Available online at www.isr-publications.com/jnsa
J. Nonlinear Sci. Appl., 10 (2017), 5038–5051
Research Article

ISSN : 2008-1898

Journal of Nonlinear Sciences and Applications

Journal Homepage: www.tjnsa.com - www.isr-publications.com/jnsa

Proximal ADMM with larger step size for two-block separable convex programming and its application to the correlation matrices calibrating problems

Hongchun Sun[a], Min Sun[b,c,*], Yiju Wang[c]

[a]School of Sciences, Linyi University, Shandong 276005, P. R. China.
[b]School of Mathematics and Statistics, Zaozhuang University, Shandong 277160, P. R. China.
[c]School of Management, Qufu Normal University, Shandong 276826, P. R. China.

Communicated by Y. H. Yao

Abstract

The alternating direction method of multipliers (ADMM) is a benchmark for solving two-block separable convex programming. However, as other first-order iteration methods, the ADMM also suffers from low convergence. In this paper, to accelerate the convergence of the ADMM, the restriction region of the Fortin and Glowinski's constant γ in the ADMM is relaxed from $\left(0, \frac{1+\sqrt{5}}{2}\right)$ to $(0, +\infty)$, thus we get a proximal ADMM with larger step size. By proving some properties of the method, we show its global convergence under mild conditions. Finally, some numerical experiments on the correlation matrices calibrating problems are given to demonstrate the efficiency and the performance of the new method.

Keywords: Alternating direction method of multipliers, the Fortin and Glowinski's constant, global convergence, the correlation matrices calibrating problems.
2010 MSC: 90C25, 90A08.

1. Introduction

In this paper, we are concerned with the following two-block separable convex programming

$$\min\{\theta_1(x_1) + \theta_2(x_2) | A_1x_1 + A_2x_2 = b, x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2\}, \tag{1.1}$$

where $\theta_i : \mathcal{R}^{n_i} \to \mathcal{R}$ $(i = 1, 2)$ are closed proper convex functions (not necessarily smooth), $A_i \in \mathcal{R}^{l\times n_i}(i = 1, 2)$, $b \in \mathcal{R}^l$, and $\mathcal{X}_i \subseteq \mathcal{R}^{n_i}(i = 1, 2)$ are nonempty closed, convex sets. Due to the mathematical generality of $\theta_i(x_i)$ $(i = 1, 2)$, the two-block separable convex programming is very inclusive and plays very important roles in statistical learning [3], image/signal processing [1, 17, 22], traffic equilibrium [24], and so on. For example, it includes the following $\ell_1 - \ell_2$ basis pursuit (BP) model of compressive sensing (CS) as a special case:

$$\begin{aligned} &\min_x \theta(x) = \|x\|_1 + \tfrac{1}{2\mu}\|y\|_2^2 \\ &\text{s.t. } Ax = b, x - y = 0, \end{aligned}$$

*Corresponding author
Email addresses: sunhc68@126.com (Hongchun Sun), ziyouxiaodou@163.com (Min Sun), yiju-wang@163.com (Yiju Wang)

doi:10.22436/jnsa.010.09.40

Received 2017-02-17

where $\mu > 0, A \in \mathcal{R}^{m\times n}(m \ll n)$ is the sensing matrix, $b \in \mathcal{R}^m$ is the observed signal, and the ℓ_1-norm and ℓ_2-norm of the vector x are defined by $\|x\|_1 = \sum_{i=1}^n |x_i|$ and $\|x\|_2 = (\sum_{i=1}^n x_i^2)^{1/2}$, respectively. Hence, in the literature, this problem has been widely studied during the last several decades [13, 17, 18, 21, 24–29, 31, 34]. There has been growing interest in searching for methods to solve (1.1), and various iteration methods have been employed, such as the alternating direction method of multiplier [14], the augmented Lagrangian method [13], the customized proximal point algorithm [19], the customized Douglas-Rachford splitting algorithms [11, 12], the first-order primal-dual proximal method [30] and some other splitting algorithms [10, 15], etc.. We refer the reader to [5] about the relationship among those iteration methods. In this paper, we are going to further study the alternating direction method of multiplier due to its nice features, such as the neat and simple iterative scheme and good numerical performance.

Let $\beta \in (0, +\infty)$ be a given parameter. The augmented Lagrangian function of the convex programming (1.1) can be written as

$$\mathcal{L}_\beta(x_1, x_2, \lambda) := \theta_1(x_1) + \theta_2(x_2) - \langle \lambda, A_1x_1 + A_2x_2 - b\rangle + \frac{\beta}{2}\left\| A_1x_1 + A_2x_2 - b\right\|^2,$$

where $\lambda \in \mathcal{R}^l$ is the Lagrangian multiplier for the linear constraints of (1.1). The classical alternating direction method of multipliers (ADMM) was proposed by Glowinski and Marrocco in 1975 [9]; its iterative scheme for solving the convex programming (1.1) consists of the following recursions: for $k = 0, 1, \ldots$, compute

$$\begin{cases} x_1^{k+1} \in \operatorname{argmin}_{x_1\in\mathcal{X}_1}\mathcal{L}_\beta(x_1, x_2^k, \lambda^k), \\ x_2^{k+1} \in \operatorname{argmin}_{x_2\in\mathcal{X}_2}\mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^k), \\ \lambda^{k+1} = \lambda^k - \beta\left(A_1x_1^{k+1} + A_2x_2^{k+1} - b\right). \end{cases} \tag{1.2}$$

Obviously, the ADMM alternately minimizes the augmented Lagrangian function $\mathcal{L}_\beta(x_1, x_2, \lambda)$ with respect to x_1, then x_2 and then updating the Lagrangian multiplier λ, and at each iteration, the two subproblems of (1.2) are the most computationally expensive task in the ADMM, which turns out to be numerically expensive and are often difficult to deal with even if the sets $\mathcal{X}_1 \neq \mathcal{R}^{n_1}$ and $\mathcal{X}_2 \neq \mathcal{R}^{n_2}$ are very simple and the resolvent operators of $\theta_1(x_1)$ and $\theta_2(x_2)$'s subdifferentials have closed-form representations [11, 12]. Then, various inexact, relaxed and accelerated variants of the ADMM and the Douglas-Rachford splitting algorithms [10–12, 14, 15, 31] are well-studied in recent years. Among them, the proximal ADMM, which goes back to at least [14], is almost the most influential. Then, based on the proximal ADMM and the variant ADMM in [6], Xu et al. [31] presented the iterative scheme of a proximal ADMM with larger step size for (1.1) as follows

$$\begin{cases} x_1^{k+1} \in \operatorname{argmin}_{x_1\in\mathcal{X}_1}\mathcal{L}_\beta(x_1, x_2^k, \lambda^k) + \frac{1}{2}\|x_1 - x_1^k\|^2_{R_1}, \\ x_2^{k+1} \in \operatorname{argmin}_{x_2\in\mathcal{X}_2}\mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^k) + \frac{1}{2}\|x_2 - x_2^k\|^2_{R_2}, \\ \lambda^{k+1} = \lambda^k - \gamma\beta\left(A_1x_1^{k+1} + A_2x_2^{k+1} - b\right), \end{cases} \tag{1.3}$$

where $\gamma \in \left(0, \frac{1+\sqrt{5}}{2}\right)$ is named as the Fortin and Glowinski's constant [8] in the literature and it is also the step size for the update of λ; R_1 and R_2 are two positive semi-definite matrices with compatible dimensionality. In fact, $R_i = \tau_i I_{n_i} - \beta A_i^\top A_i$ with $\tau_i \geqslant \beta\|A_i^\top A_i\| (i = 1, 2)$ in [31]. The iterative scheme (1.3) with $\gamma = 1, R_1 = 0$ and $R_2 = 0$ reduces to the iterative scheme (1.2). Attaching the parameters γ, R_1 and R_2 can confer some advantages for the proximal ADMM. For example, by taking some special cases of R_1 and R_2, the two subproblems of (1.3) are often simple enough to admit closed form solutions or can be solved effectively to high precision [25, 26, 29, 32]. Furthermore, $\gamma > 1$ is usually beneficial to induce fast convergence of the ADMM empirically, which has been verified by the numerical results in [8, 18].

