

# Proximal ADMM with larger step size for two-block separable convex programs

Hongchun Sun<sup>1</sup>, Min Sun<sup>2,3</sup>, Yiju Wang<sup>3</sup>

1. School of Sciences, Linyi University,

Shandong, 276005, P. R. China.

2. School of Mathematics and Statistics, Zaozhuang University,

Shandong, 277160, P. R. China.

3. School of Management, Qufu Normal University,

Shandong, 276826, P. R. China

**Abstract.** The alternating direction method of multipliers (ADMM) is a benchmark for solving two-block separable convex programs, and it finds more and more applications in many areas. However, as other first-order methods, ADMM also suffers from low convergence. In this paper, to accelerate the convergence of ADMM, we relax the restriction region of the Fortin and Glowinski's constant  $\gamma$  in ADMM from  $(0, \frac{1+\sqrt{5}}{2})$  to  $(0, +\infty)$ , and thus get a proximal ADMM with larger step size. Then, to further accelerate its efficiency, we use the convex combination of its output with the current iterate to generate the new iterate. Under mild conditions, we prove the global convergence of the new method. Some numerical experiments are given to demonstrate the advantage of large step size.

**Keywords:** alternating direction method of multipliers; the Fortin and Glowinski's constant; global convergence.

**AMS subject classifications.** 90C25, 94A08

## 1 Introduction

In this paper, we consider the following two-block separable convex programs:

$$\min\{\theta_1(x_1) + \theta_2(x_2) \mid A_1x_1 + A_2x_2 = b, x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2\}, \quad (1)$$

where  $\theta_i : \mathcal{R}^{n_i} \rightarrow \mathcal{R} (i = 1, 2)$  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. As a special case of convex optimization, (1) is the mathematical model of many practical problems encountered in diverse fields, such as statistical learning [1], image/signal processing [2, 3], traffic equilibrium [4, 5], and so on.

Let  $\beta \in (0, +\infty)$  be a given parameter. The augmented Lagrangian function for the model (1) is

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

where  $\lambda \in \mathcal{R}^l$  is the Lagrangian multiplier with the linear constraint in (1). The classical alternating direction method of multipliers (ADMM) for solving model (1) consists of the following iterations for  $k = 0, 1, \dots$

$$\begin{cases} x_1^{k+1} \in \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \mathcal{L}_\beta(x_1, x_2^k, \lambda^k), \\ x_2^{k+1} \in \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^k), \\ \lambda^{k+1} = \lambda^k - \gamma\beta (A_1x_1^{k+1} + A_2x_2^{k+1} - b), \end{cases} \quad (2)$$

where  $\gamma \in (0, \frac{1+\sqrt{5}}{2})$  is the Fortin and Glowinski's constant [6] and it is also the step size for the update of  $\lambda$ . Obviously, the efficiency of ADMM depends on how to solve its two subproblems related to the primal variables  $x_1$  and  $x_2$ . Therefore, some proximal ADMMs are proposed recently [7, 8, 9]. Such as, in [7], Xu and Wu add some proximal terms to the subproblems of (2) and get the following iterative scheme:

$$\begin{cases} x_1^{k+1} \in \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \mathcal{L}_\beta(x_1, x_2^k, \lambda^k) + \frac{1}{2} \|x_1 - x_1^k\|_{R_1}^2, \\ x_2^{k+1} \in \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^k) + \frac{1}{2} \|x_2 - x_2^k\|_{R_2}^2, \\ \lambda^{k+1} = \lambda^k - \gamma\beta (A_1x_1^{k+1} + A_2x_2^{k+1} - b), \end{cases} \quad (3)$$

where the constant  $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ ,  $R_1$  and  $R_2$  are some positive semi-definite matrices with compatible dimensionality. In many applications, by taking some special cases of  $R_1$  and  $R_2$ , the two subproblems of (3) often simple enough to admit closed form solutions [8, 9]. Furthermore, the numerical results in [10] indicate that the constant  $\gamma$  should be larger than one to accelerate the convergence of ADMM.

