Symmetric ADMM with Positive-Indefinite Proximal Regularization for Linearly Constrained Convex Optimization

Bin Gao[1] **Feng Ma**[2]

September 12, 2016

Abstract: The proximal ADMM which adds proximal regularizations to ADMM's subproblems is a popular and useful method for linearly constrained separable convex problems, especially its linearized case. A well-known requirement on guaranteeing the convergence of the method in the literature is that the proximal regularization must be positive semidefinite. Recently it was shown by He et al. (Optimization Online, 2016) that the proximal term regularized on ADMM is not necessarily positive semidefinite without any additional assumptions, while it is still unknown whether the indefinite setting is valid for the proximal version of the symmetric ADMM. In this paper, we confirm that the symmetric ADMM can also be regularized with positive-indefinite proximal term. Theoretically, we prove global convergence of the improved method and establish the worst-case nonasymptotic $\mathcal{O}(1/t)$ convergence rate result in ergodic sense, where t counts the iteration. In addition, the generalized ADMM proposed by Eckstein and Bertsekas is a special case of our discussion. Finally, we demonstrate the improvements of using the positive-indefinite proximal term by some experimental results.

Key words: Convex programming, alternating direction method of multipliers, positive indefinite proximal, iteration-complexity

1 Introduction

In this paper, we consider the following linearly constrained separable convex optimization problem

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}, \tag{1.1}$$

where $\mathcal{X} \subseteq \Re^{n_1}$ and $\mathcal{Y} \subseteq \Re^{n_2}$ are closed convex sets; $A \in \Re^{m\times n_1}$, $B \in \Re^{m\times n_2}$ and $b \in \Re^m$; $\theta_1 : \Re^{n_1} \to \Re$ and $\theta_2 : \Re^{n_2} \to \Re$ are closed proper convex functions (not necessarily smooth). Throughout, the solution set of (1.1) is assumed to be nonempty.

Let the augmented Lagrangian function of problem (1.1) be

$$\mathcal{L}_\beta(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2, \tag{1.2}$$

where $\lambda \in \Re^m$ is the Lagrange multiplier and $\beta > 0$ is a penalty parameter. A benchmark method for solving (1.1) is the alternating direction method of multipliers (ADMM) [4, 12]. At each iteration,

[1]College of Communications Engineering, PLA University of Science and Technology, Nanjing 210007, China. Email: feimaxiao123@gmail.com

[2]Corresponding author. High-Tech Institute of Xi'an, Xi'an, 710025, China. Email: mafengnju@gmail.com. This author was supported by NSFC Grant 71401176.

ADMM performs the following updates

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & \text{(1.3a)} \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, & \text{(1.3b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(1.3c)} \end{cases}$$

The ADMM updates the variables alternately by decomposing the augmented Lagrangian function (1.2) with respect to the variables in a Gauss-Seidel order, and the resulting subproblems are usually easy enough if each θ_i has any specific properties. With this advantageous feature, the ADMM has recently gained tremendous popularity in various domains such as machine learning, image processing, computer vision, signal processing, and so on. Interested readers are referred to [2, 10, 13, 14] for recent survey papers. Specifically, the convergence rate of the ADMM has been studied extensively, see e.g., [19, 20, 7]. Although ADMM is suitable for many distributed instances that formulated as (1.1), it can't directly extend to the multi-block cases (see [6] for a counterexample). We refer to [21, 15, 3, 24] for the studies of the convergence properties for the multi-block ADMM.

The ADMM (1.3) works well in situations where the subproblem regards of θ_i is simple. If one θ_i is generic without any specific properties, the resulting subproblem may not easy to evaluate, hence need to be solved iteratively. To avoid inner loops, a well known technology is to linearize the quadratic term in the subproblem by adding a proximal term, and solve the resulting subproblem instead. In this case, the proximal ADMM is as follows:

$$\text{(PD-ADMM)} \quad \begin{cases} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\big\}, & \text{(1.4a)} \\ y^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_D^2 \mid y \in \mathcal{Y}\big\}, & \text{(1.4b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(1.4c)} \end{cases}$$

where

$$D = rI_{n_2} - \beta B^T B \quad \text{and} \quad r > \beta\|B^T B\|. \tag{1.4d}$$

To see the linearization, recall (1.2) and using simple manipulation for (1.4b), the subproblem amounts to solving the following surrogate

$$y^{k+1} = \arg\min\big\{\theta_2(y) + \frac{r}{2}\|y - (y^k + \frac{1}{\tau r}q_k)\|^2 \mid y \in \mathcal{Y}\big\}, \tag{1.5a}$$

where

$$q_k = B^T[\lambda^k - \beta(Ax^{k+1} + By^k - b)]. \tag{1.5b}$$

With this change, updating y^k reduces to evaluating the resolvent operator of $\partial\theta(y)$ which is much cheaper than (1.3b). For example, when $\theta(y) = \|y\|_1$, the resulting subproblem admits a closed form solution given by the soft shrinkage operator; when $\theta(Y) = \|Y\|_*$ is the nuclear norm of matrix Y, and the quadratic penalty term is Frobenius norm correspondingly, the closed form solution can be given by the matrix shrinkage operator. The reader may consult e.g., [30, 31, 11, 25] for extensive applications of the linerization technology.