Therefore, a question can be raised: can the restriction region of γ be relaxed? In this paper, we answer this question positively. In fact, we propose a proximal ADMM with larger step size, in which the constant γ can take any values of the interval $(0, +\infty)$, and the expense is that the output of (1.3) must be corrected by the convex combination $w^{k+1} = (1-\rho)w^k + \rho\tilde{w}^k$, where $w^k = (x_1^k, x_2^k, \lambda^k)$ is the current

iterate and $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \tilde{\lambda}^k)$ is the output of (1.3); $\rho > 0$ is a constant. In this sense, our new method can be categorized to the prediction-correction ADMMs [13, 27, 33, 34]. However, our new method is much simpler than the iteration methods in [33, 34], because the later need to compute some dynamically updating step size at each iteration, and need much computational effort. Though the correction steps of the iteration methods in [13, 27] are also some convex combination forms, the feasible regions of ρ in [13, 27] is much smaller than that of our new method; see Remark 3.4 of Section 3. The correction step inevitably increases the computational complexity of the ADMM. However, this strategy also has many superiorities. For example: (1) it can accelerate the convergence of the ADMM [20, 33]; (2) it can make some divergent ADMMs become convergent [13]; (3) it can enlarge the feasible region of the constant γ as we shall show in this paper.

After we accomplished the first version of this paper, we have noticed a quite recent work [16], which presented a proximal ADMM with $R_1 = 0, R_2 = \tau\beta A_2^\top A_2$. They proved that the proximal ADMM with larger step size is global convergent when D_2 is positive definite. Specifically, the two parameters (τ, γ) in [16] are restricted into the domain

$$\mathcal{D} = \left\{(\tau,\gamma)|\tau > 0 \ \ \& \ \ 0 < \gamma < \frac{1-\tau+\sqrt{\tau^2+6\tau+5}}{2}\right\},$$

which is obviously larger than $\left(0, \frac{1+\sqrt{5}}{2}\right)$. The limit of the upper bound of γ is a monotone increasing function of τ, and

$$\lim_{\tau\to+\infty} \frac{1-\tau+\sqrt{\tau^2+6\tau+5}}{2} = 2,$$

which indicates that the feasible region of γ in [16] is still smaller than $(0, +\infty)$, the feasible region of γ in our new method. Furthermore, to get larger step size, the constant τ must be set as a very large value, which makes the proximal terms play a too heavy weight in the objective of the second subproblem, and this eventually leads to slow convergence.

The remainder of this paper is organized as follows. In Section 2, we briefly review some basic concepts and list some necessary assumptions. In Section 3, we propose a new proximal ADMM with larger step size for solving the convex programming (1.1), and prove its global convergence in detail. In Section 4, some numerical experiments on the correlation matrices calibrating problems are given to demonstrate the advantage of the new method. Before starting the next section, we first give some notations used in this paper. The notation $\mathcal{R}^{m\times n}$ stands for the set of all $m \times n$ real matrices. For any two vectors $x, y \in \mathcal{R}^n$, $\langle x, y\rangle = x^\top y$ denotes their inner product; if $G \in \mathcal{R}^{n\times n}$ is a symmetric positive definite matrix, we denote by $\|x\|_G = \sqrt{x^\top Gx}$ the G-norm of the vector x. The effective domain of a function $f : \mathcal{X} \to (-\infty, +\infty]$ is defined as $\mathrm{dom}(f) := \{x \in \mathcal{X}|f(x) < +\infty\}$. The set of all relative interior points of a given nonempty convex set $\mathcal{C}$ is denoted by $\mathrm{ri}(\mathcal{C})$.

2. Preliminaries

In this section, we briefly review some basic concepts and list some necessary assumptions for further analysis.

Definition 2.1 ([23]). A function $f : \mathcal{R}^n \to \mathcal{R}$ is convex if and only if

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y), \ \forall x, y \in \mathcal{R}^n, \alpha \in [0,1].$$

For a convex function $f : \mathcal{R}^n \to \mathcal{R}$, we have the following basic inequality

$$f(x) \geqslant f(y) + \langle \xi, x-y\rangle, \ \ \forall x, y \in \mathcal{R}^n, \xi \in \partial f(y),$$

where $\partial f(y) = \{\xi \in \mathcal{R}^n : f(\bar{y}) \geqslant f(y) + \langle \xi, \bar{y} - y\rangle$ for all $\bar{y} \in \mathcal{R}^n\}$ denotes the subdifferential of $f(\cdot)$ at the point y.

Definition 2.2 ([2]). Let $\Omega \subseteq \mathcal{R}^n$ be a nonempty set. The sequence $\{t^k\} \subseteq \mathcal{R}^n$ is said to be Fejér monotone with respect to Ω if

$$\|t^{k+1} - t\| \leqslant \|t^k - t\|, \quad \forall t \in \Omega, k \geqslant 1.$$

We make the following standard assumptions about (1.1) in this paper.

Assumption 2.3. The functions $\theta_i(\cdot)$ $(i = 1, 2)$ are convex.

Assumption 2.4. The matrices A_i $(i = 1, 2)$ are full-column rank.

Assumption 2.5. The generalized Slater's condition holds, i.e., there is a point $(\hat{x}_1, \hat{x}_2) \in \mathrm{ri}(\mathrm{dom}\theta_1 \times \mathrm{dom}\theta_2) \cap P$, where

$$P := \{x = (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 | A_1x_1 + A_2x_2 = b\}.$$

Under Assumption 2.5, it follows from Theorem 3.22 and Theorem 3.23 of [7] that the vector $x^* = (x_1^*, x_2^*) \in \mathcal{R}^{n_1+n_2}$ is an optimal solution to the convex programming (1.1) if and only if there exists a vector $\lambda^* \in \mathcal{R}^l$ such that

$$\begin{cases} (x_1^*, x_2^*) \in \mathcal{X}_1 \times \mathcal{X}_2, \\ \theta_i(x_i) - \theta_i(x_i^*) + (x_i - x_i^*)^\top(-A_i^\top\lambda^*) \geqslant 0, \quad \forall x_i \in \mathcal{X}_i, i = 1, 2, \\ A_1x_1^* + A_2x_2^* = b. \end{cases} \tag{2.1}$$

Moreover, any $\lambda^* \in \mathcal{R}^l$ satisfying (2.1) is an optimal solution to the dual of the convex programming (1.1). Obviously, the system (2.1) is equivalent to the following mixed variational inequality problem, denoted by VI$(\mathcal{W}, F, \theta)$: find a vector $w^* \in \mathcal{W}$ such that

$$\theta(x) - \theta(x^*) + (w - w^*)^\top F(w^*) \geqslant 0, \quad \forall w \in \mathcal{W},$$

where $\theta(x) = \theta_1(x_1) + \theta_2(x_2)$, $\mathcal{W} = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{R}^l$, and

$$F(w) := \begin{pmatrix} -A_1^\top\lambda \\ -A_2^\top\lambda \\ A_1x_1 + A_2x_2 - b \end{pmatrix} = \begin{pmatrix} 0 & 0 & -A_1^\top \\ 0 & 0 & -A_2^\top \\ A_1 & A_2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \lambda \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ b \end{pmatrix}.$$

The set of the solutions of VI$(\mathcal{W}, F, \theta)$, denoted by $\mathcal{W}^*$, is nonempty by Assumption 2.5. It is easy to verify that the mapping $F(\cdot)$ is not only monotone but also satisfies the following nice property

$$(w' - w)^\top(F(w') - F(w)) = 0, \quad \forall w', w \in \mathcal{W}.$$

3. Proximal ADMM with larger step size and its global convergence

In this section, we pay attention to present a proximal ADMM with larger step size to solve (1.1). Then we give an analysis of the convergence of the constructed method under Assumptions 2.3-2.5. At each iteration, the new method is composed of two steps: the prediction step and the correction step. More specifically, it first generates a trial iterate via the iterative scheme (1.3) in the prediction step, and then yields the new iterate via the convex combination of the current iterate with the trial iterate in the contraction step.

