],

$$\sigma_{k+1} = \begin{cases} \min\{2\sigma_k, 10^7\} & \text{if} \quad R_P/R_D < 0.1, \\ \max\{0.5\sigma_k, 10^{-2}\} & \text{if} \quad R_P/R_D > 10, \\ \sigma_k & \text{otherwise.} \end{cases} \tag{4.1}$$

We set $\alpha = 1$ and $\beta = 100$ to fix the lower and upper bounds of the eigenvalues of the solution, and choose $\mu = 4 + 1/n$ and $\tau = 1.6$ from Theorem 3.1.

We now describe the procedure of generating a true sparse inverse covariance matrix, and then describe how to construct the sample covariance S. We set that random variables $x \sim N(0, \Sigma)$ with given Σ and then divide it into r groups, where the number of elements in each group is about n/r. We set the connection probability between two nodes from the same group to p_1. While for any two different groups, the connection probability is set to p_2. When the different groups indeed exist connections, we set the connection probability of two nodes to p_3. In this test, we choose $p_1 = 0.8$, $p_2 = 0.2$ and $p_3 = 0.5$. We generate ζ random variable and then construct the sample covariance matrix S as well as adding some Gauss noise. Equality constraints essentially represent pre-known zeros which set about 25% of the elements to be known in this test.

Firstly, we test the effectiveness of SP-ALM with fixed $n = 150$ and different p. The results are listed in Figure 1. Comparing each plot, it is easy to see that the proposed algorithm can effectively solve the problem of sparse inverse covariance matrix estimation, the group lasso regularization with $p = 2$ and $p = \infty$ can give better results than $p = 1$.

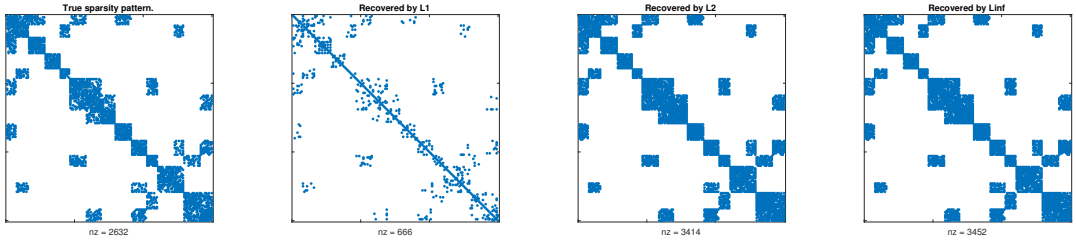

FIGURE 1. The sparsity pattern of the true inverse covariance matrix (left) and those recovered by SP-ALM with ℓ_1 (middle left), ℓ_2 (middle right), and ℓ_∞ (right), respectively. Parameters: $n = 150$, $r = 15$. The tuning parameters are set to be $w_l = w = 0.002$, 0.009, and 0.05 for ℓ_1, ℓ_2, and ℓ_∞, respectively.

Now, we further test the numerical performances of SP-ALM using random simulated data with different problem sizes. In this test, the regularization parameter is set as $w_l \equiv w = 1/n$. Then implement the algorithm with the dimensions from $n = 200$ to 2,000 and with $r = n/10$ number of groups. Since 25% of the zero elements are assumed to be known in advance, this means that the total number of the known zeros which denoted by m are also previously given. The numerical results are reported in Table 1 where "Iter" represents the number of iterations, "Time" represents the computing time, "(R_P, R_D, R_G)" are defined before. It can be seen from this table that almost all problems are solved successfully. Although the stopping criterion $5e-3$ is not achieved at some special cases, the acceptable magnitude -2 is obtained. It is also

TABLE 1. Numerical Results of the SP-ALM on random problems

n	m	p=1			p=2			p=∞		
		Iter	Time	(R_P,R_D,R_G)	Iter	Time	(R_P,R_D,R_G)	Iter	Time	(R_P,R_D,R_G)
200	8812	66	5.4	(2.8-4,3.2-3,4.8-3)	2000	183.1	(1.6-2,5.5-4,6.0-6)	2000	259.8	(6.0-2,4.3-3,8.5-4)
500	55072	35	24.9	(2.0-4,1.2-3,4.4-3)	2000	1701	(7.8-3,5.7-4,3.7-4)	2000	2156	(7.6-2,4.5-3,9.9-6)
800	141546	39	121	(4.9-3,5.0-4,4.5-3)	33	113.7	(4.7-3,2.1-3,1.1-3)	2000	7967	(5.0-2, 1.4-3, 1.9-4)
1000	222685	37	254.5	(3.8-4, 4.5-3, 3.6-3)	32	238.9	(4.9-3,2.4-3,4.6-3)	41	337.1	(7.9-3,6.5-4,4.1-4)
1200	321088	37	469.8	(1.6-3,1.6-3,1.9-3)	37	504.9	(2.5-3,2.4-4,1.4-3)	30	445.9	(2.4-3,2.0-3,1.2-3)
1500	502178	37	974.4	(1.7-3,2.7-4,3.9-4)	37	1028	(2.1-3,2.7-4,6.2-5)	30	893.8	(4.1-3,3.0-3,2.7-3)
1700	646557	37	1454	(1.8-3,2.9-4,2.9-4)	37	1512	(2.2-3,2.9-4,7.0-4)	30	1255	(3.6-3,3.3-3,3.2-3)
2000	894713	32	2079	(4.8-3,3.4-3,1.9-3)	32	2156	(4.9-3,3.4-3,2.6-3)	37	2531	(3.9-3,3.2-4,1.7-3)