Convergence of the proximal ADMM (1.4) has traditionally been proved under the condition that D is positive semidefinite. This is mainly because that when D is indefinite, it may suffer great difficulty to establish the contraction property. On the other hand, equipping (1.4) with positive definite matrices D is easy to prove the convergence, however, is not beneficial to get an improved numerical performance. This can be intuitively understood as that the positive regularization term forces the new point y^{k+1} to be close enough to y^k, and makes the primal step go smaller. So, it is desirable to ask if it is possible to regularize ADMM with positive-indefinite proximal term, while the convergence can be still ensured? To the best of our knowledge, only [22] studied this question under the assumption that $\theta_i(\cdot)$ is Lipschitz continuous. Recently, He et al [17] answered this question affirmatively. They proved that it is not necessary to ensure the positive definiteness of the D without any additional assumptions, and obtained the improved positive-indefinite ADMM.

$$
\text{(PID-ADMM)} \quad \begin{cases} x^{k+1} = \text{argmin}\left\{\mathcal{L}_\beta(x, y^k, \lambda^k) | x \in \mathcal{X}\right\}, & (1.6a) \\ y^{k+1} = \text{argmin}\left\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|^2_{D_0} | y \in \mathcal{Y}\right\}, & (1.6b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), & (1.6c) \end{cases}
$$

where

$$
D_0 = \tau r I_{n_2} - \beta B^T B \quad \text{with} \quad r > \beta\|B^T B\|, \quad \tau \in [0.8, 1). \tag{1.6d}
$$

In this case, the y-subproblem (1.6b) can be written as

$$
y^{k+1} = \arg\min\left\{\theta_2(y) + \frac{\tau r}{2}\|y - (y^k + \frac{1}{\tau r} q_k)\|^2 \mid y \in \mathcal{Y}\right\}. \tag{1.7}
$$

where q_k is defined in (1.5b). Obviously, the iterative step (1.7) only needs to attach a factor τ in the linearized subproblem (1.5a), and thus enjoy the same easiness in implementation as the scheme (1.5a). Compared with the fixed $\tau = 1$ in PD-ADMM (1.4), the small values of $\tau \in [0.8, 1)$ can ensure the proximal term has less weight for the y-subproblem (1.6b), and thus allows for larger steps. In practice, it is recommended to choose 0.8 as an "optimal" choice to obtain the fastest convergence [17].

In addition to the original ADMM (1.3), the symmetric version of ADMM which updates Lagrange multiplier twice is an important ADMM variant. We restrict our discussion to a particular symmetric ADMM:

$$
\text{(SADMM)} \quad \begin{cases} x^{k+1} = \text{argmin}\left\{\mathcal{L}_\beta(x, y^k, \lambda^k) | x \in \mathcal{X}\right\}, & (1.8a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k - b), & (1.8b) \\ y^{k+1} = \text{argmin}\left\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) | y \in \mathcal{Y}\right\}, & (1.8c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b), & (1.8d) \end{cases}
$$

where $\alpha \in (-1, 1)$ is an underdetermined relaxation factor. SADMM can be explained as an application of shrunken Peaceman–Rachford splitting method in [26, 27] to the dual of (1.1). In [18], it was proved that the sequence generated by (1.8) is strictly contractive with respect to the solution set of (1.1); thus, the global convergence and a worst-case convergence rate measured by the iteration complexity were established for (1.8) therein. We refer to [18, 16] for the more general symmetric ADMM whose dual step sizes are chosen in a domain. Inspired by He et al' work for ADMM, it is interesting to ask a question: can we also regularize the SADMM (1.8) with positive-indefinite proximal term? In this paper we answer this question affirmatively, and thus propose the linearized

SADMM with positive-indefinite proximal regularization for solving problem (1.1):

$$
\text{(PID-SADMM)} \quad \begin{cases} x^{k+1} = \text{argmin}\left\{\mathcal{L}_\beta(x, y^k, \lambda^k) | x \in \mathcal{X}\right\}, & (1.9a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k - b), & (1.9b) \\ y^{k+1} = \text{argmin}\left\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) + \frac{1}{2}\|y - y^k\|_{D_0}^2 | y \in \mathcal{Y}\right\}, & (1.9c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b). & (1.9d) \end{cases}
$$

where

$$
D_0 = \tau r - \beta B^{\mathrm{T}} B \quad \text{with} \quad r > \beta\|B^T B\| \quad \text{and} \quad \tau \in [\frac{\alpha^2 - \alpha + 4}{\alpha^2 - 2\alpha + 5}, 1). \tag{1.9e}
$$

Obviously, the PID-SADMM (1.9) includes the PID-ADMM (1.6) as a special case with $\alpha \equiv 0$. For a fixed α, we see that the PID-SADMM only requires one to choose suitable proximal parameters without using any additional computation. Due to the smaller size of the proximal parameter, the proximal term has a less weight in the objective function. It is thus possible to enlarge the step size of the y, and accelerate the numerical performance. We prove the convergence of the proposed algorithm and show that the convergence rate of the algorithm is $\mathcal{O}(1/t)$, where t is the iteration counter. To verify the improvement, we test our algorithm on two well-studied numerical problems. The computational results show that PID-SADMM works reasonably well on large-scale instances and usually outperforms its positive part schemes.

The rest of the paper is organized as follows. In Sect. 2, we summarize some useful results for further analysis. Then, in Sect. 3, we give a prediction-correction interpretation of the algorithm. The convergence and complexity results are proved in Sect. 4 and 5, respectively. In Sect. 6, we test the algorithm and report some preliminary numerical results. Concluding remarks are drawn in Sect. 7.

2 Preliminaries

2.1 Variational inequality reformulation of (1.1)

In this section, we reformulate the convex minimization model (1.1) as a variational inequality (VI) form. This reformulation will be useful for succedent algorithmic illustration and convergence analysis.

