

A symmetric version of the generalized alternating direction method of multipliers for two-block separable convex programming

Jing Liu^{1,2}, Yongrui Duan¹, Min Sun^{3,4}

1. School of Economics and Management, Tongji University,
Shanghai, 200092, P.R. China

2. School of Data Sciences, Zhejiang University of Finance and Economics,
Zhejiang, 310018, P. R. China

3. School of Mathematics and Statistics, Zaozhuang University,
Shandong, 277160, P. R. China

4. School of Management, Qufu Normal University,
Shandong, 276826, P. R. China

Abstract. This paper introduces a symmetric version of the generalized alternating direction method of multipliers for two-block separable convex programming with linear equality constraints, which inherits the superiorities of the classical alternating direction method of multipliers (ADMM), and extends the feasible set of the relaxation factor α of the generalized ADMM to the infinite interval $[1, +\infty)$. Under the conditions that the objective function is convex and the solution set is nonempty, we establish the convergence results of the proposed method, including the global convergence, the worst-case $\mathcal{O}(1/k)$ convergence rate in both the ergodic and non-ergodic senses, where k denotes the iteration counter. Numerical experiments to decode a sparse signal arising in compressed sensing are included to illustrate the efficiency of the new method.

Keywords: Alternating direction method of multipliers; convex programming; mixed variational inequalities; compressed sensing.

AMS subject classifications. 90C25, 90C30

1 Introduction

We consider the two-block separable convex programming with linear equality constraints, where the objective function is the sum of two individual functions with decoupled variables:

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

where $\theta_i : \mathcal{R}^{n_i} \rightarrow \mathcal{R} (i = 1, 2)$ are closed proper convex functions; $A_i \in \mathcal{R}^{l \times n_i} (i = 1, 2)$ and $b \in \mathcal{R}^l$, and $\mathcal{X}_i \subseteq \mathcal{R}^{n_i} (i = 1, 2)$ are given nonempty closed convex sets. The linear constrained convex problem (1) is a unified framework of many problems arising in real world, including compressed sensing, image restoration, and statistical learning, and so forth (see for example, [1, 2, 3]). An important special case of (1) is the following linear inverse problem:

$$\min_{x \in \mathcal{R}^n} \mu \|x\|_1 + \frac{1}{2} \|Ax - y\|^2, \quad (2)$$

where $A \in \mathcal{R}^{m \times n}$ and $y \in \mathcal{R}^m$ are given matrix and vector, $\mu > 0$ is a regularization parameter and $\|x\|_1$ is the ℓ_1 -norm of a vector x defined as $\|x\|_1 = \sum_{i=1}^n |x_i|$. Then setting $x_1 := Ax - y, x_2 := x$, (2) can be converted into the following two-block separable convex programming

$$\begin{aligned} \min \quad & \frac{1}{2} \|x_1\|^2 + \mu \|x_2\|_1 \\ \text{s.t.} \quad & -x_1 + Ax_2 = y, \\ & x_1 \in \mathcal{R}^m, x_2 \in \mathcal{R}^n, \end{aligned} \quad (3)$$

which is a special case of problem (1) with the following specifications:

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

1.1 Existing algorithms

In their seminal works, Glowinski et al. [4] and Gabay et al. [5] independently developed the alternating direction method of multipliers (ADMM), which is an influential first-order method for solving problem (1). ADMM can be regarded as an application of the Douglas-Rachford splitting method (DRSM) [7] to the dual of (1), or a special case of the proximal point algorithm (PPA) [8, 9] in the cyclic sense. We refer to [10] for more detailed relationship. With any initial vectors $x_1^0 \in \mathcal{X}_1, \lambda^0 \in \mathcal{R}^l$, the iterative scheme of ADMM reads as

$$\begin{cases} x_1^{k+1} \in \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \{\theta_1(x_1) - x_1^\top A_1^\top \lambda^k + \frac{\beta}{2} \|A_1 x_1 + A_2 x_2^k - b\|^2\}, \\ x_2^{k+1} \in \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \{\theta_2(x_2) - x_2^\top A_2^\top \lambda^k + \frac{\beta}{2} \|A_1 x_1^{k+1} + A_2 x_2 - b\|^2\}, \\ \lambda^{k+1} = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \end{cases} \quad (4)$$

where $\lambda \in \mathcal{R}^l$ is the Lagrangian multiplier and $\beta > 0$ is a penalty parameter. The main characteristic of ADMM is that it fully exploits the separable structure of problem (1), and updates the variables x_1, x_2, λ in an alternating order by solving a series of low-dimensional sub-problems with only one unknown variable.

In the past few decades, ADMM has received a revived interest, and it has become a research focus in optimization community, especially in the (non)convex optimization. Many efficient ADMM-type methods have been developed, including the proximal ADMM [6, 9], the generalized ADMM [11], the symmetric ADMM [12], the inertial ADMM [18], and some proximal ADMM-type methods

[13, 14, 19, 20, 21]. Specifically, the proximal ADMM attaches some proximal terms to the subproblems of ADMM (9). The generalized ADMM updates the variables x_2 and λ by including a relaxation factor $\alpha \in (0, 2)$, and $\alpha \in (1, 2)$ is often advantageous to speed up its performance. The symmetric ADMM updates the Lagrangian multiplier λ twice at each iteration and includes two relaxation factors $\alpha \in (0, 1), \beta \in (0, 1)$. Recent researches of the symmetric ADMM can be found in [12, 14, 21]. The inertial ADMM unifies the basic ideas of the inertial PPA and ADMM, which utilizes the latest two iterates to generate the new iterate, therefore it can be viewed as a multistep method. For the proximal ADMM, the objective functions of its subproblems are often strongly convex, which are often easier to be solved than those of (9). However, a new challenge has arisen for the proximal ADMM-type methods. That is how to choose a proper proximal matrix. In fact, most proximal ADMM-type methods need to estimate the matrix norm $\|A_i^\top A_i\| (i = 1, 2)$, which demands lots of calculations, especially for large $n_i (i = 1, 2)$. Quite recently, some customized Douglas-Rachford splitting algorithms [15, 16, 17], and the proximal ADMM-type methods with indefinite proximal regularization are developed [28, 29], which dissolve the above problem to some extent. All the above mentioned ADMM-type methods are generalizations of the classical ADMM, because they all reduce to iterative scheme (9) by choosing some special parameters. For more new development of the ADMM-type methods, including the convergence rate, acceleration techniques, its generalization for solving multi-block separable convex programming and nonconvex, nonsmooth programming, we refer to [30, 32, 33, 35, 36].

1.2 Contributions and organization

In this paper, we are going to further study the generalized ADMM. Note that the first subproblem in the generalized ADMM is irrelevant to the relaxation factor α . That is, the updating formula for x_1 does not incorporate the relaxation factor α explicitly. Furthermore, $\alpha \in (1, 2)$ is often advantageous for the generalized ADMM [13]. Therefore, in this paper, we are going to propose a new generalized ADMM, whose both subproblems incorporate the relaxation factor α directly. The new method generalizes the method proposed in [22] by relaxing the feasible set of α from the interval $[1, 2)$ to the infinite interval $[1, +\infty)$, and can be viewed as a symmetric version of the generalized ADMM.

The rest of the paper is organized as follows. In Section 2, we summarize some necessary preliminaries and characterize problem (1) by a mixed variational inequality problem. In Section 3, we describe the new symmetric version of the generalized ADMM and establish its convergence results in detail. In Section 4, some compressed sensing experiments are given to illustrate the efficiency of the proposed method. Some conclusions are drawn in Section 5.

2 Preliminaries

In this section, some necessary preliminaries which are useful for further discussions are presented, and to make our analysis more succinct, some positive definite or positive semi-definite block matrices are defined and their properties are investigated.

For two real matrices $A \in \mathcal{R}^{s \times m}$, $B \in \mathcal{R}^{n \times s}$, the Kronecker product of A and B is defined as $A \otimes B = (a_{ij}B)$. Let $\|\cdot\|_p$ ($p = 1, 2$) denote the standard definition of ℓ_p -norm; in particular, $\|\cdot\| = \|\cdot\|_2$. For any two vectors $x, y \in \mathcal{R}^n$, $\langle x, y \rangle$ or $x^\top y$ denote their inner product, and for any symmetric matrix $G \in \mathcal{R}^{n \times n}$, the symbol $G \succ 0$ (resp., $G \succeq 0$) denotes that G is positive definite (resp., semi-definite). For any $x \in \mathcal{R}^n$ and $G \succeq 0$, the G -norm $\|x\|_G$ of the vector x is defined as $\sqrt{x^\top G x}$. The effective domain of a closed proper function $f : \mathcal{X} \rightarrow (-\infty, +\infty]$ is defined as $\text{dom}(f) := \{x \in \mathcal{X} | f(x) < +\infty\}$, and the symbol $\text{ri}(\mathcal{C})$ denotes the set of all relative interior points of a given nonempty convex set \mathcal{C} . Furthermore, we use the following notations:

$$x = (x_1, x_2), w = (x, \lambda).$$

Definition 2.1. [23] 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

$$f(x) \geq f(y) + \langle \xi, x - y \rangle, \forall x, y \in \mathcal{R}^n, \xi \in \partial f(y), \quad (5)$$

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 .