observed that when the dimension is not less than 1,000, SP-ALM can produce the solution successfully at each p within dozens of iterations. Taking everything together, this simple test shows that the implemented algorithm performs quite well which provides an promising approach to recover high-dimensional inverse covariance matrix.

5. CONCLUSION

Estimating the high-dimensional inverse covariance matrix is a classic problem in modern multivariate analysis. In the past few years, many optimization algorithms based estimators have been proposed, analyzed, and tested. Among them, the generalized log-determinant model introduced in [16] which includes several existing estimation models as special cases. In addition, Xiao, Li and Lu [16] implemented two types of ADMMs for solving the corresponding five separable variables contained dual model. The first one is an extension of the regular ADMM and the second one is based on the combinations of symmetric Gauss-seidel and classical Jacobian iteration. In this paper, we implemented another different type of algorithm where all the dual variables are treated as a bigger group, and then updated in a parallel way, i.e., the newest value of the each variable is not used immediately until the whole loop is finished. This implemented algorithm is equivalent to the semi-proximal augmented Lagrangian method as shown in Lemma 3.2, which plays an important role in establishing the convergence result. Although the proposed method performs well as shown by the limited numerical experiments, its practical performance and testing against others solvers are deserve further investigating. At last but not at least, using other algorithms for sparse estimating such as the generalized ADMM of [15] is also an interesting research topic.

Acknowledgements
The research of the second author was supported by the National Natural Science Foundation of China (Grant No. 11971149).

REFERENCES

[1] T. W. Anderson, (An introduction to multivariate statistical analysis, 2nd Edition. Technometrics 28 (1958), 180-181.
[2] A. D'Aspremont, O. Banerjee, L.El Ghaoui, First-order methods for sparse covariance selection, SIAM J. Matrix Anal. Appl. 30 (2006), 56-66.
[3] O. Banerjee, L. El Ghaoui, A. D'Aspremont Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data, J. Machine Learning Res. 9 (2008), 485-516.

[4] C. Chen, B. He, Y. Ye, X. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent, Math. Program. 155 (2016), 57-79.
[5] Y. Cui, C. Leng, D.F. Sun, Sparse estimation of high-dimensional correlation matrices, Comput. Stat. Data Anal. 93 (2016), 390-403.
[6] J. Dahl, L. Vandenberghe, V. Roychowdhury, Covariance selection for nonchordal graphs via chordal embedding, Optim. Methods Softw. 23 (2008), 501-520.
[7] M. Fazel, T.K. Pong, D.F. Sun, P. Tseng, Hankel matrix rank minimization with applications to system identification and realization, SIAM J. Matrix Anal. Appl. 34 (2013), 946-977.
[8] M.R. Hestenes, Multiplier and gradient methods, J. Optim. Theory Appl. 4 (1969), 303-320.
[9] X.D. Li, D.F. Sun, K.-C. Toh, A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions, Math. Program. 155 (2016), 333-373.
[10] P. Li, Y. Xiao, An efficient algorithm for sparse inverse covariance matrix estimation based on dual formulation, Comput. Stat. Data Anal. 128 (2018), 292-307.
[11] M.J.D. Powell, A method for nonlinear constraints in minimization problems, Optimization 5 (1969), 283-298.
[12] R. T. Rockafellar, Convex Analysis, Princeton University Press, 1970.
[13] R. T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, Math. Oper. Res. 1 (1976), 97-116.
[14] H. Wang, C. Leng, A note on adaptive group lasso, Comput. Stat. Data Anal. 52 (2008), 5277-5286.
[15] Y. Xiao, L. Chen, D.H. Li, A generalized alternating direction method of multipliers with semi-proximal terms for convex composite conic programming, Math. Program. Comput. 10 (2018), 533-555.
[16] Y. Xiao, P. Li, S. Lu, Efficient algorithms for sparse inverse covariance matrices estimation with explicit eigenvalue constraints, submitted.
[17] J. Yang, D.F. Sun, K.-C. Toh, A proximal point algorithm for log-determinant optimization with group Lasso regularization SIAM J. Optim. 23 (2013), 857-893.
[18] M. Yuan, Y. Lin, Model selection and estimation in the Gaussian graphical model, Biometrika 94 (2007), 19-35.
[19] H. Zou, The adaptive lasso and its oracle properties, Publications of the American Statistical Association, 101 (2006), 1418-1429.