Therefore, a question can be raised: can the restriction region of  $\gamma$  be relaxed? In this paper, we answer this question positively. In fact, we propose a proximal ADMM with larger step size, and in our new method, the constant  $\gamma$  can take any values of the interval  $(0, +\infty)$ , and the expense is that the output of (3) must be corrected by some convex combination of the current iterate and the output of (3). Therefore, our new method can be categorized to the prediction-correction type ADMMs [11, 12, 13, 14]. However, compared with other similar methods, such as [11, 12], our correction step is quite simple, because the correction steps in [11, 12] need to compute a dynamically updated step size, which needs much computational effort, especially when the involved matrix is not a diagonal matrix.

The remainder of this paper is organized as follows. In next section, 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 model (1), and prove its global convergence. In Section 4, some numerical experiments are given to demonstrate the advantage of our new method.

## 2 Preliminaries

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

Firstly, we give some notation used in this paper. 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  $\|x\|_G = \sqrt{x^\top G x}$  the  $G$ -norm of the vector  $x$ . The effective domain of a function  $f : \mathcal{X} \rightarrow (-\infty, +\infty]$  is defined as  $\text{dom}(f) := \{x \in \mathcal{X} | f(x) < +\infty\}$ . The set of all relative interior points of a given nonempty convex set  $\mathcal{C}$  is denoted by  $\text{ri}(\mathcal{C})$ .

**Definition 2.1.** [17] A function  $f : \mathcal{R}^n \rightarrow \mathcal{R}$  is convex if and only if

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

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

$$f(x) \geq f(y) + \langle \xi, x - y \rangle, \forall x, y \in \mathcal{R}^n, \xi \in \partial f(y), \quad (4)$$

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

We make the following standard assumptions about (1) in this paper.

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

**Assumption 2.2.** The matrices  $A_i$  ( $i = 1, 2$ ) are full-column rank.

**Assumption 2.3.** The generalized Slater's condition holds, i.e., there is a point  $(\hat{x}_1, \hat{x}_2) \in \text{ri}(\text{dom}\theta_1 \times \text{dom}\theta_2) \cap P$ , where

$$P := \{x = (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 | A_1 x_1 + A_2 x_2 = b\}.$$

Under Assumption 2.3, it follows from Theorem 3.22 and Theorem 3.23 of [19] that the vector  $x^* = (x_1^*, x_2^*) \in \mathcal{R}^{n_1+n_2}$  is an optimal solution to problem (1) if and only if there exists a vector  $\lambda^* \in \mathcal{R}^l$  such that

$$\begin{cases} (x_1^*, x_2^*) \in \mathcal{X}_1 \times \mathcal{X}_2; \\ \theta_i(x_i) - \theta_i(x_i^*) + (x_i - x_i^*)^\top (-A_i^\top \lambda^*) \geq 0, \forall x_i \in \mathcal{X}_i, i = 1, 2; \\ A_1 x_1^* + A_2 x_2^* = b. \end{cases} \quad (5)$$

Moreover, any  $\lambda^* \in \mathcal{R}^l$  satisfying (5) is an optimal solution to the dual of problem (1). Obvious, (5) can be written as the following mixed variational inequality problem, denoted by  $\text{VI}(\mathcal{W}, F, \theta)$ : Find a vector  $w^* \in \mathcal{W}$  such that

$$\theta(x) - \theta(x^*) + (w - w^*)^\top F(w^*) \geq 0, \quad \forall w \in \mathcal{W}, \quad (6)$$

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

$$F(w) := \begin{pmatrix} -A_1^\top \lambda \\ -A_2^\top \lambda \\ A_1 x_1 + A_2 x_2 - b \end{pmatrix}. \quad (7)$$

The set of the solutions of the KKT systems  $\text{VI}(\mathcal{W}, F, \theta)$ , denoted by  $\mathcal{W}^*$ , is nonempty by the Assumption 2.3 and Remark 2.1. It is easy to verify that the mapping  $F(\cdot)$  is not only monotone but also satisfies the following desired property:

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

### 3 Proximal ADMM with large step size and its global convergence

In this section, we pay attention to present a proximal ADMM with large step size for solving (1) and prove its global convergence under Assumptions 2.1-2.3. At each iteration, the new method is composed of two steps: the prediction step and the correction step. More specifically, the new method first generates a trial iterate via the iterative scheme (3) in the prediction step, and then yields the new iterate via the convex combination of the current iterate and the trial iterate in the contraction step.