Throughout the paper, we make the following standard assumptions for problem (1).

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 $(\hat{x}_1, \hat{x}_2) \in \text{ri}(\text{dom}\theta_1 \times \text{dom}\theta_2) \cap \{x = (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 | A_1 x_1 + A_2 x_2 = b\}$.

2.1 The mixed variational inequality problem

Under Assumption 2.3, it follows from Theorem 3.22 and Theorem 3.23 of [25] that $x^* = (x_1^*, x_2^*) \in \mathcal{R}^{n_1+n_2}$ is an optimal solution to problem (1) if and only if there exists a vector $\lambda^* \in \mathcal{R}^l$ such that $(x_1^*, x_2^*, \lambda^*)$ is a solution of the following KKT system:

$$\begin{cases} 0 \in \partial\theta_i(x_i^*) - A_i^\top \lambda^* + N_{\mathcal{X}_i}(x_i^*), i = 1, 2, \\ A_1 x_1^* + A_2 x_2^* = b, \end{cases} \quad (6)$$

where $N_{\mathcal{X}_i}(x_i^*)$ is the normal cone of the convex set \mathcal{X}_i at the point x_i^* , which is defined as $N_{\mathcal{X}_i}(x_i^*) = \{z \in \mathcal{R}^{n_i} \mid \langle z, x_i - x_i^* \rangle \leq 0, \forall x_i \in \mathcal{X}_i\}$. Then, for the nonempty convex set \mathcal{X}_i and $\forall x_i \in \mathcal{X}_i$, it follows from [24] (Example 2.123) that $N_{\mathcal{X}_i}(x_i) = \partial\delta(\cdot|\mathcal{X}_i)(x_i)$, where $\delta(\cdot|\mathcal{X}_i)$ is the indicator function of the set \mathcal{X}_i , and $\partial\delta(\cdot|\mathcal{X}_i)(x_i)$ is the subdifferential mappings of $\delta(\cdot|\mathcal{X}_i)$ at the point $x_i \in \mathcal{X}_i$.

Lemma 2.1. For any vector $x_i^* \in \mathcal{R}^{n_i}, \lambda^* \in \mathcal{R}^l$, the relationship $0 \in \partial\theta_i(x_i^*) - A_i^\top \lambda^* + \partial\delta(\cdot|\mathcal{X}_i)(x_i^*)$ is equivalent to $x_i^* \in \mathcal{X}_i$ and the inequality

$$\theta_i(x_i) - \theta_i(x_i^*) + (x_i - x_i^*)^\top (-A_i^\top \lambda^*) \geq 0, \forall x_i \in \mathcal{X}_i.$$

Proof. From $0 \in \partial\theta_i(x_i^*) - A_i^\top \lambda^* + \partial\delta(\cdot|\mathcal{X}_i)(x_i^*)$, we have $x_i^* \in \mathcal{X}_i$ and there exists $\eta_i \in \partial\delta(\cdot|\mathcal{X}_i)(x_i^*)$ such that

$$A_i^\top \lambda^* - \eta_i \in \partial\theta_i(x_i^*).$$

From the subgradient inequality (10), one has

$$\theta_i(x_i) - \theta_i(x_i^*) \geq (x_i - x_i^*)^\top (A_i^\top \lambda^* - \eta_i), \forall x_i \in \mathcal{R}^{n_i}.$$

Thus,

$$\theta_i(x_i) - \theta_i(x_i^*) - (x_i - x_i^*)^\top (-A_i^\top \lambda^*) \geq (x_i - x_i^*)^\top (-\eta_i) \geq 0, \forall x_i \in \mathcal{X}_i,$$

where the second inequality comes from $x_i^* \in \mathcal{X}_i$ and $\eta_i \in \partial\delta(\cdot|\mathcal{X}_i)(x_i^*)$.

Conversely, from $\theta_i(x_i) - \theta_i(x_i^*) + (x_i - x_i^*)^\top (-A_i^\top \lambda^*) \geq 0, \forall x_i \in \mathcal{X}_i$, we have

$$\theta_i(x_i) + x_i^\top (-A_i^\top \lambda^*) \geq \theta_i(x_i^*) + (x_i^*)^\top (-A_i^\top \lambda^*), \forall x_i \in \mathcal{X}_i,$$

which together with $x_i^* \in \mathcal{X}_i$ implies that

$$x_i^* = \operatorname{argmin}_{x_i \in \mathcal{X}_i} \{\theta_i(x_i) + x_i^\top (-A_i^\top \lambda^*)\}.$$

From this and Theorem 3.22 of [25], we have $0 \in \partial\theta_i(x_i^*) - A_i^\top \lambda^* + \partial\delta(\cdot|\mathcal{X}_i)(x_i^*)$. This completes the proof. \square

Remark 2.1. Based on (11) and Lemma 2.1, 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

$$\begin{cases} (x_1^*, x_2^*) \in \mathcal{X}_1 \times \mathcal{X}_2; \\ \theta_i(x_i) - \theta_i(x_i^*) + (x_i - x_i^*)^\top (-A_i^\top \lambda^*) \geq 0, \forall x_i \in \mathcal{X}_i, i = 1, 2; \\ A_1 x_1^* + A_2 x_2^* = b. \end{cases} \quad (7)$$

Moreover, any $\lambda^* \in \mathcal{R}^l$ satisfying (12) is an optimal solution to the dual of problem (1). Obvious, (12) 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

$$\theta(x) - \theta(x^*) + (w - w^*)^\top F(w^*) \geq 0, \quad \forall w \in \mathcal{W}, \quad (8)$$

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

$$F(w) := \begin{pmatrix} -A_1^\top \lambda \\ -A_2^\top \lambda \\ A_1 x_1 + A_2 x_2 - b \end{pmatrix} = \begin{pmatrix} 0 & 0 & -A_1^\top \\ 0 & 0 & -A_2^\top \\ A_1 & A_2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \lambda \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ b \end{pmatrix}. \quad (9)$$

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

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

2.2 Three matrices and their properties

To present our analysis in a compact way, now let us define some matrices. For any $R_i \in \mathcal{R}^{n_i \times n_i}$ ($i = 1, 2$) $\succeq 0$, set

$$M = \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & -\beta A_2 & I_l \end{pmatrix} \quad (10)$$

and for $\alpha \in [1, +\infty)$, set

$$Q = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 + (2\alpha - 1)\beta A_2^\top A_2 & \frac{1-\alpha}{\alpha} A_2^\top \\ 0 & -A_2 & \frac{1}{\alpha\beta} I_l \end{pmatrix}, \quad H = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 + \frac{2\alpha^2 - 2\alpha + 1}{\alpha} \beta A_2^\top A_2 & \frac{1-\alpha}{\alpha} A_2^\top \\ 0 & \frac{1-\alpha}{\alpha} A_2 & \frac{1}{\alpha\beta} I_l \end{pmatrix}. \quad (11)$$

The above defined three matrices M, Q, H satisfy the following properties.

Lemma 2.2. If $\alpha \in \mathcal{R}$ and $R_i \succeq 0$ ($i = 1, 2$), then the matrix H defined in (16) is positive semi-definite.

Proof. Set $t = 2\alpha^2 - 2\alpha + 1$, which is positive for any $\alpha \in \mathcal{R}$. By (16), we have

$$H = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{t\beta}{\alpha} A_2^\top A_2 & \frac{1-\alpha}{\alpha} A_2^\top \\ 0 & \frac{1-\alpha}{\alpha} A_2 & \frac{1}{\alpha\beta} I_l \end{pmatrix}$$

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

$$\frac{1}{\alpha} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sqrt{\beta} A_2^\top & 0 \\ 0 & 0 & \frac{1}{\sqrt{\beta}} I_l \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & tI_l & (1-\alpha)I_l \\ 0 & (1-\alpha)I_l & I_l \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sqrt{\beta} A_2 & 0 \\ 0 & 0 & \frac{1}{\sqrt{\beta}} I_l \end{pmatrix}.$$

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

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & t & 1 - \alpha \\ 0 & 1 - \alpha & 1 \end{pmatrix} \otimes I_l,$$

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

$$\begin{pmatrix} t & 1 - \alpha \\ 1 - \alpha & 1 \end{pmatrix}$$

is positive semi-definite. In fact,

$$t - (1 - \alpha)^2 = \alpha^2 \geq 0.$$

Therefore, the matrix H is positive semi-definite. The proof is then complete.