The Lagrangian function of (1.1) can be written as

$$
\mathcal{L}(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^{\mathrm{T}}(Ax + By - b), \tag{2.1}
$$

where $\lambda \in \Re^m$ is a Lagrangian multiplier. Seeking a saddle point of $L(x, y, \lambda)$ is to finding (x^*, y^*, λ^*) such that

$$
\mathcal{L}_{\lambda \in \Re^m}(x^*, y^*, \lambda) \le \mathcal{L}(x^*, y^*, \lambda^*) \le \mathcal{L}_{x \in \mathcal{X}, y \in \mathcal{Y}}(x, y, \lambda^*). \tag{2.2}
$$

Then we get

$$
\begin{cases} \theta_1(x) - \theta_1(x^*) + (x - x^*)^{\mathrm{T}}(-A^{\mathrm{T}}\lambda^*) \ge 0, \quad \forall\, x \in \mathcal{X}, & (2.3a) \\ \theta_2(y) - \theta_2(y^*) + (y - y^*)^{\mathrm{T}}(-B^{\mathrm{T}}\lambda^*) \ge 0, \quad \forall\, y \in \mathcal{Y}, & (2.3b) \\ (\lambda - \lambda^*)^{\mathrm{T}}(Ax^* + By^* - b) \ge 0, \quad \forall\, \lambda \in \Re^m. & (2.3c) \end{cases}
$$

More compactly, the above inequalities can be written into the following mixed variational inequalities

$$\mathrm{VI}(\Omega, F, \theta): \theta(u) - \theta(u^*) + (w - w^*)^{\mathrm{T}} F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{2.4a}$$

where

$$u := \begin{pmatrix} x \\ y \end{pmatrix}, \quad w := \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad \theta(u) := \theta_1(x) + \theta_2(y), \tag{2.4b}$$

and

$$F(w) := \begin{pmatrix} -A^{\mathrm{T}}\lambda \\ -B^{\mathrm{T}}\lambda \\ Ax + By - b \end{pmatrix}, \quad \Omega := \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{2.4c}$$

Note that the mapping $F(w)$ defined in (2.4c) is affine with a skew-symmetric matrix, it is monotone. We denote by Ω^* the solution set of $\mathrm{VI}(\Omega, F, \theta)$.

2.2 Some notations

For a real symmetric matrix G, we mark $G \succeq 0$ ($G \succ 0$) if G is positive semidefinite (positive definite). We use $\|\cdot\|$ to denote the 2-norm of a vector and let $\|z\|_G^2 = z^{\mathrm{T}} G z$. For notational simplicity, we introduce the following matrices:

$$M = \begin{pmatrix} I_{n_2} & 0 \\ -\beta B & (1+\alpha) I_m \end{pmatrix}, \qquad Q = \begin{pmatrix} \tau r I_{n_2} & -\alpha B^{\mathrm{T}} \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \tag{2.5}$$

and

$$H = \begin{pmatrix} \tau r I_{n_2} - \frac{\alpha}{\alpha+1}\beta B^{\mathrm{T}} B & -\frac{\alpha}{1+\alpha} B^{\mathrm{T}} \\ -\frac{\alpha}{1+\alpha} B & \frac{1}{(1+\alpha)\beta} I_m \end{pmatrix}. \tag{2.6}$$

These defined matrices have some useful properties for further analysis. We summarize them in the following lemma.

Lemma 2.1. *Suppose the matrix B in (1.1) is full column rank and let*

$$r > \beta \|B^T B\|, \quad \tau \in [\frac{\alpha^2 - \alpha + 4}{\alpha^2 - 2\alpha + 5}, 1), \quad \alpha \in (-1, 1). \tag{2.7}$$

Then the matrices Q, M and H defined respectively in (2.5) and (2.6) satisfy:

$$H \succ 0, \tag{2.8}$$

and

$$HM = Q. \tag{2.9}$$

Proof. The equation (2.9) can be easily verified, we omit here. Now we prove (2.8).

Since $r > \beta\|B^T B\|$, we know that $H \succ 0$ is guaranteed if the following matrix is positive definite

$$\tilde{H} = \begin{pmatrix} (\tau - \frac{\alpha}{1+\alpha})\beta B^{\mathrm{T}} B & -\frac{\alpha}{1+\alpha} B^{\mathrm{T}} \\ -\frac{\alpha}{1+\alpha} B & \frac{1}{(1+\alpha)\beta} I_m \end{pmatrix}. \tag{2.10}$$

Note that $\tilde{H}$ can be expressed as

$$\tilde{H} = \frac{1}{1+\alpha}\begin{pmatrix} B^T & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} (\tau+\alpha\tau-\alpha)\beta & -\alpha \\ -\alpha & \frac{1}{\beta} \end{pmatrix}\begin{pmatrix} B & 0 \\ 0 & I \end{pmatrix}. \tag{2.11}$$

With the assumption that B is full column rank and $\alpha \in (-1,1)$, we only need to check

$$\begin{pmatrix} (\tau+\alpha\tau-\alpha)\beta & -\alpha \\ -\alpha & \frac{1}{\beta} \end{pmatrix} \succ 0, \tag{2.12}$$

and it is equivalent to proving

$$\tau + \alpha\tau - \alpha - \alpha^2 = (1+\alpha)(\tau-\alpha) > 0. \tag{2.13}$$