First, let's define an important constant  $\eta$ , which is used in our new method.

$$\eta = \begin{cases} \gamma, & \text{if } 0 < \gamma \leq 1, \\ \frac{1}{\gamma}, & \text{if } \gamma > 1. \end{cases} \quad (8)$$

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

The iterative scheme of the new proximal ADMM for solving the model (1) is as follows.

**Algorithm 1.** Proximal ADMM

**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 initial point  $w^0 = (x_1^0, x_2^0, \lambda^0) \in \mathcal{W}$  arbitrarily. 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

$$\begin{cases} \tilde{x}_1^k \in \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \mathcal{L}_\beta(x_1, x_2^k, \lambda^k) + \frac{1}{2} \|x_1 - x_1^k\|_{R_1}^2, \\ \tilde{x}_2^k \in \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \mathcal{L}_\beta(\tilde{x}_1^k, x_2, \lambda^k) + \frac{1}{2} \|x_2 - x_2^k\|_{R_2}^2, \\ \tilde{\lambda}^k = \lambda^k - \gamma\beta (A_1 \tilde{x}_1^k + A_2 \tilde{x}_2^k - b), \end{cases} \quad (9)$$

**Step 2.** (Stopping conditions) If  $\|w^k - \tilde{w}^k\| \leq \varepsilon$ , 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  $\eta$  is defined by (8). Go to Step 1.

**Remark 3.1.** To accelerate Algorithm 1, the Fortin and Glowinski's constant  $\gamma$  is attached to the update of  $\lambda$  in the prediction step, whose feasible region is  $(0, +\infty)$ , and is relaxed remarkably compared with the Fortin and Glowinski's constant  $\gamma \in (0, \frac{1+\sqrt{5}}{2})$  in the methods of [7, 10].

**Remark 3.2.** From  $\eta \in (0, 1], \rho \in (0, \eta]$ , it is obvious that the new iterate  $w^{k+1}$  is a convex combination of  $w^k$  with  $\tilde{w}^k$ .

**Remark 3.3.** 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. [14] 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  $\rho$  in [13] is restricted in  $(0, 2(1 - \sqrt{6}/3))$  and the constant  $\rho$  in [14] is restricted in  $(0, 2(1 - \sqrt{2}/2))$ . Obviously, the feasible set of  $\rho$  in Algorithm 1 is larger than these in [13, 14] when  $2 - \sqrt{2} < \gamma < 1/(2 - \sqrt{2})$ .

If Algorithm 1 stops at Step 2, then the current iterate  $w^k$  is a proximal solution of  $\text{VI}(\mathcal{W}, F, \theta)$ . Thus, it is assumed, without loss of generality, that Algorithm 1 generates two infinite sequences  $\{w^k\}$  and  $\{\tilde{w}^k\}$ .

Before proving the global convergence of Algorithm 1, we define two matrices to simplify our notation in the later analysis.

$$M = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & \beta A_2^\top A_2 + R_2 & 0 \\ 0 & 0 & \frac{1}{\beta\gamma} I_l \end{pmatrix}, \quad Q = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & \beta A_2^\top A_2 + R_2 & \frac{1}{2\gamma} A_2^\top \\ 0 & \frac{1}{2\gamma} A_2 & \frac{1}{\beta\gamma^2} I_l \end{pmatrix}. \quad (10)$$

The two matrices  $M, Q$  just defined has the following properties.

**Lemma 3.1.** When  $R_1$  and  $R_2$  are two positive semi-definite matrices, we have

- (i). the matrices  $M$  and  $Q$  are positive semi-definite;
- (ii). the matrix  $H_1 = 2Q - \gamma M$  is positive semi-definite if  $0 < \gamma \leq 1$ , 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

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

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  $0 < \gamma \leq 1$ , 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 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 show that

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

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

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  $\|w^k - \tilde{w}^k\| \leq \varepsilon$  to terminate Algorithm 1.

**Lemma 3.2.** If  $A_i x_i^k = A_i \tilde{x}_i^k$  ( $i = 1, 2$ ),  $\lambda^k = \tilde{\lambda}^k$ , then the vector  $(x_1^k, x_2^k, \lambda^k)$  is a solution of VI( $\mathcal{W}, F, \theta$ ).