First, let us define an important constant η, which is used in our new method.

$$\eta = \begin{cases} \gamma, & \text{if } 0 < \gamma \leqslant 1, \\ \frac{1}{\gamma}, & \text{if } \gamma > 1. \end{cases} \tag{3.1}$$

Obviously, the constant $\eta \in (0, 1]$.

The iterative scheme of the new proximal ADMM with larger step size for solving the model (1.1) is given as follows.

Algorithm 3.1 (Proximal ADMM with larger step size)**.**

Step 0. (Initialization) Select $\beta > 0, \gamma \in (0, +\infty), \varepsilon > 0$ and two positive semi-definite matrices $R_i \in \mathcal{R}^{n_i \times n_i} (i = 1, 2)$. Choose an arbitrarily initial point $w^0 = (x_1^0, x_2^0, \lambda^0) \in \mathcal{R}^{n_1+n_2+l}$. Set $k = 0$.

Step 1. (Prediction step) Compute the new iterate $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \tilde{\lambda}^k)$ via

$$\begin{cases} \tilde{x}_1^k \in \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \mathcal{L}_\beta(x_1, x_2^k, \lambda^k) + \frac{1}{2}\|x_1 - x_1^k\|_{R_1}^2, \\ \tilde{x}_2^k \in \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \mathcal{L}_\beta(\tilde{x}_1^k, x_2, \lambda^k) + \frac{1}{2}\|x_2 - x_2^k\|_{R_2}^2, \\ \tilde{\lambda}^k = \lambda^k - \gamma\beta\left(A_1\tilde{x}_1^k + A_2\tilde{x}_2^k - b\right), \end{cases} \tag{3.2}$$

Step 2. (Stopping conditions) If $\|w^k - \tilde{w}^k\| \leqslant \varepsilon$, then stop; otherwise, go to Step 3.

Step 3. (Correction step) Set $w^{k+1} = w^k + \rho(\tilde{w}^k - w^k)$, where $\rho \in (0, \eta)$, and η is a constant defined by (3.1). Go to Step 1.

Remark 3.2. To accelerate Algorithm 3.1, the Fortin and Glowinski's constant γ is attached to the updating formula for λ in the prediction step, and its feasible region is $(0, +\infty)$, which is relaxed remarkably compared with $\gamma \in \left(0, \frac{1+\sqrt{5}}{2}\right)$ in the methods of [18, 31].

Remark 3.3. From $\eta \in (0, 1]$ and $\rho \in (0, \eta)$, it is obvious that the new iterate w^{k+1} is a convex combination of w^k with $\tilde{w}^k$. As we have analyzed in Section 1, $\gamma > 1$ is usually beneficial to induce fast convergence of the ADMM, and the following numerical results indicate that $\gamma \in (1, 2)$ can accelerate the convergence of Algorithm 3.1. But the parameter η becomes small when the parameter γ becomes big, which determines the weight of the new information $\tilde{w}^k$ in the updating formula for w, and $\eta \to 0^+$ when $\gamma \to +\infty$, therefore we cannot choose large γ, though it can still ensure the convergence of Algorithm 3.1. The choice of the optimal γ is problem-dependent and maybe generally difficult to obtain in theory or practice.

Remark 3.4. In [13], He et al. proposed a fully Jacobian decomposition of the augmented Lagrangian method for multi-block separable convex programming. Subsequently, Wang et al. [27] developed a proximal partially parallel splitting method for the same problem. Both methods can also use the convex combination $w^{k+1} = w^k + \rho(\tilde{w}^k - w^k)$ to generate the new iterate. However, the constant ρ in [13] is restricted into the domain $(0, 2(1 - \sqrt{6}/3))$ and the constant ρ in [27] is restricted into the domain $(0, 2(1 - \sqrt{2}/2))$. Obviously, the feasible set of ρ in Algorithm 3.1 is larger than these in [13, 27] when $2 - \sqrt{2} < \gamma < 1/(2 - \sqrt{2})$.

If Algorithm 3.1 stops at Step 2, then the current iterate w^k is a proximal solution of $\mathrm{VI}(\mathcal{W}, F, \theta)$. Thus, it is assumed, without loss of generality, that Algorithm 3.1 generates two infinite sequences $\{w^k\}$ and $\{\tilde{w}^k\}$.

Before proving the global convergence of Algorithm 3.1, we define two matrices to simplify our notation in the later analysis.

$$M = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & \beta A_2^\top A_2 + R_2 & 0 \\ 0 & 0 & \frac{1}{\beta\gamma} I_l \end{pmatrix}, \quad Q = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & \beta A_2^\top A_2 + R_2 & \frac{1}{2\gamma} A_2^\top \\ 0 & \frac{1}{2\gamma} A_2 & \frac{1}{\beta\gamma^2} I_l \end{pmatrix}.$$

The following lemma gives some interesting properties of the two matrices M, Q just defined. These properties are crucial in the convergence analysis of Algorithm 3.1.

Lemma 3.5. *When R_1 and R_2 are two positive semi-definite matrices, we have*

(i). *Both matrices M and Q are positive semi-definite;*

(ii). *the matrix $H_1 = 2Q - \gamma M$ is positive semi-definite if $0 < \gamma \leqslant 1$, and the matrix $H_2 = 2\gamma Q - M$ is positive semi-definite if $\gamma > 1$.*

Proof.

(i). For any $w = (x_1, x_2, \lambda)$, we have

$$w^\top M w = \|x_1\|^2_{R_1} + \beta\|A_2 x_2\|^2 + \|x_2\|^2_{R_2} + \frac{1}{\beta\gamma}\|\lambda\|^2 \geqslant 0.$$

Therefore, the matrix M is positive semi-definite. The matrix Q can be partitioned into

$$Q = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta A_2^\top A_2 & \frac{1}{2\gamma}A_2^\top \\ 0 & \frac{1}{2\gamma}A_2 & \frac{1}{\beta\gamma^2}I_l \end{pmatrix}.$$

Obviously, the first part is positive semi-definite, and we only need to prove the second part is also positive semi-definite. In fact, it can be written as

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & \sqrt{\beta}A_2^\top & 0 \\ 0 & 0 & \frac{1}{\sqrt{\beta}}I_l \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_l & \frac{1}{2\gamma}I_l \\ 0 & \frac{1}{2\gamma}I_l & \frac{1}{\gamma^2}I_l \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sqrt{\beta}A_2 & 0 \\ 0 & 0 & \frac{1}{\sqrt{\beta}}I_l \end{pmatrix}.$$

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

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & \frac{1}{2\gamma} \\ 0 & \frac{1}{2\gamma} & \frac{1}{\gamma^2} \end{pmatrix} \otimes I_l,$$

where $\otimes$ denotes the matrix Kronecker product. The matrix Kronecker product has a nice property: for any two matrices X and Y, the eigenvalue of $X \otimes Y$ equals the product of $\lambda(X)\lambda(Y)$, where $\lambda(X)$ and $\lambda(Y)$ are the eigenvalue of X and Y, respectively. Therefore, we only need to show the 2-by-2 matrix

$$\begin{pmatrix} 1 & \frac{1}{2\gamma} \\ \frac{1}{2\gamma} & \frac{1}{\gamma^2} \end{pmatrix}$$

is positive semi-definite. In fact,

$$1 \times \frac{1}{\gamma^2} - \frac{1}{2\gamma} \times \frac{1}{2\gamma} = \frac{3}{4\gamma^2} > 0.$$

Therefore, the matrix Q is positive semi-definite.