Since $\alpha \in (-1,1)$, we have $1+\alpha > 0$. We focus on checking $\tau - \alpha > 0$. Note that

$$\begin{aligned} \tau - \alpha &\geq \frac{\alpha^2-\alpha+4}{\alpha^2-2\alpha+5} - \alpha = \frac{3\alpha^2-\alpha^3+4-6\alpha}{\alpha^2-2\alpha+5} \\ &= \frac{(\alpha^2-2\alpha+4)(1-\alpha)}{\alpha^2-2\alpha+5}. \end{aligned}$$

For any $\alpha \in (-1,1)$, we have

$$\frac{(\alpha^2-2\alpha+4)(1-\alpha)}{\alpha^2-2\alpha+5} > 0. \tag{2.14}$$

Thus we get

$$\tau - \alpha > 0. \tag{2.15}$$

Using the above inequality, the assertion (2.12) is verified, and the positive definiteness of H is followed accordingly. □

3 Prediction-correction interpretation

In this section, we interpret PID-SADMM (1.9) as a prediction-correction scheme in the VI form. This interpretation does not change the algorithmic structure of PID-SADMM, but can make our convergence analysis more compact.

First, in order to alleviate the notation in the following analysis, we define an auxiliary sequence $\tilde{w}^k$ as

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1}+By^k-b) \end{pmatrix}, \tag{3.1}$$

where $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is generated by the sheme (1.9). Note that the variable x plays an intermediate role while (y,λ) are essential variables [2]. Accordingly, we use the notations

$$v = \begin{pmatrix} y \\ \lambda \end{pmatrix}, \qquad v^k = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix}, \quad \text{and} \quad \tilde{v}^k = \begin{pmatrix} \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix}, \tag{3.2}$$

to collect the essential variables. Moreover, we use $\mathcal{V}^*$ to denote the set of v^* for all subvectors of w^* in Ω^*.

In the following, the auxiliary variable $\tilde{v}^k$ denotes the predicted variable and v^{k+1} denotes the corrected variable, respectively. The following lemma shows how to generate $\tilde{v}^k$ in the PID-SADMM scheme, and it also characterizes how accurate the point $\tilde{v}^k$ defined is to a solution point of VI(Ω, F, θ).

Lemma 3.1. *Let the sequence $\{w^k\}$ be generated by PID-SADMM (1.9) and $\{\tilde{w}^k\}$ be defined in (3.1). Then we have*

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^{\mathrm{T}} F(\tilde{w}^k) \geq (v - \tilde{v}^k)^{\mathrm{T}} Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{3.3}$$

where Q is defined in (2.5).

Proof. Recall the optimality condition of x-subproblem in (1.9a), we have

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^{\mathrm{T}}\{-A^T\lambda^k + \beta A^T(Ax^{k+1} + By^k - b)\} \geq 0, \ \forall x \in \mathcal{X}. \tag{3.4}$$

Using the definition of $\tilde{x}^k$ and $\tilde{\lambda}^k$ in (3.1), the above inequality can be written as

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^{\mathrm{T}}\{-A^T\tilde{\lambda}^k\} \geq 0, \tag{3.5}$$

Analogously, the optimality condition of y-subproblem in (1.9c) is

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^{\mathrm{T}}\{-B^T\lambda^{k+\frac{1}{2}} + \beta B^T(Ax^{k+1} + By^{k+1} - b) + D_0(y^{k+1} - y^k)\} \geq 0, \ \forall y \in \mathcal{Y}. \tag{3.6}$$

Using the definitions of $\lambda^{k+\frac{1}{2}}, \tilde{y}^k$ and $\tilde{\lambda}^k$, we get

$$\begin{aligned} & \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b) \\ = \ & \lambda^k - \alpha\beta(Ax^{k+1} + By^k - b) - \beta(Ax^{k+1} + By^k - b) - \beta B(y^{k+1} - y^k) \\ = \ & \lambda^k - \beta(Ax^{k+1} + By^k - b) - \alpha(\lambda^k - \tilde{\lambda}^k) - \beta B(\tilde{y}^k - y^k) \\ = \ & \tilde{\lambda}^k - \alpha(\lambda^k - \tilde{\lambda}^k) - \beta B(\tilde{y}^k - y^k). \end{aligned} \tag{3.7}$$

Substituting it into (3.6) and recalling the definition of D_0 (see (1.9e)), we have

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^{\mathrm{T}}\{-B^T\tilde{\lambda}^k + \alpha B^T(\lambda^k - \tilde{\lambda}^k) + \tau r(\tilde{y}^k - y^k)\} \geq 0, \ \forall y \in \mathcal{Y}. \tag{3.8}$$

In addition, follows from (3.1) we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. \tag{3.9}$$

Combining (3.5), (3.8), and (3.9), we obtain

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^{\mathrm{T}} \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} - \begin{pmatrix} 0 \\ \tau r(y^k - \tilde{y}^k) - \alpha B^{\mathrm{T}}(\lambda^k - \tilde{\lambda}^k) \\ -B(y^k - \tilde{y}^k) + \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix} \right\} \geq 0, \quad \forall w \in \Omega. \tag{3.10}$$

Recall the notation Q in (2.5), and $F(\cdot)$ in (2.4c), the above inequality is exactly (3.3). ☐

Now, we establish the relationship between the output iterates v^{k+1}, v^k generated by PID-SADMM (1.9) and the auxiliary variable $\tilde{v}^k$ defined in (3.2). We show how to generate the corrected point v^{k+1} via using the predicted variable $\tilde{v}^k$ and the previous point v^k.

Lemma 3.2. *Let the sequence $\{w^k\}$ be generated by PID-SADMM (1.9) and $\{\tilde{w}^k\}$ be defined in (3.1). Then, we have*

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{3.11}$$

where M is defined in (2.5).