Lemma 2.3. If $\alpha \in [1, +\infty)$ and $R_i \geq 0 (i = 1, 2)$, then the matrices M, Q, H defined, respectively, in (15), (16) satisfy the following relationships:

$$HM = Q \tag{12}$$

and

$$Q^\top + Q - M^\top HM \succeq \frac{\alpha - 1}{2\alpha} M^\top HM. \tag{13}$$

Proof. From (15) and (16), we have

$$\begin{aligned} HM &= \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 + \frac{2\alpha^2 - 2\alpha + 1}{\alpha} \beta A_2^\top A_2 & \frac{1 - \alpha}{\alpha} A_2^\top \\ 0 & \frac{1 - \alpha}{\alpha} A_2 & \frac{1}{\alpha\beta} I_l \end{pmatrix} \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & -\beta A_2 & I_l \end{pmatrix} \\ &= \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 + (2\alpha - 1)\beta A_2^\top A_2 & \frac{1 - \alpha}{\alpha} A_2^\top \\ 0 & -A_2 & \frac{1}{\alpha\beta} I_l \end{pmatrix} = Q. \end{aligned}$$

Then the first assertion is proved. For (18), by some simple manipulations, we obtain

$$\begin{aligned} M^\top HM &= M^\top Q \\ &= \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & -\beta A_2^\top \\ 0 & 0 & I_l \end{pmatrix} \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 + (2\alpha - 1)\beta A_2^\top A_2 & \frac{1 - \alpha}{\alpha} A_2^\top \\ 0 & -A_2 & \frac{1}{\alpha\beta} I_l \end{pmatrix} \\ &= \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 + 2\alpha\beta A_2^\top A_2 & -A_2^\top \\ 0 & -A_2 & \frac{1}{\alpha\beta} I_l \end{pmatrix}. \end{aligned}$$

We now break up the proof into two cases. First, if $\alpha = 1$, then

$$(Q^\top + Q) - M^\top H M = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & \frac{1}{\beta} I_l \end{pmatrix} \succeq 0.$$

Therefore, (18) holds. Second, if $\alpha \in (1, +\infty)$, then

$$\begin{aligned} & (Q^\top + Q) - M^\top H M \\ &= \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 + (2\alpha - 2)\beta A_2^\top A_2 & \frac{1-\alpha}{\alpha} A_2^\top \\ 0 & \frac{1-\alpha}{\alpha} A_2 & \frac{1}{\alpha\beta} I_l \end{pmatrix} \\ &= \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + (2\alpha - 2) \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta A_2^\top A_2 & -\frac{1}{2\alpha} A_2^\top \\ 0 & -\frac{1}{2\alpha} A_2 & \frac{1}{\alpha\beta(2\alpha-2)} I_l \end{pmatrix}. \end{aligned} \quad (14)$$

Note that

$$\begin{aligned} & 4\alpha \begin{pmatrix} \beta A_2^\top A_2 & -\frac{1}{2\alpha} A_2^\top \\ -\frac{1}{2\alpha} A_2 & \frac{1}{\alpha\beta(2\alpha-2)} I_l \end{pmatrix} - \begin{pmatrix} 2\alpha\beta A_2^\top A_2 & -A_2^\top \\ -A_2 & \frac{1}{\alpha\beta} I_l \end{pmatrix} \\ &= \begin{pmatrix} 2\alpha\beta A_2^\top A_2 & -A_2^\top \\ -A_2 & \frac{\alpha+1}{\alpha\beta(\alpha-1)} I_l \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\beta} A_2^\top & 0 \\ 0 & \frac{1}{\sqrt{\beta}} I_l \end{pmatrix} \begin{pmatrix} 2\alpha I_l & -I_l \\ -I_l & \frac{\alpha+1}{\alpha(\alpha-1)} I_l \end{pmatrix} \begin{pmatrix} \sqrt{\beta} A_2 & 0 \\ 0 & \frac{1}{\sqrt{\beta}} I_l \end{pmatrix}. \end{aligned} \quad (15)$$

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

$$\begin{pmatrix} 2\alpha & -1 \\ -1 & \frac{\alpha+1}{\alpha(\alpha-1)} \end{pmatrix} \otimes I_l,$$

Since

$$\begin{pmatrix} 2\alpha & -1 \\ -1 & \frac{\alpha+1}{\alpha(\alpha-1)} \end{pmatrix} \succeq 0, \quad \forall \alpha > 1,$$

the right-hand side of (20) is also positive semi-definite. Thus, we have

$$\begin{pmatrix} \beta A_2^\top A_2 & -\frac{1}{2\alpha} A_2^\top \\ -\frac{1}{2\alpha} A_2 & \frac{1}{\alpha\beta(2\alpha-2)} I_l \end{pmatrix} \succeq \frac{1}{4\alpha} \begin{pmatrix} 2\alpha\beta A_2^\top A_2 & -A_2^\top \\ -A_2 & \frac{1}{\alpha\beta} I_l \end{pmatrix}. \quad (16)$$

Substituting (21) into (19) and by the expression of $M^\top H M$, we obtain (18). The lemma is proved.

3 Algorithm and convergence results

In this section, we first describe the symmetric version of the generalized alternating direction method of multipliers (SGADMM) for VI(\mathcal{W}, F, θ) formally, and then we prove its global convergence

in a contraction perspective and establish its worst-case $\mathcal{O}(1/k)$ convergence rate in both the ergodic and non-ergodic senses step by step, where k denotes the iteration counter.

3.1 Algorithm

Algorithm 3.1. SGADMM

Step 0. Choose the parameters $\alpha \in [1, +\infty)$, $\beta > 0$, $R_i \in \mathcal{R}^{n_i \times n_i} \succeq 0$ ($i = 1, 2$), the tolerance $\varepsilon > 0$ and the initial iterate $(x_1^0, x_2^0, \lambda^0) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{R}^l$. Set $k := 0$.

Step 1. Generate the new iterate $w^{k+1} = (x_1^{k+1}, x_2^{k+1}, \lambda^{k+1})$ by

$$\begin{cases} x_1^{k+1} \in \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \{ \theta_1(x_1) - x_1^\top A_1^\top \lambda^k + \frac{\alpha\beta}{2} \|A_1 x_1 + A_2 x_2^k - b\|^2 + \frac{1}{2} \|x_1 - x_1^k\|_{R_1}^2 \}, \\ x_2^{k+1} \in \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \{ \theta_2(x_2) - x_2^\top A_2^\top \lambda^k + \frac{(2\alpha-1)\beta}{2} \|A_1 x_1^{k+1} + A_2 x_2 - b\|^2 + \frac{1}{2} \|x_2 - x_2^k\|_{R_2}^2 \}, \\ \lambda^{k+1} = \lambda^k - \beta[\alpha A_1 x_1^{k+1} - (1-\alpha)(A_2 x_2^k - b) + A_2 x_2^{k+1} - b]. \end{cases} \quad (17)$$

Step 2. If

$$\max\{\|R_1 x_1^k - R_1 x_1^{k+1}\|, \|R_2 x_2^k - R_2 x_2^{k+1}\|, \|A_2 x_2^k - A_2 x_2^{k+1}\|, \|\lambda^k - \lambda^{k+1}\|\} < \varepsilon, \quad (18)$$

then stop and return an approximate solution $(x_1^{k+1}, x_2^{k+1}, \lambda^{k+1})$ of $\text{VI}(\mathcal{W}, F, \theta)$; else set $k := k + 1$, and goto Step 1.

Remark 3.1. Obviously, the iterative scheme (22) reduces to the generalized ADMM when $\alpha = 1$, and further reduces to (9) when $R_i = 0$ ($i = 1, 2$). That is to say, if the parameters $\alpha = 1$ and $R_i = 0$ ($i = 1, 2$), then the classical ADMM is recovered. Since the convergence results of the (proximal) ADMM have been established in the literature [26, 27, 29], in the following, we only consider $\alpha \in (1, +\infty)$.