(ii). If $0 < \gamma \leqslant 1$, we have

$$\begin{aligned} H_1 = 2Q - \gamma M &= \begin{pmatrix} (2-\gamma)R_1 & 0 & 0 \\ 0 & (2-\gamma)(\beta A_2^\top A_2 + R_2) & \frac{1}{\gamma}A_2^\top \\ 0 & \frac{1}{\gamma}A_2 & \frac{2-\gamma^2}{\beta\gamma^2}I_l \end{pmatrix} \\ &= (2-\gamma)\begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & (2-\gamma)\beta A_2^\top A_2 & \frac{1}{\gamma}A_2^\top \\ 0 & \frac{1}{\gamma}A_2 & \frac{2-\gamma^2}{\beta\gamma^2}I_l \end{pmatrix}. \end{aligned}$$

Obviously, the first part is positive semi-definite, and we only need to show the second part is also positive semi-definite. In fact, it can be written as

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & \sqrt{\beta}A_2^\top & 0 \\ 0 & 0 & \frac{1}{\sqrt{\beta}}I_l \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & (2-\gamma)I_l & \frac{1}{\gamma}I_l \\ 0 & \frac{1}{\gamma}I_l & \frac{2-\gamma^2}{\gamma^2}I_l \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sqrt{\beta}A_2 & 0 \\ 0 & 0 & \frac{1}{\sqrt{\beta}}I_l \end{pmatrix},$$

and similar to the proof of (i), we only need to show

$$(2-\gamma)\times\frac{2-\gamma^2}{\gamma^2}-\frac{1}{\gamma}\times\frac{1}{\gamma}=\frac{(1-\gamma)(3+\gamma-\gamma^2)}{\gamma^2}\geqslant 0.$$

Therefore, the matrix H_1 is positive semi-definite if $0<\gamma\leqslant 1$.

Now, we show that the matrix H_2 is positive semi-definite if $\gamma>1$.

$$\begin{aligned}H_2=2\gamma Q-M&=\begin{pmatrix}(2\gamma-1)R_1 & 0 & 0\\ 0 & (2\gamma-1)(\beta A_2^\top A_2+R_2) & A_2^\top\\ 0 & A_2 & \frac{1}{\beta\gamma}I_l\end{pmatrix}\\&=(2\gamma-1)\begin{pmatrix}R_1&0&0\\0&R_2&0\\0&0&0\end{pmatrix}+\begin{pmatrix}0&0&0\\0&(2\gamma-1)\beta A_2^\top A_2&A_2^\top\\0&A_2&\frac{1}{\beta\gamma}I_l\end{pmatrix}.\end{aligned}$$

Similarly, we only need to show the second part is positive semi-definite. In fact, it can be written as

$$\begin{pmatrix}0&0&0\\0&\sqrt{\beta}A_2^\top&0\\0&0&\frac{1}{\sqrt{\beta}}I_l\end{pmatrix}\begin{pmatrix}0&0&0\\0&(2\gamma-1)I_l&I_l\\0&I_l&\frac{1}{\gamma}I_l\end{pmatrix}\begin{pmatrix}0&0&0\\0&\sqrt{\beta}A_2&0\\0&0&\frac{1}{\sqrt{\beta}}I_l\end{pmatrix},$$

and

$$(2\gamma-1)\times\frac{1}{\gamma}-1\times 1=\frac{\gamma-1}{\gamma}>0.$$

Therefore, the matrix H_2 is positive semi-definite if $\gamma>1$. The proof is complete. □

Now, we shall show that it is reasonable to use $\|w^k-\tilde{w}^k\|\leqslant\varepsilon$ to terminate Algorithm 3.1, in which $\varepsilon>0$ is a chosen fixed threshold specified by the user.

Lemma 3.6. *If $A_ix_i^k=A_i\tilde{x}_i^k$ $(i=1,2)$, $\lambda^k=\tilde{\lambda}^k$, then the vector (x_1^k,x_2^k,λ^k) is a solution of VI$(\mathcal{W},F,\theta)$.*

Proof. By the first-order optimality condition for x_1-subproblem in (3.2), one has

$$\theta_1(x_1)-\theta_1(\tilde{x}_1^k)+(x_1-\tilde{x}_1^k)^\top\left\{-A_1^\top\lambda^k+\beta A_1^\top\left(A_1\tilde{x}_1^k+A_2x_2^k-b\right)+R_1(\tilde{x}_1^k-x_1^k)\right\}\geqslant 0,\quad\forall x_1\in\mathcal{X}_1.$$

Substituting $\lambda^k=\tilde{\lambda}^k+\gamma\beta\left(\sum_{i=1}^{2}A_i\tilde{x}_i^k-b\right)$ into the above inequality, we have

$$\begin{aligned}\theta_1(x_1)-\theta_1(\tilde{x}_1^k)+(x_1-\tilde{x}_1^k)^\top\{&-A_1^\top\tilde{\lambda}^k+\beta(1-\gamma)A_1^\top\left(A_1\tilde{x}_1^k+A_2\tilde{x}_2^k-b\right)\\&+\beta A_1^\top A_2(x_2^k-\tilde{x}_2^k)+R_1(\tilde{x}_1^k-x_1^k)\}\geqslant 0,\quad\forall x_1\in\mathcal{X}_1,\end{aligned}$$

i.e.,

$$\begin{aligned}&\theta_1(x_1)-\theta_1(\tilde{x}_1^k)+(x_1-\tilde{x}_1^k)^\top\left\{-A_1^\top\tilde{\lambda}^k+\frac{1-\gamma}{\gamma}A_1^\top(\lambda^k-\tilde{\lambda}^k)+\beta A_1^\top A_2(x_2^k-\tilde{x}_2^k)\right\}\\&\geqslant(x_1-\tilde{x}_1^k)^\top R_1(x_1^k-\tilde{x}_1^k),\quad\forall x_1\in\mathcal{X}_1.\end{aligned}\tag{3.3}$$

Similarly, for the other variable x_2, by the first-order optimality condition for x_2-subproblem in (3.2), one has

$$\theta_2(x_2)-\theta_2(\tilde{x}_2^k)+(x_2-\tilde{x}_2^k)^\top\left\{-A_2^\top\lambda^k+\beta A_2^\top\left(A_1\tilde{x}_1^k+A_2\tilde{x}_2^k-b\right)+R_2(\tilde{x}_2^k-x_2^k)\right\}\geqslant 0,\quad\forall x_2\in\mathcal{X}_2,$$

and substituting $\lambda^k = \tilde{\lambda}^k + \gamma\beta\left(\sum_{i=1}^{2} A_i\tilde{x}_i^k - b\right)$ into the above inequality gives us

$$\theta_2(x_2) - \theta_2(\tilde{x}_2^k) + (x_2 - \tilde{x}_2^k)^\top \left\{ -A_2^\top\tilde{\lambda}^k + \beta(1-\gamma)A_2^\top\left(A_1\tilde{x}_1^k + A_2\tilde{x}_2^k - b\right) + R_2(\tilde{x}_2^k - x_2^k)\right\} \geqslant 0, \quad \forall x_2 \in \mathcal{X}_2,$$

i.e.,

$$\theta_2(x_2) - \theta_2(\tilde{x}_2^k) + (x_2 - \tilde{x}_2^k)^\top \left\{-A_2^\top\tilde{\lambda}^k + \frac{1-\gamma}{\gamma}A_2^\top(\lambda^k - \tilde{\lambda}^k)\right\} \geqslant (x_2 - \tilde{x}_2^k)^\top R_2(x_2^k - \tilde{x}_2^k), \quad \forall x_2 \in \mathcal{X}_2. \tag{3.4}$$