**Proof.** By the first-order optimization condition for  $x_1$ -subproblem in (9) and Lemma 2.1, we have

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

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

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

i.e.,

$$\begin{aligned} \theta_1(x_1) - \theta_1(\tilde{x}_1^k) + (x_1 - \tilde{x}_1^k)^\top \left\{ -A_1^\top \tilde{\lambda}^k + \frac{1 - \gamma}{\gamma} A_1^\top (\lambda^k - \tilde{\lambda}^k) + \beta A_1^\top A_2(x_2^k - \tilde{x}_2^k) \right\} \\ \geq (x_1 - \tilde{x}_1^k)^\top R_1(x_1^k - \tilde{x}_1^k), \forall x_1 \in \mathcal{X}_1. \end{aligned} \quad (11)$$

Similarly, for the other variable  $x_2$ , by the first-order optimization condition for  $x_2$ -subproblem in (9) and Lemma 2.1, we have

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

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, we have

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

i.e.,

$$\begin{aligned} \theta_2(x_2) - \theta_2(\tilde{x}_2^k) + (x_2 - \tilde{x}_2^k)^\top \left\{ -A_2^\top \tilde{\lambda}^k + \frac{1-\gamma}{\gamma} A_2^\top (\lambda^k - \tilde{\lambda}^k) \right\} \\ \geq (x_2 - \tilde{x}_2^k)^\top R_2(x_2^k - \tilde{x}_2^k), \forall x_2 \in \mathcal{X}_2. \end{aligned} \quad (12)$$

Furthermore, the updating formula of  $\lambda$  in (9) can be written as the following variational inequality problem

$$(\lambda - \tilde{\lambda}^k)^\top \left\{ \left( \sum_{i=1}^2 A_i \tilde{x}_i^k - b \right) + \frac{1}{\beta\gamma} (\tilde{\lambda}^k - \lambda^k) \right\} \geq 0, \forall \lambda \in \mathcal{R}^l. \quad (13)$$

Then, adding (11), (12) and (13), we get

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (w - \tilde{w}^k)^\top G(w^k - \tilde{w}^k), \forall w \in \mathcal{W}, \quad (14)$$

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 (13), we get

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

This together with Assumption 2.2 implies that

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

which indicates that  $(x_1^k, x_2^k, \lambda^k)$  is a solution of  $\text{VI}(\mathcal{W}, F, \theta)$ . 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 3.3.** Let  $\{w^k\}$  and  $\{\tilde{w}^k\}$  be the two sequences generated by Algorithm 1. Then, we have

$$(w^k - w^*)^\top M(w^k - \tilde{w}^k) \geq \|w^k - \tilde{w}^k\|_Q^2, \forall w^* \in \mathcal{W}^*. \quad (15)$$

**Proof.** Since  $w^* \in \mathcal{W}^* \subseteq \mathcal{W}$ , it follows from (14) that

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

Thus,

$$\begin{aligned} & (\tilde{w}^k - w^*)^\top G(w^k - \tilde{w}^k) \\ & \geq \theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^\top F(\tilde{w}^k) \\ & \geq 0, \end{aligned} \quad (16)$$

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

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

$$\begin{aligned} & (\tilde{w}^k - w^*)^\top G(w^k - \tilde{w}^k) \\ & = (\tilde{w}^k - w^*)^\top M(w^k - \tilde{w}^k) + (\tilde{w}^k - w^*)^\top \begin{pmatrix} 0 & -\beta A_1^\top A_2 & -\frac{1-\gamma}{\gamma} A_1^\top \\ 0 & -\beta A_2^\top A_2 & -\frac{1-\gamma}{\gamma} A_2^\top \\ 0 & 0 & 0 \end{pmatrix} (w^k - \tilde{w}^k) \\ & = (\tilde{w}^k - w^*)^\top M(w^k - \tilde{w}^k) - \beta (A_1 \tilde{x}_1 + A_2 \tilde{x}_2 - b)^\top A_2 (x_2^k - \tilde{x}_2^k) \\ & \quad - \frac{1-\gamma}{\gamma} (A_1 \tilde{x}_1 + A_2 \tilde{x}_2 - b)^\top (\lambda^k - \tilde{\lambda}^k) \\ & = (\tilde{w}^k - w^*)^\top M(w^k - \tilde{w}^k) - \frac{1}{\gamma} (\lambda^k - \tilde{\lambda}^k)^\top A_2 (x_2^k - \tilde{x}_2^k) - \frac{1-\gamma}{\beta\gamma^2} \|\lambda^k - \tilde{\lambda}^k\|^2, \end{aligned}$$

where the second inequality follows from  $w^* \in \mathcal{W}^*$ , and the second equality comes from (9).