3.2 Global convergence

For further analysis, we need to define an auxiliary sequence $\{\hat{w}^k\}$ as follows

$$\hat{w}^k = \begin{pmatrix} \hat{x}_1^k \\ \hat{x}_2^k \\ \hat{\lambda}^k \end{pmatrix} = \begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ \lambda^k - \alpha\beta(A_1 x_1^{k+1} + A_2 x_2^k - b) \end{pmatrix}. \quad (19)$$

Lemma 3.1. Let $\{\lambda^{k+1}\}$ and $\{\hat{\lambda}^k\}$ be the two sequences generated by SGADMM. Then

$$\lambda^{k+1} = \hat{\lambda}^k - \beta(A_2 \hat{x}_2^k - A_2 x_2^k), \quad (20)$$

and

$$\hat{\lambda}^k - \left(\frac{1}{\alpha} - 1\right)(\hat{\lambda}^k - \lambda^k) = \lambda^k - (2\alpha - 1)\beta(A_1 \hat{x}_1^k + A_2 x_2^k - b). \quad (21)$$

Proof. From the definition of λ^{k+1} , we get

$$\begin{aligned}\lambda^{k+1} &= \lambda^k - \beta[\alpha A_1 \hat{x}_1^k - (1 - \alpha)(A_2 x_2^k - b) + A_2 \hat{x}_2^k - b] \\ &= \lambda^k - \beta[\alpha(A_1 \hat{x}_1^k + A_2 x_2^k - b) + (A_2 \hat{x}_2^k - A_2 x_2^k)] \\ &= \hat{\lambda}^k - \beta(A_2 \hat{x}_2^k - A_2 x_2^k).\end{aligned}$$

Then (25) is proved. For (26), we have

$$\begin{aligned}& \hat{\lambda}^k - \left(\frac{1}{\alpha} - 1\right)(\hat{\lambda}^k - \lambda^k) \\ &= \lambda^k - \alpha\beta(A_1 \hat{x}_1^k + A_2 x_2^k - b) + \left(\frac{1}{\alpha} - 1\right)\alpha\beta(A_1 \hat{x}_1^k + A_2 x_2^k - b) \\ &= \lambda^k - (2\alpha - 1)\beta(A_1 \hat{x}_1^k + A_2 x_2^k - b).\end{aligned}$$

Therefore (26) is also right. This completes the proof. \square

Thus, based on (24) and (25), the two sequences $\{w^k\}$ and $\{\hat{w}^k\}$ satisfies the following relationship:

$$w^{k+1} = w^k - M(w^k - \hat{w}^k), \quad (22)$$

where M is defined in (15).

The following lemma shows that the stopping criterion (23) of SGADMM is reasonable.

Lemma 3.2. If $R_i x_i^k = R_i x_i^{k+1}$ ($i = 1, 2$), $A_2 x_2^k = A_2 x_2^{k+1}$ and $\lambda^k = \lambda^{k+1}$, then the iterate $\hat{w}^k = (\hat{x}_1^k, \hat{x}_2^k, \hat{\lambda}^k)$ produced by SGADMM is a solution of $\text{VI}(\mathcal{W}, F, \theta)$.

Proof. By invoking the optimality condition of the three subproblems in (4), we have the following mixed variational inequality problems: for any $w = (x_1, x_2, \lambda) \in \mathcal{W}$,

$$\begin{cases} \theta_1(x_1) - \theta_1(\hat{x}_1^k) + (x_1 - \hat{x}_1^k)^\top \left\{ -A_1^\top [\lambda^k - \alpha\beta(A_1 \hat{x}_1^k + A_2 x_2^k - b)] + R_1(\hat{x}_1^k - x_1^k) \right\} \geq 0, \\ \theta_2(x_2) - \theta_2(\hat{x}_2^k) + (x_2 - \hat{x}_2^k)^\top \left\{ -A_2^\top [\lambda^k - (2\alpha - 1)\beta(A_1 \hat{x}_1^k + A_2 \hat{x}_2^k - b)] + R_2(\hat{x}_2^k - x_2^k) \right\} \geq 0, \\ (\lambda - \hat{\lambda}^k)^\top [\alpha A_1 \hat{x}_1^k - (1 - \alpha)(A_2 x_2^k - b) + A_2 \hat{x}_2^k - b - (\lambda^k - \lambda^{k+1})/\beta] \geq 0. \end{cases}$$

Then, adding the above three inequalities and by (25), (26), we get

$$\begin{aligned}& \theta(x) - \theta(\hat{x}^k) + (w - \hat{w}^k)^\top \left\{ \begin{pmatrix} -A_1^\top \hat{\lambda}^k \\ -A_2^\top \hat{\lambda}^k \\ A_1 \hat{x}_1^k + A_2 \hat{x}_2^k - b \end{pmatrix} \right. \\ & \left. + \begin{pmatrix} R_1(\hat{x}_1^k - x_1^k) \\ (2\alpha - 1)\beta A_2^\top (A_2 \hat{x}_2^k - A_2 x_2^k) + (1 - \alpha)A_2^\top (\hat{\lambda}^k - \lambda^k)/\alpha + R_2(\hat{x}_2^k - x_2^k) \\ (1 - \alpha)(A_2 \hat{x}_2^k - A_2 x_2^k)/\alpha + (\lambda^{k+1} - \lambda^k)/(\alpha\beta) \end{pmatrix} \right\} \geq 0.\end{aligned}$$

Then by (24), we obtain

$$\theta(x) - \theta(\hat{x}^k) + (w - \hat{w}^k)^\top \left\{ F(\hat{w}^k) + \begin{pmatrix} R_1(\hat{x}_1^k - x_1^k) \\ (2\alpha - 1)\beta A_2^\top(A_2\hat{x}_2^k - A_2x_2^k) + (1 - \alpha)A_2^\top(\hat{\lambda}^k - \lambda^k)/\alpha + R_2(\hat{x}_2^k - x_2^k) \\ -(A_2\hat{x}_2^k - A_2x_2^k) + (\hat{\lambda}^k - \lambda^k)/(\alpha\beta) \end{pmatrix} \right\} \geq 0.$$

Then, by (16) (the definition of Q), the above inequality can be rewritten as

$$\theta(x) - \theta(\hat{x}^k) + (w - \hat{w}^k)^\top F(\hat{w}^k) \geq (w - \hat{w}^k)^\top Q(w^k - \hat{w}^k), \quad (23)$$

for any $w \in \mathcal{W}$. Therefore, if $R_i x_i^k = R_i x_i^{k+1}$ ($i = 1, 2$), $A_2 x_2^k = A_2 x_2^{k+1}$ and $\lambda^k = \lambda^{k+1}$, then by (25), we have $\lambda^{k+1} = \hat{\lambda}^k$. Then $\hat{\lambda}^k = \lambda^k$. Thus, we have

$$Q(w^k - \hat{w}^k) = 0,$$

which together with (28) implies that

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

This indicates that the vector \hat{w}^k is a solution of $\text{VI}(\mathcal{W}, F, \theta)$. This completes the proof. \square

Lemma 3.3. Let $\{w^k\}$ and $\{\hat{w}^k\}$ be two sequences generated by SGADMM. Then, for any $w \in \mathcal{W}$, we have

$$(w - \hat{w}^k)^\top Q(w^k - \hat{w}^k) \geq \frac{1}{2}(\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2) + \frac{\alpha - 1}{2\alpha}\|w^k - w^{k+1}\|_H^2. \quad (24)$$

Proof. Applying the identity

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

with

$$a = w, b = \hat{w}^k, c = w^k, d = w^{k+1},$$

we obtain

$$(w - \hat{w}^k)^\top H(w^k - w^{k+1}) = \frac{1}{2}(\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2) + \frac{1}{2}(\|w^k - \hat{w}^k\|_H^2 - \|w^{k+1} - \hat{w}^k\|_H^2).$$

This together with (17) and (27) implies that

$$(w - \hat{w}^k)^\top Q(w^k - \hat{w}^k) = \frac{1}{2}(\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2) + \frac{1}{2}(\|w^k - \hat{w}^k\|_H^2 - \|w^{k+1} - \hat{w}^k\|_H^2). \quad (25)$$