Proof. First, follows from the iterative scheme (3.7) and the notation (3.1), we have

$$\begin{aligned} \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \beta(A\tilde{x}^k + B\tilde{y}^k - b) \\ &= \tilde{\lambda}^k - \alpha(\lambda^k - \tilde{\lambda}^k) - \beta B(\tilde{y}^k - y^k) \\ &= \lambda^k - [(1+\alpha)(\lambda^k - \tilde{\lambda}^k) - \beta B(y^k - \tilde{y}^k)]. \end{aligned} \tag{3.12}$$

together with $y^{k+1} = \tilde{y}^k$, we have the following useful relationship

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\beta B & (1+\alpha)I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{3.13}$$

Recall the definitions of v in (3.2) and M in (2.5). The above inequality can be written as (3.11). The proof is complete. □

4 Global convergence

In this section, we analyze the convergence of PID-SADMM (1.9). We prove its global convergence under the contraction perspective. In Lemma 3.1, Observe that if $(v - \tilde{v}^k)^{\mathrm{T}} Q(v^k - \tilde{v}^k) = 0$ in the right-hand side of (3.3), then $\tilde{w}^k$ is a solution point of $\mathrm{VI}(\Omega, F, \theta)$. Therefore, here we want to estimate this term and bound it as the sum of some quadratic terms. This is given in the following lemma.

Lemma 4.1. *Let the sequence $\{w^k\}$ be generated by PID-SADMM (1.9) and $\{\tilde{w}^k\}$ be defined in (3.1). Then for any $w \in \Omega$, we have*

$$(v - \tilde{v}^k)^{\mathrm{T}} Q(v^k - \tilde{v}^k) = \frac{1}{2}(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2}(v^k - \tilde{v}^k)^{\mathrm{T}} G(v^k - \tilde{v}^k), \tag{4.1}$$

where H is defined in (2.6) and

$$G = Q^T + Q - M^T HM. \tag{4.2}$$

Proof. By using $Q = HM$ and $M(v^k - \tilde{v}^k) = (v^k - v^{k+1})$ (see (3.11)), it holds that

$$(v - \tilde{v}^k)^{\mathrm{T}} Q(v^k - \tilde{v}^k) = (v - \tilde{v}^k)^{\mathrm{T}} HM(v^k - \tilde{v}^k) = (v - \tilde{v}^k)^{\mathrm{T}} H(v^k - v^{k+1}). \tag{4.3}$$

For any vectors $a, b, c, d \in \Re^{n_2+m}$ and $H \succ 0$, it follows that

$$(a - b)^{\mathrm{T}} H(c - d) = \frac{1}{2}(\|a - d\|_H^2 - \|a - c\|_H^2) + \frac{1}{2}(\|c - b\|_H^2 - \|d - b\|_H^2). \tag{4.4}$$

Applying the above identity with $a = \upsilon$, $b = \tilde{\upsilon}^k$, $c = \upsilon^k$, and $d = \upsilon^{k+1}$ gives

$$(\upsilon - \tilde{\upsilon}^k)^{\mathrm{T}} H(\upsilon^k - \upsilon^{k+1}) = \frac{1}{2}(\|\upsilon - \upsilon^{k+1}\|_H^2 - \|\upsilon - \upsilon^k\|_H^2) + \frac{1}{2}(\|\upsilon^k - \tilde{\upsilon}^k\|_H^2 - \|\upsilon^{k+1} - \tilde{\upsilon}^k\|_H^2). \quad (4.5)$$

For the last term of (4.5), we have

$$\begin{aligned}
\|\upsilon^k - \tilde{\upsilon}^k\|_H^2 - \|\upsilon^{k+1} - \tilde{\upsilon}^k\|_H^2 &= \|\upsilon^k - \tilde{\upsilon}^k\|_H^2 - \|(\upsilon^k - \tilde{\upsilon}^k) - (\upsilon^k - \upsilon^{k+1})\|_H^2 \\
&\overset{(3.11)}{=} \|\upsilon^k - \tilde{\upsilon}^k\|_H^2 - \|(\upsilon^k - \tilde{\upsilon}^k) - M(\upsilon^k - \tilde{\upsilon}^k)\|_H^2 \\
&= 2(\upsilon^k - \tilde{\upsilon}^k)^{\mathrm{T}} HM(\upsilon^k - \tilde{\upsilon}^k) - (\upsilon^k - \tilde{\upsilon}^k)^{\mathrm{T}} M^{\mathrm{T}} HM(\upsilon^k - \tilde{\upsilon}^k) \\
&\overset{(2.9)}{=} (\upsilon^k - \tilde{\upsilon}^k)^{\mathrm{T}}(Q^{\mathrm{T}} + Q - M^{\mathrm{T}} HM)(\upsilon^k - \tilde{\upsilon}^k) \\
&\overset{(4.2)}{=} (\upsilon^k - \tilde{\upsilon}^k)^{\mathrm{T}} G(\upsilon^k - \tilde{\upsilon}^k). \qquad (4.6)
\end{aligned}$$

Now, combining (4.3), (4.5) and (4.6), the assertion (4.1) is proved. □

Lemma 4.2. *Let the sequence* $\{w^k\}$ *be generated by PID-SADMM* (1.9) *and* $\{\tilde{w}^k\}$ *be defined in* (3.1)*. Then, we have*

$$\|\upsilon^{k+1} - \upsilon^*\|_H^2 \leq \|\upsilon^k - \upsilon^*\|_H^2 - (\upsilon^k - \tilde{\upsilon}^k)^{\mathrm{T}} G(\upsilon^k - \tilde{\upsilon}^k), \quad \forall w \in \Omega. \quad (4.7)$$