Then, substituting the above relationship into the left side of (16), we get

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

i.e.,

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

where the equality follows from the definition (8). The lemma is proved.

The inequality (15) and Lemma 2.1 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 1 is Fejér monotone with respect to the solution set  $\mathcal{W}^*$  of  $\text{VI}(\mathcal{W}, F, \theta)$ .

**Theorem 3.1.** For any  $w^* = (x_1^*, x_2^*, x_3^*, \lambda^*) \in \mathcal{W}^*$ , there exists a positive scalar  $\eta$  such that the sequence  $\{w^k\}$  generated by PFPSM satisfies

$$\|w^{k+1} - w^*\|_M^2 \leq \|w^k - w^*\|_M^2 - \rho \|w^k - \tilde{w}^k\|_H^2, \quad (17)$$

where the matrix  $H$  is defined by

$$H = \begin{cases} (\gamma - \rho)M, & \text{if } 0 < \gamma \leq 1, \\ (\frac{1}{\gamma} - \rho)M, & \text{if } \gamma > 1. \end{cases} \quad (18)$$

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

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

where the first inequality follows from (15), and the second inequality comes from Lemma 2.1.

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

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

where the second inequality also comes from Lemma 2.1. The assertion (17) is proved.

**Remark 3.4.** By the definition of  $\eta$  in (8), the matrix  $H$  in (18) is positive semi-definite.

Based on the Fejér monotonicity of the sequence  $\{w^k\}$ , the global convergence of Algorithm 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 2.2, both matrices  $\tilde{M}$  and  $\tilde{H}$  are positive definite.

**Theorem 3.2.** The sequence  $\{\tilde{w}^k\}$  generated Algorithm 1 converges to some  $w^\infty \in \mathcal{W}^*$ .

**Proof.** Using (17) and noting  $\rho > 0$ , we have

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

Then  $\{v^k\}$  is a bounded sequence by the positive definiteness of  $\tilde{M}$ . By (17) again, we have

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

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

$$\lim_{k \rightarrow \infty} \|x_1^k - \tilde{x}_1^k\|_{R_1} = \lim_{k \rightarrow \infty} \|v^k - \tilde{v}^k\| = 0. \quad (19)$$

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

$$\begin{aligned} & \|A_1 \tilde{x}_1^k\| \\ &= \left\| \frac{1}{\gamma\beta} (\lambda^k - \tilde{\lambda}^k) - A_2 \tilde{x}_2^k + b \right\| \\ &\leq \frac{1}{\gamma\beta} (\|\lambda^k\| + \|\tilde{\lambda}^k\|) + \|A_2\| \|\tilde{x}_2^k\| + \|b\|, \end{aligned}$$

and Assumption 2.2, we have the sequence  $\{\tilde{x}_1^k\}$  is bounded. Then, the sequence  $\{\tilde{w}^k\}$  is bounded. Then, 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

$$A_1 x_1^\infty + A_2 x_2^\infty - b = 0.$$

Furthermore, taking limits on both sides of (11) and (12), and using (17), we obtain

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

and

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

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

Then, from  $\lim_{k \rightarrow \infty} (v^k - \tilde{v}^k) = 0$  and  $\{\tilde{v}^{k_j}\} \rightarrow v^\infty$ , 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  $k \geq k_l$ , it follows from the above two inequalities and (17) that

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

which indicates that the sequence  $\{v^k\}$  converges globally to the point  $v^\infty$ , and thus the sequence  $\{\tilde{v}^k\}$  also converges globally to the point  $v^\infty$ . 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 2.2 again, it can be inferred that the sequence  $\{\tilde{x}_1^k\}$  converges to  $x_1^\infty$ . Overall, the sequence  $\{\tilde{w}^k\}$  generated Algorithm 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 least-squares problems to verify the efficiency of Algorithm 1. All the code 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

$$\min\left\{\frac{1}{2}\|X - C\|_F^2 \mid X \in \mathcal{S}_+^n \cap \mathcal{S}_B\right\}, \quad (20)$$

where

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

and

$$\mathcal{S}_B = \{H \in \mathcal{R}^{n \times n} \mid H^\top = H, H_L \leq H \leq H_U\}.$$