Now let us deal with the last term in (30), which can be written as

$$\begin{aligned}
& \|w^k - \hat{w}^k\|_H^2 - \|w^{k+1} - \hat{w}^k\|_H^2 \\
&= \|w^k - \hat{w}^k\|_H^2 - \|(w^k - \hat{w}^k) - (w^k - w^{k+1})\|_H^2 \\
&= \|w^k - \hat{w}^k\|_H^2 - \|(w^k - \hat{w}^k) - M(w^k - \hat{w}^k)\|_H^2 \text{(Using(27))} \\
&= 2(w^k - \hat{w}^k)^\top HM(w^k - \hat{w}^k) - (w^k - \hat{w}^k)^\top M^\top HM(w^k - \hat{w}^k) \\
&= (w^k - \hat{w}^k)(Q^\top + Q - M^\top HM)(w^k - \hat{w}^k) \\
&\geq \frac{\alpha - 1}{2\alpha} (w^k - \hat{w}^k)^\top M^\top HM(w^k - \hat{w}^k) \text{(Using(18))} \\
&= \frac{\alpha - 1}{2\alpha} \|w^k - w^{k+1}\|_H^2 \text{(Using(27))}.
\end{aligned}$$

Substituting the above inequality into (30), the assertion of this lemma is proved. \square

Theorem 3.1. Let $\{w^k\}$ and $\{\hat{w}^k\}$ be two sequences generated by SGADMM. Then, for any $w \in \mathcal{W}$, we have

$$\theta(x) - \theta(\hat{x}^k) + (w - \hat{w}^k)^\top F(w) \geq \frac{1}{2} (\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2) + \frac{\alpha - 1}{2\alpha} \|w^k - w^{k+1}\|_H^2. \quad (26)$$

Proof. First, combining (28) and (29), we get

$$\theta(x) - \theta(\hat{x}^k) + (w - \hat{w}^k)^\top F(\hat{w}^k) \geq \frac{1}{2} (\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2) + \frac{\alpha - 1}{2\alpha} \|w^k - w^{k+1}\|_H^2.$$

From the monotonicity of $F(\cdot)$, we have

$$(w - \hat{w}^k)^\top (F(w) - F(\hat{w}^k)) \geq 0.$$

Adding the above two inequalities, we obtain the assertion (31). The proof is completed. \square

With the above theorem in hand, we are ready to establish the global convergence of SGADMM for solving VI(\mathcal{W}, F, θ).

Theorem 3.2. Let $\{w^k\}$ be the sequence generated by SGADMM. If $\alpha > 1$, $R_i + \beta A_i^\top A_i \succ 0$ ($i = 1, 2$), then the corresponding sequence $\{w^k\}$ converges to some w^∞ , which belongs to \mathcal{W}^* .

Proof. Setting $w = w^*$ in (31), we have

$$\begin{aligned}
& \|w^k - w^*\|_H^2 - \frac{\alpha - 1}{\alpha} \|w^k - w^{k+1}\|_H^2 \\
&\geq 2\{\theta(\hat{x}^k) - \theta(x^*) + (\hat{w}^k - w^*)^\top F(w^*)\} + \|w^{k+1} - w^*\|_H^2 \\
&\geq \|w^{k+1} - w^*\|_H^2,
\end{aligned}$$

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

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \frac{\alpha - 1}{\alpha} \|w^k - w^{k+1}\|_H^2. \quad (27)$$

Summing over $k = 0, 1, \dots, \infty$, it yields

$$\sum_{k=0}^{\infty} \|w^k - w^{k+1}\|_H^2 \leq \frac{\alpha}{\alpha - 1} \|w^0 - w^*\|_H^2.$$

By $\alpha > 1$ and the positive semi-definite of H , the above inequality implies that

$$\lim_{k \rightarrow \infty} \|w^k - w^{k+1}\|_H^2 = 0.$$

Thus, by the definition of H , we have

$$\lim_{k \rightarrow \infty} \|x_1^k - x_1^{k+1}\|_{R_1}^2 = \lim_{k \rightarrow \infty} \|v^k - v^{k+1}\|_{H_1}^2 = 0, \quad (28)$$

where

$$H_1 = \begin{pmatrix} R_2 + \frac{2\alpha^2 - 2\alpha + 1}{\alpha} \beta A_2^\top A_2 & \frac{1 - \alpha}{\alpha} A_2^\top \\ \frac{1 - \alpha}{\alpha} A_2 & \frac{1}{\alpha \beta} I_l \end{pmatrix},$$

is positive definite by $R_2 + \beta A_2^\top A_2 \succ 0$. From (32) again, we have

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^0 - w^*\|_H^2,$$

which indicates that the sequence $\{Hw^k\}$ is bounded. Thus, $\{R_1 x_1^k\}_{k=0}^{\infty}$ and $\{H_1 v^k\}_{k=0}^{\infty}$ are both bounded. Then, $\{v^k\}_{k=0}^{\infty}$ is bounded. If $R_1 \succ 0$, $\{x_1^k\}_{k=0}^{\infty}$ is bounded; otherwise, $A_1^\top A_1 \succ 0$, that is A_1 is full column-rank, which together with $A_1 x_1 = (\lambda^k - \lambda^{k+1})/(\alpha\beta) + (1 - \alpha)(A_2 x_2^k - b)/\alpha - (A_2 x_2^{k+1} - b)/\alpha$ implies that $\{x_1^k\}_{k=0}^{\infty}$ is bounded. In conclusion, $\{w^k\}_{k=0}^{\infty}$ is bounded.

Then, from (33) and $H_1 \succ 0$, the sequence $\{v^k\}$ is convergent. Suppose it converges to v^∞ . Let $w^\infty = (x_1^\infty, v^\infty)$ be a cluster point of $\{w^k\}$ and $\{w^{k_j}\}$ be the corresponding subsequence. On the other hand, by (25) and (33), we have

$$\lim_{k \rightarrow \infty} R_1(x_1^k - \hat{x}_1^k) = 0, \quad \lim_{k \rightarrow \infty} (x_2^k - \hat{x}_2^k) = 0$$

and

$$\lim_{k \rightarrow \infty} (\lambda^k - \hat{\lambda}^k) = \lim_{k \rightarrow \infty} (\lambda^k - \lambda^{k+1} + \beta(A_2 \hat{x}_2^k - A_2 x_2^k)) = 0.$$

Thus,

$$\lim_{k \rightarrow \infty} Q(w^k - \hat{w}^k) = 0. \quad (29)$$

Then, taking limit along the subsequence $\{w^{k_j}\}$ in (28) and using (34), for any $w \in \mathcal{W}$, we obtain

$$\theta(x) - \theta(x^\infty) + (w - w^\infty)^\top F(w^\infty) \geq 0,$$

which indicates that w^∞ is a solution of $\text{VI}(\mathcal{W}, F, \theta)$. Then, since w^* in (32) is arbitrary, we can set $w^* = w^\infty$ and conclude that the whole generated sequence $\{w^k\}$ converges by $R_i + \beta A_i^\top A_i \succ 0 (i = 1, 2)$. This completes the proof. \square

3.3 Convergence rate

Now, we are going to prove the worst-case $\mathcal{O}(1/t)$ convergence rate of SGADMM in both the ergodic and non-ergodic senses.

Theorem 3.3. Let $\{w^k\}$ and $\{\hat{w}^k\}$ be the sequences generated by SGADMM, and set

$$\bar{w}^t = \frac{1}{t+1} \sum_{k=0}^t \hat{w}^k.$$

Then, for any integer $t \geq 0$, we have $\bar{w}^t \in \mathcal{W}$, and

$$\theta(\bar{x}_t) - \theta(x) + (\bar{w}_t - w)^\top F(w) \leq \frac{1}{2(t+1)} \|w - w^0\|_H^2, \quad \forall w \in \mathcal{W}. \quad (30)$$

Proof. From (22) and the convexity of the set \mathcal{W} , we have $\bar{w}^k \in \mathcal{W}$. From (31), we have

$$\theta(x) - \theta(\hat{x}^k) + (w - \hat{w}^k)^\top F(w) + \frac{1}{2} \|w - w^k\|_H^2 \geq \frac{1}{2} \|w - w^{k+1}\|_H^2, \quad \forall w \in \mathcal{W}.$$

Summing the above inequality over $k = 0, 1, \dots, t$, we get

$$(t+1)\theta(x) - \sum_{k=0}^t \theta(\hat{x}^k) + \left((t+1)w - \sum_{k=0}^t \hat{w}^k \right)^\top F(w) + \frac{1}{2} \|w - w^0\|_H^2 \geq 0, \quad \forall w \in \mathcal{W}.$$

By the definition of \bar{w}^t and the convexity of $\theta(\cdot)$, the assertion (35) follows immediately from the above inequality. This completes the proof. \square

The proof of the next two lemmas is referred to those of Lemmas 5.1 and 5.2 in [30]. For completeness, we give the detail proof.