Furthermore, the updating formula for λ in (3.2) can be written as the following variational inequality problem

$$(\lambda - \tilde{\lambda}^k)^\top \left\{ \left(\sum_{i=1}^{2} A_i\tilde{x}_i^k - b\right) + \frac{1}{\beta\gamma}(\tilde{\lambda}^k - \lambda^k)\right\} \geqslant 0, \forall \lambda \in \mathcal{R}^l. \tag{3.5}$$

Then, adding (3.3), (3.4), and (3.5), we get

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geqslant (w - \tilde{w}^k)^\top G(w^k - \tilde{w}^k), \quad \forall w \in \mathcal{W}, \tag{3.6}$$

where

$$G = \begin{pmatrix} R_1 & -\beta A_1^\top A_2 & -\frac{1-\gamma}{\gamma}A_1^\top \\ 0 & R_2 & -\frac{1-\gamma}{\gamma}A_2^\top \\ 0 & 0 & \frac{1}{\beta\gamma}I_l \end{pmatrix}.$$

Therefore, substituting $A_i x_i^k = A_i\tilde{x}_i^k (i = 1, 2), \lambda^k = \tilde{\lambda}^k$ into (3.5), we get

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geqslant 0, \quad \forall w \in \mathcal{W}.$$

This together with Assumption 2.4 implies that

$$\theta(x) - \theta(x^k) + (w - w^k)^\top F(w^k) \geqslant 0, \quad \forall w \in \mathcal{W},$$

which indicates that $(x_1^k, x_2^k, \lambda^k)$ is a solution of $VI(\mathcal{W}, F, \theta)$. This completes the proof. □

Now, we intend to prove that $-(w^k - \tilde{w}^k)$ is a descent direction of the merit function $\frac{1}{2}\|w - w^*\|_M^2$ at the point $w = w^k$, where $w^* \in \mathcal{W}^*$.

Lemma 3.7. *Let $\{w^k\}$ and $\{\tilde{w}^k\}$ be two sequences generated by Algorithm 3.1. Then, we have*

$$(w^k - w^*)^\top M(w^k - \tilde{w}^k) \geqslant \|w^k - \tilde{w}^k\|_Q^2, \quad \forall w^* \in \mathcal{W}^*. \tag{3.7}$$

Proof. Since $w^* \in \mathcal{W}^* \subseteq \mathcal{W}$, it follows from (3.6) that

$$\theta(x^*) - \theta(\tilde{x}^k) + (w^* - \tilde{w}^k)^\top F(\tilde{w}^k) \geqslant (w^* - \tilde{w}^k)^\top G(w^k - \tilde{w}^k).$$

Thus,

$$(\tilde{w}^k - w^*)^\top G(w^k - \tilde{w}^k) \geqslant \theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^\top F(\tilde{w}^k) \geqslant 0, \tag{3.8}$$

where the second inequality follows from $w^* \in \mathcal{W}^*$.

On the other hand, from the definitions of M and G, we have

$$\begin{aligned}(\tilde{w}^k - w^*)^\top G(w^k - \tilde{w}^k) &= (\tilde{w}^k - w^*)^\top M(w^k - \tilde{w}^k) + (\tilde{w}^k - w^*)^\top \begin{pmatrix} 0 & -\beta A_1^\top A_2 & -\frac{1-\gamma}{\gamma}A_1^\top \\ 0 & -\beta A_2^\top A_2 & -\frac{1-\gamma}{\gamma}A_2^\top \\ 0 & 0 & 0 \end{pmatrix}(w^k - \tilde{w}^k) \\ &= (\tilde{w}^k - w^*)^\top M(w^k - \tilde{w}^k) - \beta(A_1\tilde{x}_1 + A_2\tilde{x}_2 - b)^\top A_2(x_2^k - \tilde{x}_2^k)\end{aligned}$$

$$
\begin{aligned}
&-\frac{1-\gamma}{\gamma}(A_1\tilde{x}_1+A_2\tilde{x}_2-b)^\top(\lambda^k-\tilde{\lambda}^k)\\
&=(\tilde{w}^k-w^*)^\top M(w^k-\tilde{w}^k)-\frac{1}{\gamma}(\lambda^k-\tilde{\lambda}^k)^\top A_2(x_2^k-\tilde{x}_2^k)-\frac{1-\gamma}{\beta\gamma^2}\|\lambda^k-\tilde{\lambda}^k\|^2,
\end{aligned}
$$

where the second inequality follows from $w^* \in \mathcal{W}^*$, and the second equality comes from (3.2).

Then, substituting the above relationship into the left-hand side of (3.8), one has

$$
(\tilde{w}^k-w^*)^\top M(w^k-\tilde{w}^k)-\frac{1}{\gamma}(\lambda^k-\tilde{\lambda}^k)^\top A_2(x_2^k-\tilde{x}_2^k)-\frac{1-\gamma}{\beta\gamma^2}\|\lambda^k-\tilde{\lambda}^k\|^2\geqslant 0,
$$

i.e.,

$$
\begin{aligned}
(w^k-w^*)^\top M(w^k-\tilde{w}^k)&\geqslant (w^k-\tilde{w}^k)^\top M(w^k-\tilde{w}^k)+\frac{1}{\gamma}(\lambda^k-\tilde{\lambda}^k)^\top A_2(x_2^k-\tilde{x}_2^k)+\frac{1-\gamma}{\beta\gamma^2}\|\lambda^k-\tilde{\lambda}^k\|^2\\
&=\|w^k-\tilde{w}^k\|_Q^2,
\end{aligned}
$$

where the equality follows from definition (3.1). The lemma is proved. □

The inequality (3.7) and Lemma 3.5 imply that $-(w^k-\tilde{w}^k)$ is a descent direction of the merit function $\frac{1}{2}\|w-w^*\|_M^2$ at $w=w^k$. Therefore, it is reasonable to get the new iterate by $w^{k+1}=w^k+\rho(\tilde{w}^k-w^k)$. The following theorem shows that the sequence $\{w^k\}$ generated by Algorithm 3.1 is Fejér monotone with respect to the solution set $\mathcal{W}^*$ of $\mathrm{VI}(\mathcal{W},F,\theta)$.