We use the following Matlab scripts to generate the problem data:

$$\text{rand('state', 0); } C = \text{rand}(n, n); C = (C' + C) - \text{ones}(n, n) + \text{eye}(n);$$

$$HU = \text{ones}(n) * 0.1; HL = -HU; \text{for } i = 1 : n \text{ } HU(i, i) = 1; HL(i, i) = 1; \text{end};$$

The numerical results of Algorithm 1 are listed in Table 1. The numerical results in Table 1 indicate

$n$	$\beta$	$\gamma$	No. It	CPU Sec.
50	5	1.0	45	0.5772
50	5	1.1	51	0.7956
100	5	1.0	46	3.4944
100	5	1.1	49	3.6972
200	10	1.0	54	22.6357
200	10	1.1	58	24.2582

that Algorithm 1 can successfully solved all the tested scenarios.

## Acknowledgment

This work is supported by the foundation of Scientific Research Project of Shandong Universities (J15LI11).

## References

- [1] Boyd S., Parikh N., Chu E., Peleato B. and Eckstein J., Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, (2011) 3: 1-122.

- [2] Bardsley J.M., Knepper S., Nagy J., Structured linear algebra problems in adaptive optics imaging. *Adv. Comput. Math.*, 35 (2011) 35: 103-117.
- [3] He B.S, Tao M., Yuan X.M., A splitting method for separable convex programming, *IMA J. Numer. Anal.*, (2015) 35(1): 394-426.
- [4] Bnouhachem A., Xu M.H., An inexact LQP alternating direction method for solving a class of structured variational inequalities. *Comput. Math. Appl.*, (2014)67: 671-680.
- [5] Sun M., Liu J., An inexact generalized PRSM with LQP regularization for structured variational inequalities and its applications to traffic equilibrium problems. *Journal of Inequalities and Applications*, December 2016, 150.
- [6] Glowinski R., *Numerical Methods for Nonlinear Variational Problems*. Springer, New York (1984).
- [7] Xu M.H., Wu T., A class of linearized proximal alternating direction methods. *J. Optim. Theory Appl.*, (2011) 155: 321-337.
- [8] Yang J.F., Zhang Y., Alternating direction algorithms for  $\ell_1$ -problems in compressive sensing, *SIAM J. Sci. Comput.*, (2011) 332: 250-278.
- [9] Zhang X.Q., Burger M., Osher S., A unified primal-dual algorithm framework based on Bregman iteration. *J. Sci. Comput.*, (2010) 6: 20-46.
- [10] He B.S., Wang S.L., Yang H., A modified variable-penalty alternating direction method for monotone variational inequalities. *J. Comput. Math.*, (2003) 21: 495-204.
- [11] Ye C.H., Yuan X.M., A descent method for structured monotone variational inequalities. *Optimization Methods and Software*, (2007) 22(2): 329-338.
- [12] Yuan X.M., An improved proximal alternating direction method for monotone variational inequalities with separable structure. *Comput. Optim. Appl.*, (2011) 49(1): 17-29.
- [13] He B.S., Hou L.S., Yuan X.M., On full Jacobian decomposition of the augmented lagrangian method for separable convex programming. *SIAM J. Optim.*, (2015) 25(4): 2274-2312.
- [14] Wang K., Desai J., He H.J., A proximal partially parallel splitting method for separable convex programs. *Optimization Methods and Software*, (2016) (Accepted).
- [15] Sun M., Wang Y.J., Liu J., Generalized Peaceman-Rachford splitting method for multiple-block separable convex programming with applications to robust PCA, *Calcolo*, (2016) (Accepted).

- [16] Sun M., Liu J., Generalized Peaceman-Rachford splitting method for separable convex programming with applications to image processing, *Journal of Applied Mathematics and Computing*, (2016) 51(1): 605-622.
- [17] Rockafellar, R.T., *Convex Analysis*. Princeton: Princeton University Press, 1970.
- [18] Bonnans J.F., Shapiro A., *Perturbation Analysis of Optimization Problems*. Springer-Verlag, 2000.
- [19] Fukushima M., *Fundamentals of Nonlinear Optimization*. Asakura Shoten, Tokyo, 2001.