Lemma 3.4. Let $\{w^k\}$ be the sequence generated by SGADMM. Then, we have

$$(w^k - w^{k+1})^\top H \{(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})\} \geq \frac{3\alpha - 1}{4\alpha} \|(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})\|_H^2. \quad (31)$$

Proof. Setting $w = \hat{w}^{k+1}$ in (28), we have

$$\theta(\hat{x}^{k+1}) - \theta(\hat{x}^k) + (\hat{w}^{k+1} - \hat{w}^k)^\top F(\hat{w}^k) \geq (\hat{w}^{k+1} - \hat{w}^k)^\top Q(w^k - \hat{w}^k).$$

Similarly setting $w = \hat{w}^k$ in (28) for $k := k+1$, we get

$$\theta(\hat{x}^k) - \theta(\hat{x}^{k+1}) + (\hat{w}^k - \hat{w}^{k+1})^\top F(\hat{w}^{k+1}) \geq (\hat{w}^k - \hat{w}^{k+1})^\top Q(w^{k+1} - \hat{w}^{k+1}).$$

Then, adding the above two inequalities and using the monotonicity of the mapping $F(\cdot)$, we get

$$(\hat{w}^k - \hat{w}^{k+1})^\top Q \{(w^k - \hat{w}^k) - (w^{k+1} - \hat{w}^{k+1})\} \geq 0. \quad (32)$$

By (37), we have

$$\begin{aligned} & (w^k - w^{k+1})^\top Q \{(w^k - \hat{w}^k) - (w^{k+1} - \hat{w}^{k+1})\} \\ &= \{(w^k - \hat{w}^k) - (w^{k+1} - \hat{w}^{k+1}) + (\hat{w}^k - \hat{w}^{k+1})\}^\top Q \{(w^k - \hat{w}^k) - (w^{k+1} - \hat{w}^{k+1})\} \\ &= \|(w^k - \hat{w}^k) - (w^{k+1} - \hat{w}^{k+1})\|_Q^2 + (\hat{w}^k - \hat{w}^{k+1})^\top Q \{(w^k - \hat{w}^k) - (w^{k+1} - \hat{w}^{k+1})\} \\ &\geq \|(w^k - \hat{w}^k) - (w^{k+1} - \hat{w}^{k+1})\|_Q^2. \end{aligned}$$

Using (18), (27) and $Q = HM$ on both side of the above inequality, we get

$$\begin{aligned}
& (w^k - w^{k+1})^\top H \{(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})\} \\
= & (w^k - w^{k+1})^\top Q M^{-1} \{(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})\} \\
= & (w^k - w^{k+1})^\top Q \{(w^k - \hat{w}^k) - (w^{k+1} - \hat{w}^{k+1})\} \\
\geq & \|(w^k - \hat{w}^k) - (w^{k+1} - \hat{w}^{k+1})\|_Q^2 \\
= & [(w^k - \hat{w}^k) - (w^{k+1} - \hat{w}^{k+1})]^\top Q [(w^k - \hat{w}^k) - (w^{k+1} - \hat{w}^{k+1})] \\
= & [(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})]^\top M^{-1} Q M^{-1} [(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})] \\
\geq & \frac{3\alpha - 1}{4\alpha} [(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})]^\top M^{-1} M H M M^{-1} [(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})] \\
= & \frac{3\alpha - 1}{4\alpha} \|(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})\|_H^2.
\end{aligned}$$

Then, we get the assertion (36). The proof is completed. \square

Lemma 3.5. Let $\{w^k\}$ be the sequence generated by SGADMM. Then, we have

$$\|w^{k+1} - w^{k+2}\|_H^2 \leq \|w^k - w^{k+1}\|_H^2 - \frac{\alpha - 1}{2\alpha} \|(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})\|_H^2. \quad (33)$$

Proof. Setting $a := (w^k - w^{k+1})$ and $b := (w^{k+1} - w^{k+2})$ in the identity

$$\|a\|_H^2 - \|b\|_H^2 = 2a^\top H(a - b) - \|a - b\|_H^2,$$

we can derive

$$\begin{aligned}
& \|w^k - w^{k+1}\|_H^2 - \|w^{k+1} - w^{k+2}\|_H^2 \\
= & 2(w^k - w^{k+1})^\top H \{(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})\} - \|(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})\|_H^2 \\
\geq & \frac{3\alpha - 1}{2\alpha} \|(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})\|_H^2 - \|(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})\|_H^2 \\
= & \frac{\alpha - 1}{2\alpha} \|(w^k - w^{k+1}) - (w^{k+1} - w^{k+2})\|_H^2,
\end{aligned}$$

which completes the proof of the lemma. \square

Based on Lemma 3.5, now we establish the worst-case $\mathcal{O}(1/t)$ convergence rate of SGADMM in a non-ergodic sense.

Theorem 3.4. Let $\{w^k\}$ be the sequence generated by SGADMM. Then, for any $w^* \in \mathcal{W}^*$ and integer $t \geq 0$, we have

$$\|w^t - w^{t+1}\|_H^2 \leq \frac{\alpha}{(t+1)(\alpha-1)} \|w^0 - w^*\|_H^2. \quad (34)$$

Proof. By (32), we get

$$\frac{\alpha - 1}{\alpha} \sum_{k=0}^t \|w^k - w^{k+1}\|_H^2 \leq \|w^0 - w^*\|_H^2.$$

This and (38) imply that

$$\frac{(t+1)(\alpha-1)}{\alpha} \|w^t - w^{t+1}\|_H^2 \leq \|w^0 - w^*\|_H^2.$$

Therefore, the assertion of this theorem comes from the above inequality immediately. The proof is completed. \square

Remark 3.2. From (39), we have that as the larger α , the smaller $\frac{\alpha}{\alpha-1}$, which controls the upper bounds of $\|w^t - w^{t+1}\|_H^2$. Therefore, it seems that larger values of α is more beneficial for speeding up the convergence of SGADMM.

4 Numerical experiments

In this section, we present some numerical experiments to verify the efficiency of SGADMM for solving compressed sensing. Those numerical experiments are performed in Matlab R2010a on a ThinkPad computer equipped with Windows XP, 997MHz and 2 GB of memory.

Compressed sensing (CS) is to recover a sparse signal $\bar{x} \in \mathcal{R}^n$ from an undetermined linear system $b = A\bar{x}$, where $A \in \mathcal{R}^{m \times n}$ ($m \ll n$), can be depicted as problem (2).

Obviously, Problem (2) is equivalent to the following two models:

(a) Model 1: Problem (3).

(b) Model 2:

$$\begin{aligned} \min \mu \|x_1\|_1 + \frac{1}{2} \|Ax_2 - y\|^2 \\ \text{s.t. } -x_1 - x_2 = 0, \\ x_1 \in \mathcal{R}^n, x_2 \in \mathcal{R}^n. \end{aligned} \tag{35}$$

4.1 The iterative schemes for (3) and (40)

Since (3) and (40) are both some concrete models of (1), SGADMM are applicable to them. Below, we elaborate on how to derive the closed-form solutions for the subproblems of resulted by SGADMM.

For problem (3), its first two subproblems resulted by SGADMM is as follows.

- With the given x_2^k and λ^k , the x_1 -subproblem in (22) is (Here $R_1 = 0$)

$$x_1^{k+1} = \operatorname{argmin}_{x_1 \in \mathcal{R}^n} \left\{ \frac{1}{2} \|x_1\|_2^2 + x_1^\top \lambda + \frac{\alpha\beta}{2} \|x_1 - Ax_2^k + y\|^2 \right\},$$

which has the following closed-form solution:

$$x_1^{k+1} = \frac{1}{1 + \alpha\beta} (\alpha\beta(Ax_2^k - y) - \lambda^k).$$

- With the updated x_1^{k+1} , the x_2 -subproblem in (22) is (Here $R_2 = \tau I_n - (2\alpha - 1)\beta A^\top A$ with $\tau \geq (2\alpha - 1)\beta \|A^\top A\|$)

$$x_2^{k+1} = \operatorname{argmin}_{x_2 \in \mathcal{R}^n} \left\{ \mu \|x_2\|_1 - x_2^\top A^\top \lambda^k + \frac{(2\alpha - 1)\beta}{2} \| -x_1^{k+1} + Ax_2 - y \|^2 + \frac{1}{2} \|x_2 - x_2^k\|_{R_2}^2 \right\},$$

and its closed-form solution is given by

$$x_2^{k+1} = \operatorname{shrink}_{\frac{\mu}{\tau}} \left((2\alpha - 1)\beta A^\top (x_1^{k+1} + y) / \tau + (\tau I_n - (2\alpha - 1)\beta A^\top A) x_2^k / \tau + A^\top \lambda^k / \tau \right),$$

where, for any $c > 0$, $\operatorname{shrink}_c(\cdot)$ is defined as

$$\operatorname{shrink}_c(g) := g - \min\{c, |g|\} \frac{g}{|g|}, \quad \forall g \in \mathcal{R}^n,$$

and $(g/|g|)_i$ should be taken 0 if $|g|_i = 0$.