Theorem 3.8. *For any $w^*=(x_1^*,x_2^*,\lambda^*)\in\mathcal{W}^*$, there exists a positive scalar η such that the sequence $\{w^k\}$ generated by Algorithm 3.1 satisfies*

$$
\|w^{k+1}-w^*\|_M^2\leqslant\|w^k-w^*\|_M^2-\rho\|w^k-\tilde{w}^k\|_H^2, \tag{3.9}
$$

where the matrix H is defined by

$$
H=(\eta-\rho)M. \tag{3.10}
$$

Proof. If $0<\gamma\leqslant 1$, it follows from the correction step of Algorithm 3.1 that

$$
\begin{aligned}
\|w^{k+1}-w^*\|_M^2&=\|w^k+\rho(\tilde{w}^k-w^k)-w^*\|_M^2\\
&=\|w^k-w^*\|_M^2+2\rho(w^k-w^*)^\top(\tilde{w}^k-w^k)+\rho^2\|\tilde{w}^k-w^k\|_M^2\\
&\leqslant\|w^k-w^*\|_M^2-2\rho\|\tilde{w}^k-w^k\|_Q^2+\rho^2\|\tilde{w}^k-w^k\|_M^2\\
&=\|w^k-w^*\|_M^2-2\rho\|\tilde{w}^k-w^k\|^2_{\frac{H_1}{2}+\frac{\gamma}{2}M}+\rho^2\|\tilde{w}^k-w^k\|_M^2\\
&\leqslant\|w^k-w^*\|_M^2-\rho\|\tilde{w}^k-w^k\|^2_{(\gamma-\rho)M},
\end{aligned}
$$

where the first inequality follows from (3.7), and the second inequality comes from Lemma 3.5.

If $\gamma\geqslant 1$, we similarly have

$$
\begin{aligned}
\|w^{k+1}-w^*\|_M^2&\leqslant\|w^k-w^*\|_M^2-2\rho\|\tilde{w}^k-w^k\|^2_{\frac{H_2}{2\gamma}+\frac{M}{2\gamma}}+\rho^2\|\tilde{w}^k-w^k\|_M^2\\
&\leqslant\|w^k-w^*\|_M^2-\rho\|\tilde{w}^k-w^k\|^2_{(\frac{1}{\gamma}-\rho)M},
\end{aligned}
$$

where the second inequality also comes from Lemma 3.5. The assertion (3.9) is proved. □

Remark 3.9. By the definition of η in (3.1), the matrix H in (3.10) is positive semi-definite.

Based on the Fejér monotonicity of the sequence $\{w^k\}$, the global convergence of Algorithm 3.1 can be proved as follows, which is standard, and we include it for completeness. For convenience, we set $v=(x_2,\lambda)$, and the low right sub-blocks of the matrices M, H are denoted by $\tilde{M}, \tilde{H}$, respectively. Obviously, by Assumption 2.4, both matrices $\tilde{M}$ and $\tilde{H}$ are positive definite.

Theorem 3.10. *The sequence $\{\tilde{w}^k\}$ generated by Algorithm 3.1 converges to some $w^\infty \in \mathcal{W}^*$.*

Proof. Using (3.9) and noting $\rho > 0$, we have

$$\|v^{k+1} - v^*\|_{\tilde{M}}^2 \leqslant \|w^{k+1} - w^*\|_M^2 \leqslant \|w^k - w^*\|_M^2 \leqslant \cdots \leqslant \|w^0 - w^*\|_M^2.$$

Then $\{v^k\}$ is a bounded sequence by the positive definiteness of $\tilde{M}$. By (3.9) again, we have

$$(\eta - \rho)\sum_{k=0}^{\infty}\|x_1^k - \tilde{x}_1^k\|_{R_1}^2 + \sum_{k=0}^{\infty}\|v^k - \tilde{v}^k\|_{\tilde{H}}^2 = \sum_{k=0}^{\infty}\|w^k - \tilde{w}^k\|_H^2 \leqslant \frac{1}{\rho}\|w^0 - w^*\|_M^2,$$

which together with the positive definiteness of $\tilde{H}$ yields

$$\lim_{k\to\infty}\|x_1^k - \tilde{x}_1^k\|_{R_1} = \lim_{k\to\infty}\|v^k - \tilde{v}^k\| = 0. \tag{3.11}$$

Thus $\{\tilde{v}^k\}$ is also a bounded sequence. From

$$\|A_1\tilde{x}_1^k\| = \|\frac{1}{\gamma\beta}(\lambda^k - \tilde{\lambda}^k) - A_2\tilde{x}_2^k + b\| \leqslant \frac{1}{\gamma\beta}(\|\lambda^k\| + \|\tilde{\lambda}^k\|) + \|A_2\|\|\tilde{x}_2^k\| + \|b\|,$$

and Assumption 2.4, we have that the sequence $\{\tilde{x}_1^k\}$ is bounded. Then the sequence $\{\tilde{w}^k\}$ is also bounded, thus it has at least a cluster point, saying $(x_1^\infty, x_2^\infty, \lambda^\infty)$, and suppose the subsequence $\{(\tilde{x}_1^{k_i}, \tilde{x}_2^{k_i}, \tilde{\lambda}^{k_i})\}$ converges to $(x_1^\infty, x_2^\infty, \lambda^\infty)$. Taking limits on both sides of $A_1\tilde{x}_1^k + A_2\tilde{x}_2^k - b = \frac{1}{\gamma\beta}(\lambda^k - \tilde{\lambda}^k)$, we have

$$A_1x_1^\infty + A_2x_2^\infty - b = 0.$$

Furthermore, taking limits on both sides of (3.3) and (3.4), and using (3.11), we obtain

$$\theta_1(x_1) - \theta_1(x_1^\infty) + (x_1 - x_1^\infty)^\top(-A_1^\top\lambda^\infty) \geqslant 0, \quad \forall x_1 \in \mathcal{X}_1,$$

and

$$\theta_2(x_2) - \theta_2(x_2^\infty) + (x_2 - x_2^\infty)^\top(-A_2^\top\lambda^\infty) \geqslant 0, \quad \forall x_2 \in \mathcal{X}_2.$$

Therefore, $(x_1^\infty, x_2^\infty, \lambda^\infty) \in \mathcal{W}^*$.

Then, from $\lim_{k\to\infty}(v^k - \tilde{v}^k) = 0$ and $\{\tilde{v}^{k_j}\} \to v^\infty$, for any given $\epsilon > 0$, there exists an integer l, such that

$$\|v^{k_l} - \tilde{v}^{k_l}\|_{\tilde{M}} < \frac{\epsilon}{2}, \quad \text{and} \quad \|\tilde{v}^{k_l} - \hat{v}\|_{\tilde{M}} < \frac{\epsilon}{2}.$$

Therefore, for any $k \geqslant k_l$, it follows from the above two inequalities and (3.9) that

$$\|v^k - v^\infty\|_{\tilde{M}} \leqslant \|v^{k_l} - v^\infty\|_{\tilde{M}} \leqslant \|v^{k_l} - \tilde{v}^{k_l}\|_{\tilde{M}} + \|\tilde{v}^{k_l} - v^\infty\|_{\tilde{M}} < \epsilon,$$

which indicates that the sequence $\{v^k\}$ converges globally to the point v^∞, and thus the sequence $\{\tilde{v}^k\}$ also converges globally to the point v^∞. Then, from $A_1\tilde{x}_1^k = \frac{1}{\gamma\beta}(\lambda^k - \tilde{\lambda}^k) - A_2\tilde{x}_2^k + b$ and Assumption 2.4 again, it can be inferred that the sequence $\{\tilde{x}_1^k\}$ converges to x_1^∞. Overall, the sequence $\{\tilde{w}^k\}$ generated by Algorithm 3.1 converges to some $w^\infty \in \mathcal{W}^*$. The proof is completed. □

4. Numerical results

In this section, we conduct some numerical experiments about the correlation matrices calibrating problems to verify the efficiency of Algorithm 3.1, and compare the performance with some existing iteration methods including the inexact alternating directions method (IADM) developed by He et al. [14], the full Jacobian decomposition of the augmented Lagrangian method (JDALM) proposed by He et al. [13], and the customized proximal point algorithm (CPPA) introduced by He et al. [19]. All codes were

written by Matlab R2010a and performed on a ThinkPad computer equipped with Windows XP, 997MHz and 2 GB of memory.