Proof. Using the monotonicity of F, we obtain

$$(w - \tilde{w}^k)^{\mathrm{T}} F(w) \geq (w - \tilde{w}^k)^{\mathrm{T}} F(\tilde{w}^k). \quad (4.8)$$

Substituting the above inequality into (3.3), we have

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^{\mathrm{T}} F(w) \geq (\upsilon - \tilde{\upsilon}^k)^{\mathrm{T}} Q(\upsilon^k - \tilde{\upsilon}^k), \quad \forall w \in \Omega, \quad (4.9)$$

Setting $w = w^*$ in (4.9) and recall (2.4), we get

$$(\tilde{\upsilon}^k - \upsilon^*)^{\mathrm{T}} Q(\upsilon^k - \tilde{\upsilon}^k) \geq 0, \quad \forall w \in \Omega, \quad (4.10)$$

which together with (4.1), the assertion (4.7) is immediately obtained. □

If the matrix G defined in (4.2) is positive definite, then the sequence $\{\upsilon^k\}$ is strictly contractive with respect to the solution set $\mathcal{V}^*$ in H-norms, and the convergence of $\{\upsilon^k\}$ generated by (1.9) can be easily established following the contraction perspective. Now, we first deduce the exact expression of matrix G.

$$G := Q^{\mathrm{T}} + Q - M^{\mathrm{T}} HM = \begin{pmatrix} \tau r I_{n_2} - \beta B^{\mathrm{T}} B & 0 \\ 0 & \frac{(1-\alpha)}{\beta} I_m \end{pmatrix} = \begin{pmatrix} D_0 & 0 \\ 0 & \frac{(1-\alpha)}{\beta} I_m \end{pmatrix}. \quad (4.11)$$

Obviously, when $\tau < 1$ (see D_0 defined in (1.9e)), G may not be positive definite, and thus the inequality (4.7) in Lemma 4.2 does not imply the convergence of the sequence $\{\upsilon^k\}$. We need further investigate the term $(\upsilon^k - \tilde{\upsilon}^k)^{\mathrm{T}} G(\upsilon^k - \tilde{\upsilon}^k)$. Our main purpose is to find a lower bound in terms of the discrepancy between some non-negative quadratic terms. The following three lemmas are for this purpose.

Lemma 4.3. *Let the sequence $\{w^k\}$ be generated by PID-SADMM (1.9) and $\{\tilde{w}^k\}$ be defined in (3.1). We have the identity*

$$(\upsilon^k-\tilde{\upsilon}^k)^{\mathrm{T}}G(\upsilon^k-\tilde{\upsilon}^k)=\tau\|y^k-y^{k+1}\|_D^2+\big(\tau-1+\frac{1-\alpha}{(1+\alpha)^2}\big)\beta\|B(y^k-y^{k+1})\|^2 + \frac{1-\alpha}{\beta(1+\alpha)^2}\|\lambda^k-\lambda^{k+1}\|^2+\frac{2(1-\alpha)}{(1+\alpha)^2}(\lambda^k-\lambda^{k+1})^{\mathrm{T}}B(y^k-y^{k+1}). \quad (4.12)$$

Proof. First, recall the definition of G (4.11), we have

$$(\upsilon^k-\tilde{\upsilon}^k)^{\mathrm{T}}G(\upsilon^k-\tilde{\upsilon}^k)=\tau r\|y^k-y^{k+1}\|^2-\beta\|B(y^k-y^{k+1})\|^2+\frac{(1-\alpha)}{\beta}\|\lambda^k-\tilde{\lambda}^k\|^2. \quad (4.13)$$

On the other hand, using (3.13), we have

$$(\lambda^k-\tilde{\lambda}^k)=\frac{1}{1+\alpha}\big((\lambda^k-\lambda^{k+1})+\beta B(y^k-y^{k+1})\big), \quad (4.14)$$

and thus we get

$$\|\lambda^k-\tilde{\lambda}^k\|^2 = \frac{1}{(1+\alpha)^2}\|\lambda^k-\lambda^{k+1}\|^2+\frac{\beta^2}{(1+\alpha)^2}\|B(y^k-y^{k+1})\|^2 +\frac{2\beta}{(1+\alpha)^2}(\lambda^k-\lambda^{k+1})^{\mathrm{T}}B(y^k-y^{k+1}). \quad (4.15)$$

Substituting (4.15) into (4.13) and recalling the definition of D, we obtain (4.12). □

The last crossing term in (4.12) still needs careful analysis, and the following result presents the lower bound on it. Similar technologies are used in [17].