Similarly, for problem (40), its first two subproblems resulted by SGADMM is as follows.

- With the given x_2^k and λ^k , the x_1 -subproblem in (22) is (Here $R_1 = 0$)

$$x_1^{k+1} = \operatorname{argmin}_{x_1 \in \mathcal{R}^n} \left\{ \mu \|x_1\|_1 + \frac{\alpha\beta}{2} \|x_1 - (x_2^k + \frac{1}{\alpha\beta}\lambda^k)\|^2 \right\},$$

and its closed-form solution is given by

$$x_1^{k+1} = \operatorname{shrink}_{\frac{\mu}{\alpha\beta}} \left(x_2^k + \frac{1}{\alpha\beta}\lambda^k \right).$$

- With the updated x_1^{k+1} , the x_2 -subproblem in (22) is (Here $R_2 = \tau I_n - A^\top A$ with $\tau \geq \|A^\top A\|$)

$$x_2^{k+1} = \operatorname{argmin}_{x_2 \in \mathcal{R}^n} \left\{ \frac{1}{2} \|Ax_2 - y\|^2 + x_2^\top \lambda^k + \frac{(2\alpha - 1)\beta}{2} \|x_2 - x_1^{k+1}\|^2 + \frac{1}{2} \|x_2 - x_2^k\|_{R_2}^2 \right\},$$

and its closed-form solution is given by

$$x_2^{k+1} = \frac{1}{\tau + (2\alpha - 1)\beta} (A^\top y - \lambda^k + (2\alpha - 1)\beta x_1^{k+1} + A^\top A x_2^k).$$

Obviously, the above two iterative schemes both need to compute $A^\top A$ and $A^\top y$, which is quite time consuming if n is large. However, noting that these two terms are invariant during the iteration process, therefore we need only compute them once before all iterations.

Regarding the penalty parameter β and the constant α in SGADMM, any $\beta > 0$ and $\alpha \geq 1$ can ensure the convergence of SGADMM in theory. There are two traditional method to determine them in practice. One is the tentative method, which is easy to execute. The other is the self-adaptive adjustment method, which needs much computation. In this experiment, for β and α , we use the tentative method to determine their suitable values. For β , Xiao et al. [34] set $\beta = \operatorname{mean}(\operatorname{abs}(y))$ for ADMM. Motivated by this choice, we set $\beta = \operatorname{mean}(\operatorname{abs}(y))/(2\alpha - 1)$ in our algorithm. As for

the parameter α , we have pointed out in Remark 3.2 that larger values of α maybe beneficial for our algorithm. Here, we use (3) to do a little experiment to test this Remark. We choose different values of α in the interval $[1, 2]$. Specifically, we choose $\alpha \in \{1.0, 1.1, \dots, 2\}$. Other data about this experiment are: the proximal parameter τ is set as $\tau = 1.01(2\alpha - 1)\beta\|A^\top A\|$; the observed signal y is set as $y = Ax + 0.01 \times \text{randn}(m, 1)$ in Matlab; the sensing matrix A and the original signal x are generated by

$$\bar{A} = \text{randn}(m, n), [Q, R] = \text{qr}(\bar{A}', 0), A = Q',$$

and

$$x = \text{zeros}(n, 1); p = \text{randperm}(n); x(p(1 : k)) = \text{randn}(k, 1).$$

Then, the observed signal y is further set as $(R^\top)^{-1}y$. The initial points are set as $x_2^0 = A^\top y, \lambda^0 = Ax_2^0$. In addition, we set the regularization parameter $\mu = 0.01$, and the dimensions of the problem are set as $n = 1000, m = 300, k = 60$, where k denotes the number of the non-zeros in the original signal x . To evaluate the quality of the recovered signal, let us define the quantity ‘RelErr’ as follows:

$$\text{RelErr} = \frac{\|\tilde{x} - x\|}{\|x\|},$$

where \tilde{x} denotes the recovered signal. The stopping criterion is

$$\frac{\|f_k - f_{k-1}\|}{\|f_{k-1}\|} < 10^{-5},$$

where f_k denotes the function value of (2) at the iterate x_k .

4.2 Numerical results

The numerical results are graphically shown in Figure 1. Clearly, numerical results in Figure 1 indicate that Remark 3.2 is reasonable. Both CPU time and number of iterations are descent with respective to α . Then, in the following, we set $\alpha = 1.4$, which is a moderate choice for α .

Now, let us graphically show the recovered results of SGADMM for (3) and (40). The proximal parameter τ is set as $\tau = 1.01(2\alpha - 1)\beta\|A^\top A\|$ for (3), and $\tau = 1.01\|A^\top A\|$ for (40). The initial points are set as $x_2^0 = A^\top y, \lambda^0 = Ax_2^0$ for (3), and $x_2^0 = A^\top y, \lambda^0 = x_2^0$ for (40). Other parameters are set the same as above. Figure 2 reports the numerical results of SGADMM for (3) and (40).

The bottom two subplots in Figure 2 indicate that our new method SGADMM can be used to solve (3) and (40).

In the following, we do some numerical comparisons to illustrate the advantage of our new method and to analyze which one is more suitable to compressed sensing (2) between the two models (3) and (40). SGADMM for (3) is denoted by SGADMM1, SGADMM for (40) is denoted by SGADMM2. We also compare SGADMM with the classical ADMM. The numerical results are listed in Table 1, where ‘Time’ denotes the CPU time (in seconds), and ‘Iter’ denotes the number of iterations

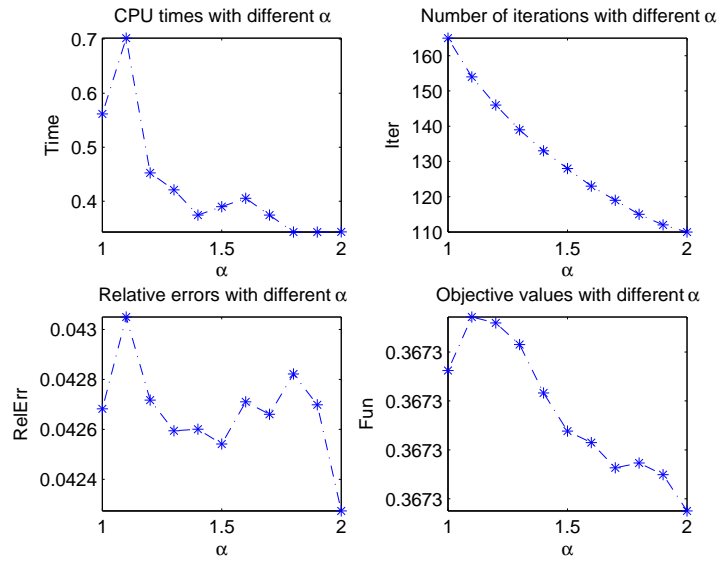


Figure 1: Sensitivity test on the parameter α

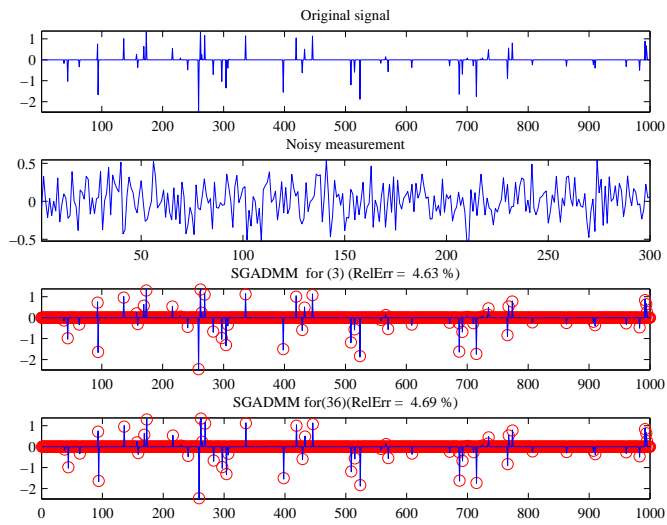


Figure 2: The top: the original signal; The second: the noisy measurement; The bottom two: recovered signal.

required for the whole recovering process, $m = \text{floor}(\gamma n), k = \text{floor}(\sigma m)$. The numerical results are the average of the numerical results of ten runs with different combinations of γ and σ .