The least-squares problem is to find a matrix $X \in \mathcal{R}^{n\times n}$ such that

$$\min\left\{\frac{1}{2}\|X-C\|_F^2 | X \in \mathcal{S}_+^n \cap \mathcal{S}_{\mathcal{B}}\right\}, \tag{4.1}$$

where

$$\mathcal{S}_+^n = \{H \in \mathcal{R}^{n\times n} | H^\top = H, H \succeq 0\},$$

and

$$\mathcal{S}_{\mathcal{B}} = \{H \in \mathcal{R}^{n\times n} | H^\top = H, H_L \leqslant H \leqslant H_U\}.$$

(4.1) is quite similar to the problem of computing the nearest correlation matrix, which appears in then financial mathematics where the correlations are between stocks. Introducing an auxiliary variable Y, (4.1) can be written as

$$\begin{aligned} &\min\left\{\tfrac{1}{2}\|X-C\|_F^2 + \tfrac{1}{2}\|Y-C\|_F^2\right\} \\ &\text{s.t. } X - Y = 0, \\ &\quad X \in \mathcal{S}_+^n, Y \in \mathcal{S}_{\mathcal{B}}. \end{aligned} \tag{4.2}$$

Then, Algorithm 3.1 is applicable to (4.2). By some simple manipulations, we have

$$\tilde{X}^k = P_{S_+^n}\left\{\frac{1}{1+\beta}(\beta Y^k + \lambda^k + C)\right\},$$

where $P_{S_+^n}(A) = U\Lambda^+U^\top, [U,\Lambda] = \mathtt{eig}(A), \Lambda^+ = \max\{\Lambda, 0\}$, and the computational load of the above procedure is $10n^3$ flops. $\tilde{Y}^k$ is generated by

$$\tilde{Y}^k = P_{S_{\mathcal{B}}}\left\{\frac{1}{1+\beta}(\beta \tilde{X}^k - \lambda^k + C)\right\},$$

where $P_{S_{\mathcal{B}}}(A) = \min\{\max\{H_L, A\}, H_U\}$ and its computational load is bounded above by $2n^2$ flops. Obviously, the main computational cost in Algorithm 3.1 for (4.1) is the eigenvalue decomposition, which is accomplished by the built-in function `eig` of the software Matlab. When n is large, the computation of `eig` is quite time-consuming. Due to the limitations of the memory and speed in a regular computer, we only solve some medium scale cases of (4.1).

We use the following Matlab scripts to generate the problem data

```
rand('state',0); C=rand(n,n); C=(C'+C)-ones(n,n)+eye(n);

HU=ones(n)*0.1; HL=-HU; for  i=1:n  HU(i,i)=1; HL(i,i)=1;  end.
```

All the initial matrices are set as zero. The stopping criterion is

$$\mathrm{Err} = \max\left\{\frac{\|Y_k - Y_0\|}{\|Y_0\|}, \frac{\|\lambda_k - \lambda_0\|}{\|\lambda_0\|}\right\} < 10^{-6}.$$

First, by solving a small scale case of (4.1) with $n = 100$, we show that the relaxation parameter $\gamma > 1$ maybe accelerate the convergence of Algorithm 3.1. We choose different values of γ in the interval [1, 2]. Specifically, we choose $\gamma \in \{1.0, 1.1, \ldots, 2\}$. The numerical results are graphically shown in Figure 1.

The two subplots in Figure 1 indicate that both the number of iterations (Iter.) and the CPU time (Time) decrease obviously as the relaxation parameter γ increases. Specifically, the minimum of Iter. is obtained at $\gamma = 1.8$, and the minimum of Time is gotten at $\gamma = 1.8$. Then, in the following, we set $\gamma = 1.8$.

We now do some numerical comparisons to illustrate the advantage of Algorithm 3.1, and the numerical results are listed in Table 1.

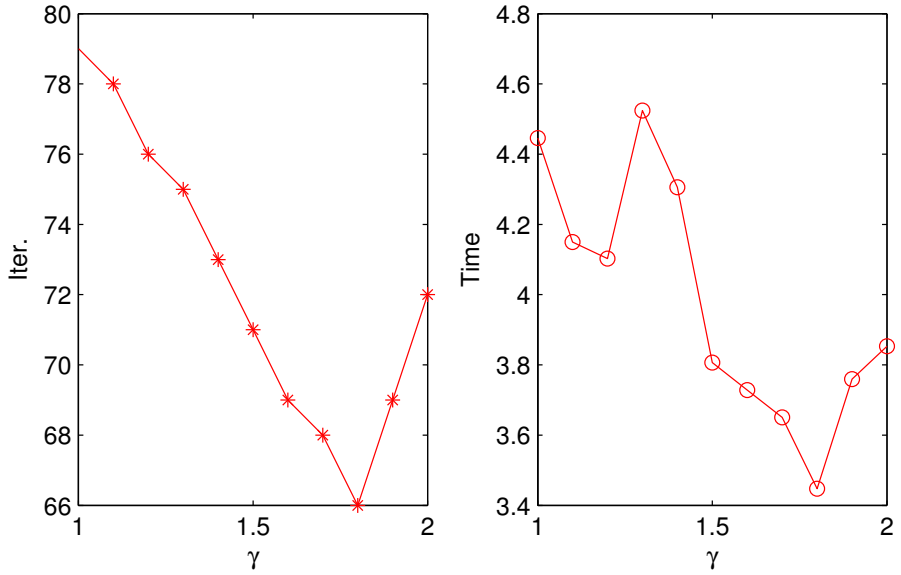

Figure 1: Sensitivity test on the parameter γ with $n = 100, \beta = 3.5$.

Table 1: Numerical results for the problem (4.1).

n	β	Method	Iter.	Time	Err
100	3.5	IADM	79	4.3992	9.1033e-007
		JDALM	184	10.1089	9.8351e-007
		CPPA	68	4.1184	8.7915e-007
		Algorithm 3.1	66	3.6192	9.4386e-007
200	6	IADM	62	17.7061	8.3922e-007
		JDALM	197	58.6876	9.5547e-007
		CPPA	68	21.9805	8.0165e-007
		Algorithm 3.1	53	16.4893	8.3736e-007
300	6	IADM	73	58.8904	8.6578e-007
		JDALM	195	150.4318	9.9112e-007
		CPPA	66	65.7388	9.5922e-007
		Algorithm 3.1	53	41.4027	8.6329e-007
400	6	IADM	78	163.1146	9.1689e-007
		JDALM	197	429.1744	9.9938e-007
		CPPA	66	165.3299	8.0608e-007
		Algorithm 3.1	53	106.2211	9.0843e-007
500	6	IADM	81	318.8036	9.2148e-007
		JDALM	201	750.7704	9.4523e-007
		CPPA	65	287.2914	9.4195e-007
		Algorithm 3.1	53	204.2677	9.0167e-007

The numerical results in Table 1 indicate that the tested four iteration methods are applicable to solve (4.1), and Algorithm 3.1 significantly decreases both the number of iterations and CPU time comparing with the other three iteration methods. Figure 2 displays the evolution of Err with respect to the iteration counter k, from which we can see that the IADM performs the best firstly. When $k \geqslant 40$, Algorithm 3.1 outperforms the IADM, and when $k \geqslant 57$, the CPPA also outperforms the IADM, but still performs a little worse than Algorithm 3.1.