Lemma 4.4. *Let the sequence $\{w^k\}$ be generated by PID-SADMM (1.9). Then, we have*

$$2(\lambda^k-\lambda^{k+1})^{\mathrm{T}}B(y^k-y^{k+1}) \geq \tau\|y^k-y^{k+1}\|_D^2-\tau\|y^k-y^{k-1}\|_D^2-3(1-\tau)\beta\|B(y^k-y^{k+1})\|^2 -(1-\tau)\beta\|B(y^k-y^{k-1})\|^2. \quad (4.16)$$

Proof. From the update (1.9d), the optimality condition of y-subproblem (1.9c) is

$$\theta(y)-\theta(y^{k+1})+(y-y^{k+1})^{\mathrm{T}}\{-B^{\mathrm{T}}\lambda^{k+1}+D_0(y^{k+1}-y^k)\}\geq 0. \quad (4.17)$$

Similarly, for the previous iterate, we get

$$\theta(y)-\theta(y^{k})+(y-y^{k})^{\mathrm{T}}\{-B^{\mathrm{T}}\lambda^{k}+D_0(y^{k}-y^{k-1})\}\geq 0. \quad (4.18)$$

Setting $y=y^k$ and $y=y^{k+1}$ in (4.17) and (4.18) respectively, and then adding them, we obtain

$$(y^k-y^{k+1})^{\mathrm{T}}\{B^{\mathrm{T}}(\lambda^k-\lambda^{k+1})+D_0[(y^{k+1}-y^k)-(y^k-y^{k-1})]\}\geq 0. \quad (4.19)$$

Using notation D_0 defined in (1.9e), we have

$$\begin{aligned} & 2(\lambda^k-\lambda^{k+1})^{\mathrm{T}}B(y^k-y^{k+1}) \\ \geq\ & 2\tau\|y^k-y^{k+1}\|_D^2-2\tau(y^k-y^{k+1})^{\mathrm{T}}D(y^k-y^{k-1})-2(1-\tau)\beta\|B(y^k-y^{k+1})\|^2 \\ & +2(1-\tau)\beta(y^k-y^{k+1})^{\mathrm{T}}B^{\mathrm{T}}B(y^k-y^{k-1}) \\ \geq\ & \tau\|y^k-y^{k+1}\|_D^2-\tau\|y^k-y^{k-1}\|_D^2-3(1-\tau)\beta\|B(y^k-y^{k+1})\|^2-(1-\tau)\beta\|B(y^k-y^{k-1})\|^2. \end{aligned}$$

The last inequality is obtained by using the Cauchy-Schwarz inequality. This completes the proof. □

Next, by substituting (4.16) into (4.12), we obtain the following lemma to get the lower bound of $(\upsilon^k - \tilde{\upsilon}^k)^{\mathrm{T}} G(\upsilon^k - \tilde{\upsilon}^k)$.

Lemma 4.5. *Let the sequence $\{w^k\}$ be generated by PID-SADMM (1.9) and $\{\tilde{w}^k\}$ be defined in (3.1). Then*

$$
\begin{aligned}
(\upsilon^k - \tilde{\upsilon}^k)^{\mathrm{T}} G(\upsilon^k - \tilde{\upsilon}^k) \geq{}& \tau\|y^k - y^{k+1}\|_D^2 + C_0\beta\|B(y^k - y^{k+1})\|^2 + \frac{1-\alpha}{\beta(1+\alpha)^2}\|\lambda^k - \lambda^{k+1}\|^2 \\
&+ \frac{1-\alpha}{(1+\alpha)^2}\big[\tau(\|y^k - y^{k+1}\|_D^2 - \|y^k - y^{k-1}\|_D^2) \\
&+ (1-\tau)(\|B(y^k - y^{k+1})\|^2 - \|B(y^k - y^{k-1})\|^2)\big]. \qquad (4.20)
\end{aligned}
$$

where $C_0 = \frac{(\alpha^2 - 2\alpha + 5)\tau - (\alpha^2 - \alpha + 4)}{(1+\alpha)^2}$.

Note that when

$$
\tau \in [\frac{\alpha^2 - \alpha + 4}{\alpha^2 - 2\alpha + 5}, 1), \qquad (4.21)
$$

we have $C_0 \geq 0$, and thus

$$
C_0\beta\|B(y^k - y^{k+1})\|^2 \geq 0.
$$

Based on Lemma 4.2 and Lemma 4.5, we now establish the following results:

Lemma 4.6. *Let the sequence $\{w^k\}$ be generated by PID-SADMM (1.9). Then for any $\tau \in [\frac{\alpha^2-\alpha+4}{\alpha^2-2\alpha+5}, 1)$, we have*

$$
\begin{aligned}
&\|\upsilon^{k+1} - \upsilon^*\|_H^2 + \frac{1-\alpha}{(1+\alpha)^2}(\tau\|y^k - y^{k+1}\|_D^2 + (1-\tau)\beta\|B(y^k - y^{k+1})\|^2) \\
&\leq \|\upsilon^k - \upsilon^*\|_H^2 + \frac{1-\alpha}{(1+\alpha)^2}(\tau\|y^k - y^{k-1}\|_D^2 + (1-\tau)\beta\|B(y^k - y^{k-1})\|^2) \\
&- \Big(\tau\|y^k - y^{k-1}\|_D^2 + \frac{1-\alpha}{\beta(1+\alpha)^2}\|\lambda^k - \lambda^{k+1}\|^2\Big). \qquad (4.22)
\end{aligned}
$$

The above lemma describe the iterative contraction of the sequence $\{\|\upsilon^{k+1} - \upsilon^*\|_H^2 + \frac{1-\alpha}{(1+\alpha)^2}(\tau\|y^k - y^{k+1}\|_D^2 + (1-\tau)\beta\|B(y^k - y^{k+1})\|^2)\}$. Using Lemma 4.6, we are able to establish the convergence result of PID-SADMM (1.9).