Table 1: Comparison of SGADMM1, SGADMM2 and ADMM

n	γ	σ	SGADMM1			SGADMM2			ADMM		
			Time	Iter	RelErr	Time	Iter	RelErr	Time	Iter	RelErr
1000	0.3	0.2	0.6911	92.4	0.0387	0.9578	266.0	0.0394	0.9812	264.0	0.0393
	0.2	0.2	0.6661	118.6	0.0825	1.3915	421.8	0.0915	1.3603	419.6	0.0915
	0.2	0.1	0.5008	85.3	0.0609	0.4758	139.0	0.0579	0.5008	138.0	0.0539
2000	0.3	0.2	2.1965	90.0	0.0437	3.6535	267.7	0.0447	3.5412	265.6	0.0447
	0.2	0.2	2.2339	109.6	0.0785	5.2182	431.4	0.0874	5.1543	429.0	0.0874
	0.2	0.1	1.5428	79.9	0.0534	1.7893	142.8	0.0467	1.7613	140.8	0.0513

4.3 Discussion

The numerical results in Table 1 indicate that: (1) From the criterion ‘RelErr’, all methods successfully solved all the cases; (2) From the criteria ‘Time’ and ‘Iter’, SGADMM1 performs better than the other two methods. Especially the number of iterations of SGADMM1 is about at most two-thirds of the other two methods. This experiment also indicate that the model (3) is also an effective model for compressed sensing, and sometimes it is more efficient than the model (40), though they are equivalent in theory. In conclusion, by choosing some relaxation factor $\alpha \in [1, +\infty)$, SGADMM maybe more efficient than the classical ADMM.

5 Conclusions

In this paper, we have proposed a symmetric version of the generalized ADMM (SGADMM), which generalizes the feasible set of the relaxation factor α from $(0, 2)$ to $[1, +\infty)$. Under the same conditions, we have proved the convergence results of the new method. Some numerical results illustrate that it may perform better than the classical ADMM. In the future, we shall study SGADMM with $\alpha \in (0, 1)$ to perfect the theoretical system.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

The first author has proved the convergence results; the second author has accomplished the numerical experiment; and the third author has proposed the motivation of the manuscript. All authors read and approved the final manuscript.

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. 11601475,71532015), the foundation of First Class Discipline of Zhejiang-A(Zhejiang University of Finance and Economics-Statistics).

References

- [1] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.*, 2, (2009), pp. 183-202.
- [2] J.F. Yang, Y. Zhang, W.T. Yin, A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data, *IEEE J. Selected Topics in Signal Process.*, 4, (2010), pp. 288-297.
- [3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Mach. Learning*, 3, (2011), pp. 1-122.
- [4] R. Glowinski, A. Marrocco, Sur l'approximation, par éléments fins d'ordre n , et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires, *Rev. Fr. Autom. Inform. Rech. Oper.* 9, (1975), pp. 41-76.
- [5] D. Gabay, B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite-element approximations, *Comput. Math. Appl.* 2, (1976), pp. 17-40.
- [6] B.S. He, L.Z. Liao, D. Han, and H. Yang, A new inexact alternating directions method for monotone variational inequalities, *Math. Program.*, 92, (2002), pp. 103-118.
- [7] P.L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.* 16, (1979), pp. 964-979.

- [8] D. Gabay, Applications of the method of multipliers to variational inequalities, in *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, M. Fortin and R. Glowinski, eds., North-Holland, Amsterdam, (1988), pp. 299-331.
- [9] J. Eckstein, D.P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Program.*, 55, (1992), pp. 293-318.
- [10] E. Esser, Applications of Lagrangian-based alternating direction methods and connections to split Bregman. TR. 09-31, CAM, UCLA (2009).
- [11] E.G. Golshtein, N.V. Tretyakov, Modified Lagrangian in convex programming and their generalizations, *Math. Program. Study*, 10, (1979), pp. 86-97.
- [12] B.S. He, F. Ma, X.M. Yuan, Convergence Study on the Symmetric Version of ADMM with Larger Step Sizes, *SIAM J. Imaging Sci.*, 9(3), (2016), pp. 1467-1501.
- [13] E.X. Fang, B.S. He, H. Liu, X.M. Yuan, Generalized alternating direction method of multipliers: new theoretical insights and applications, *Mathematical Programming Computation*, 7(2), (2015), pp. 149-187.
- [14] M. Sun, J. Liu, A proximal Peaceman-Rachford splitting method for compressive sensing, *Journal of Applied Mathematics and Computing*, 50, (2016), pp. 349-363.
- [15] 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.
- [16] H.J. He, X.J. Cai, D.R. Han, A fast splitting method tailored for Dantzig selectors, *Comput. Optim. Appl.*, 62, 2015, pp. 347-372.
- [17] H.J. He, D.R. Han, A distributed Douglas-Rachford splitting method for multi-block convex minimization problems, *Adv. Comput. Math.*, 42, (2016), pp. 27-53.
- [18] C.H. Chen, S.Q. Ma, J.F. Yuan, A general inertial proximal point algorithm for mixed variational inequality problem, *SIAM J. Optim.*, 25(4), (2015), pp. 2120-2142.
- [19] C.H. Chen, R.H. Chan, S.Q. Ma, J.F. Yuan, Inertial proximal ADMM for linearly constrained separable convex optimization, *SIAM J. Imaging Sci.*, 8(4), (2015), pp. 2239-2267.
- [20] H.C. Sun, M. Sun, H.C. Zhou, A proximal splitting method for separable convex programming and its application to compressive sensing, *J. Nonlinear Sci. Appl.* 9, (2016), pp. 392-403.
- [21] 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(1), (2017), pp. 77-94.

- [22] H.C. Sun, M. Sun, H.C. Zhou, A new generalized alternating direction method of multiplier for separable convex programming and its applications to image deblurring with wavelets, *ICIC Express Letters*, 10(2), (2016), pp. 271-277.
- [23] R.T. Rockafellar, *Convex Analysis*. Princeton: Princeton University Press, 1970.
- [24] J.F. Bonnans, A. Shapiro, *Perturbation Analysis of Optimization Problems*. Springer-Verlag, 2000.
- [25] M. Fukushima, *Fundamentals of Nonlinear Optimization*. Asakura Shoten, Tokyo, 2001.
- [26] B.S. He, X.M. Yuan, On the $O(1/n)$ Convergence rate of the DouglasCRachford alternating direction method, *SIAM J. Numer. Anal.*, 50(2), (2012), pp. 700-709.
- [27] B.S. He, X.M. Yuan, On non-ergodic convergence rate of DouglasCRachford alternating direction method of multipliers, *Numer. Math.* 130, (2015), pp. 567-577.
- [28] B.S. He, F. Ma, X.M. Yuan, Linearized alternating direction method of multipliers via positive-indefinite proximal regularization for convex programming, 2016, manuscript.
- [29] M. Sun, J. Liu, The convergence rate of the proximal alternating direction method of multipliers with indefinite proximal regularization, *Journal of Inequalities and Applications*, December 2017, 2017: 19.
- [30] B.S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman-Rachford splitting method for convex programming, *SIAM J. Optim*, 24(3), (2014), pp. 1011-1040.
- [31] B.S. He, M. Tao, X.M. Yuan, A splitting method for separable convex programming, *IMA J. Numer. Anal.*, 35(1), (2014), pp. 394-426.
- [32] L.S. Hou, H.J. He, J.F. Yang, A partially parallel splitting method for multiple-block separable convex programming with applications to robust PCA, *Comput. Optim. Appl.*, 63(1), (2016), pp. 273-303.
- [33] K. Wang, J. Desai, H.J. He, A proximal partially-parallel splitting method for separable convex programs, *Optim. Method Softw.*, 32(1), (2017), pp. 39-68.
- [34] Y.H. Xiao, H. Zhu, S.Y. Wu, Primal and dual alternating direction algorithms for ℓ_1 - ℓ_1 -norm minimization problems in compressive sensing, *Comput. Optim. Appl.*, 54, (2013), pp. 441-459.
- [35] Y.Y. Xu, Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming, 2016, manuscript.
- [36] K. Guo, D.R. Han, David Z.W. Wang, T.T. Wu, Convergence of ADMM for multi-block non-convex separable optimization models, *Frontiers of Mathematics in China*, 2017, On line.