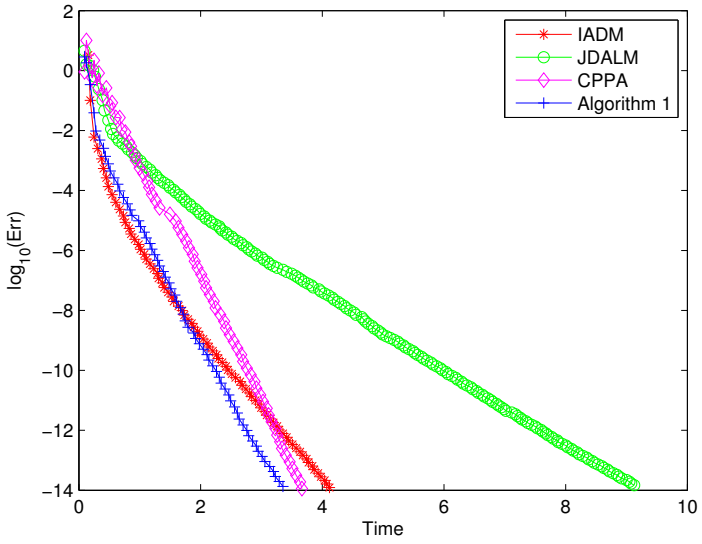

Figure 2: The relative error with $n = 100, \beta = 3.5$.

5. Conclusions

In this paper, a proximal ADMM with larger step size for two-block separable convex programming is proposed. Under mild conditions, we have established its global convergence. Some numerical results are given, which illustrate that the new method performs better than some state-of-the-art solvers. In the numerical experiments, we find that the proposed method is sensitive to the relaxation factor γ under certain conditions, thus the choice of this factor needs to be further studied.

Acknowledgment

The authors gratefully acknowledge the valuable comments of the anonymous reviewers. This work is supported by the National Natural Science Foundation of China (No. 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) and Scientific Research Project of Shandong Universities (No. J15LI11).

References

[1] J. M. Bardsley, S. Knepper, J. Nagy, *Structured linear algebra problems in adaptive optics imaging*, Adv. Comput. Math., **35** (2011), 103–117. 1
[2] H. H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, (2011). 2.2
[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning, **3** (2011), 1–122. 1
[4] H. B. Chen, Y. J. Wang, H. g. Zhao, *Finite convergence of a projected proximal point algorithm for generalized variational inequalities*, Oper. Res. Lett., **40** (2012), 303–305.
[5] E. Esser, *Applications of Lagrangian-based alternating direction methods and connections to split Bregman*, CAM report, (2009). 1
[6] M. Fortin, R. Glowinski, *Augmented Lagrangian Methods: Applications to the Numerical Solutions of Boundary Value Problems*, Studies in Applied Mathematics, Amsterdam, (1983). 1
[7] M. Fukushima, *Fundamentals of Nonlinear Optimization*, Asakura Shoten, Tokyo, (2001). 2
[8] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer, New York, (1984). 1

[9] R. Glowinski, A. Marrocco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non-linéaires*, Rev. Francaise Autom. Inf. Recherche Opérationnelle Anal. Numér, **9** (1975), 41–76. 1

[10] H.-J. He, X.-J. Cai, D. Han, *A fast splitting method tailored for Dantzig selectors*, Comput. Optim. Appl., **62** (2015), 347–372. 1, 1

[11] H.-J. He, D. Han, *A distributed Douglas-Rachford splitting method for multi-block convex minimization problems*, Adv. Comput. Math., **42** (2016), 27–53. 1, 1

[12] D. R. Han, H. J. He, H. Yang, X. M. Yuan, *A customized Douglas-Rachford splitting algorithm for separable convex minimization with linear constraints*, Numer. Math., **127** (2014), 167–200. 1, 1

[13] B.-S. He, L.-S. Hou, X.-M. Yuan, *On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming*, SIAM J. Optim., **25** (2015), 2274–2312. 1, 1, 3.4, 4

[14] B.-S. He, L.-Z. Liao, D. R. Han, H. Yang, *A new inexact alternating directions method for monotone variational inequalities*, Math. Program., **92** (2002), 103–118. 1, 1, 4

[15] H.-J. He, C. Ling, H.-K. Xu, *An implementable splitting algorithm for the ℓ_1-norm regularized split feasibility problem*, J. Sci. Comput., **67** (2016), 281–298. 1, 1

[16] B.-S. He, F. Ma, *Convergence study on the proximal alternating direction method with larger step size*, manuscript, (2017). 1

[17] B.-S. He, M. Tao, X.-M. Yuan , *A splitting method for separable convex programming*, IMA J. Numer. Anal., **35** (2015), 394–426. 1

[18] B.-S. He, S.-L. Wang, H. Yang, *A modified variable-penalty alternating direction method for monotone variational inequalities*, J. Comput. Math., **21** (2003), 495–504. 1, 1, 3.2

[19] B.-S. He, X.-M. Yuan, W.-X. Zhang, *A customized proximal point algorithm for convex minimization with linear constraints*, Comput. Optim. Appl., **56** (2013), 559–572. 1, 4

[20] M. Li, *A hybrid LQP-based method for structured variational inequalities*, Int. J. Comput. Math., **89** (2012), 1412–1425. 1

[21] J. Liu, Y.-R. Duan, M. Sun, *A symmetric version of the generalized alternating direction method of multipliers for two-block separable convex programming*, J. Inequal. Appl., **2017** (2017), 21 pages. 1

[22] H. Qiu, X. Chen, W. Liu, G. Zhou, Y.-J. Wang, J.-H. Lai, *A fast ℓ_1-solver and its applications to robust face recognition*, J. Ind. Manag. Optim., **8** (2012), 163–178. 1

[23] R. T. Rockafellar, *Convex Analysis*, University Press, Princeton, (1970). 2.1

[24] M. Sun, J. Liu, *An inexact generalized PRSM with LQP regularization for structured variational inequalities and its applications traffic equilibrium problems*, J. Inequal. Appl., **2016** (2016), 17 pages. 1

[25] M. Sun, J. Liu, *Generalized Peaceman-Rachford splitting method for separable convex programming with applications to image processing*, J. Appl. Math. Comput., **51** (2016), 605–622. 1

[26] M. Sun, Y.-J. Wang, J. Liu, *Generalized Peaceman-Rachford splitting method for multiple-block separable convex programming with applications to robust PCA*, Calcolo, **54** (2017), 77–94. 1

[27] K. Wang, J. Desai, H.-J. He, *A proximal partially parallel splitting method for separable convex programs*, Optim. Methods Softw., **32** (2017), 39–68. 1, 3.4

[28] Y. Wang, W. Liu, L. Caccetta, G. Zhou, *Parameter selection for nonnegative ℓ_1 matrix/tensor sparse decomposition*, Operations Research Lett., **43** (2015), 423–426.

[29] Y. Wang, G. Zhou, L. Caccetta, W. Liu, *An alternative Lagrange-dual based algorithm for sparse signal reconstruction*, IEEE Transactions on Signal Processing, **59** (2011), 1895–1901. 1, 1

[30] Y.-Y. Xu, *Accelerated first-order primal-dual proximal methods for linear constrained composite convex programming*, SIAM Journal on Optimization, **27** (2017), 1459–1484. 1

[31] M. H. Xu, T. Wu, *A class of linearized proximal alternating direction methods*, J. Optim. Theory Appl., **151** (2011), 321–337. 1, 1, 1, 3.2

[32] J.-F. Yang, Y. Zhang, *Alternating direction algorithms for ℓ_1-problems in compressive sensing*, SIAM J. Sci. Comput., **332** (2011), 250–278. 1

[33] C.-H. Ye, X.-M. Yuan, *A descent method for structured monotone variational inequalities*, Optim. Methods Softw., **22** (2007), 329–338. 1

[34] X.-M. Yuan, *An improved proximal alternating direction method for monoton variational inequalities with separable structure*, Computational Optimization and Applications, **49** (2011), 17–29. 1, 1