Theorem 4.1. *Let the sequence $\{w^k\}$ be generated by PID-SADMM (1.9). Then for any $\tau \in [\frac{\alpha^2-\alpha+4}{\alpha^2-2\alpha+5}, 1)$, the sequence $\{\upsilon^k\}$ converges to some $\upsilon^\infty \in \mathcal{V}^*$.*

Proof. First, it follows from (4.22) that

$$
\begin{aligned}
&\tau\|y^k - y^{k-1}\|_D^2 + \frac{1-\alpha}{\beta(1+\alpha)^2}\|\lambda^k - \lambda^{k+1}\|^2 \\
&\leq \Big(\|\upsilon^{k+1} - \upsilon^*\|_H^2 + \frac{1-\alpha}{(1+\alpha)^2}\big(\tau\|y^k - y^{k+1}\|_D^2 + (1-\tau)\beta\|B(y^k - y^{k+1})\|^2\big)\Big) \\
&- \Big(\|\upsilon^k - \upsilon^*\|_H^2 + \frac{1-\alpha}{(1+\alpha)^2}\big(\tau\|y^k - y^{k-1}\|_D^2 + (1-\tau)\beta\|B(y^k - y^{k-1})\|^2\big)\Big). \qquad (4.23)
\end{aligned}
$$

Summing (4.23) from $k = 1$ to ∞ gives

$$\sum_{k=1}^{\infty}\Big(\tau\|y^k - y^{k-1}\|_D^2 + \frac{1-\alpha}{\beta(1+\alpha)^2}\|\lambda^k - \lambda^{k+1}\|^2\Big)$$
$$\le \|\upsilon^1 - \upsilon^*\|_H^2 + \frac{1-\alpha}{(1+\alpha)^2}\Big(\tau\|y^0 - y^1\|_D^2 + (1-\tau)\beta\|B(y^0 - y^1)\|^2\Big). \tag{4.24}$$

Recall D is positive definite, and thus

$$\lim_{k\to\infty}\|\upsilon^k - \upsilon^{k+1}\| = 0. \tag{4.25}$$

In addition, letting $\upsilon^* \in \mathcal{V}^*$ be an arbitrarily fixed point, it follows from (4.22) that

$$\|\upsilon^{k+1} - \upsilon^*\|_H^2 \le \|\upsilon^k - \upsilon^*\|_H^2 + \frac{1-\alpha}{(1+\alpha)^2}(\tau\|y^k - y^{k-1}\|_D^2 + (1-\tau)\beta\|B(y^k - y^{k-1})\|^2)$$
$$\le \|\upsilon^1 - \upsilon^*\|_H^2 + \frac{1-\alpha}{(1+\alpha)^2}(\tau\|y^0 - y^1\|_D^2 + (1-\tau)\beta\|B(y^0 - y^1)\|^2), \quad \forall\, k \ge 1. \tag{4.26}$$

Thus, the sequence $\{\upsilon^k\}$ must be bounded. Note that M defined in (2.5) is nonsingular, according to (3.11), $\{\tilde{\upsilon}^k\}$ is also bounded and must have a cluster point. Let υ^∞ be a cluster point of $\{\tilde{\upsilon}^k\}$ and the subsequence $\{\tilde{\upsilon}^{k_j}\}$ converges to υ^∞. Let x^∞ be the vector accompanied with $(y^\infty, \lambda^\infty) \in \mathcal{V}$. Then, it follows from (3.3) and (4.25) that

$$w^\infty \in \Omega, \quad \theta(u) - \theta(u^\infty) + (w - w^\infty)^{\mathrm{T}} F(w^\infty) \ge 0, \quad \forall w \in \Omega. \tag{4.27}$$

which means that w^∞ is a solution of $\mathrm{VI}(\Omega, F, \theta)$, and its essential part $\upsilon^\infty \in \mathcal{V}^*$. Since $\upsilon^\infty \in \mathcal{V}^*$, it follows from (4.26) that

$$\|\upsilon^{k+1} - \upsilon^\infty\|_H^2 \le \|\upsilon^k - \upsilon^\infty\|_H^2 + \frac{1-\alpha}{(1+\alpha)^2}\Big(\tau\|y^k - y^{k-1}\|_D^2 + (1-\tau)\beta\|B(y^k - y^{k-1})\|^2\Big). \tag{4.28}$$

Thus the sequence $\{\|\upsilon^{k+1} - \upsilon^\infty\|_H^2\}$ is monotonically nonincreasing, and the sequence $\{\upsilon^k\}$ can't have more than one cluster point. Therefore, the sequence $\{\upsilon^k\}$ converges to υ^∞ and the proof is complete. □

5 Convergence rate

To estimate the convergence rate in terms of the iteration complexity, we need a characterization of the solution set of VI (2.4), which is described in the following theorem. The proof can be found in [19] (Theorem 2.1).

Theorem 5.1. *The solution set of $VI(\Omega, F, \theta)$ is convex and it can be characterized as*

$$\Omega^* = \bigcap_{w\in\Omega}\big\{\tilde{w} \in \Omega : \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \ge 0\big\}. \tag{5.1}$$

Therefore, for a given accuracy $\epsilon > 0$, $\tilde{w} \in \Omega$ is called an ϵ-approximate solution point of $\mathrm{VI}(\Omega, F, \theta)$ if it satisfies

$$\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w) \le \epsilon, \;\; \forall\, w \in \mathcal{D}_{(\tilde{w})},$$

where

$$\mathcal{D}_{(\tilde{w})} = \{w \in \Omega \,|\, \|w - \tilde{w}\| \le 1\}.$$

To estimate the convergence rate in terms of the iteration complexity for a sequence $\{w^k\}$, we need to show that for given $\epsilon > 0$, after t iterations, this sequence can offer a point $\tilde{w} \in \Omega$ such that

$$\tilde{w} \in \Omega \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \big\{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\big\} \le \epsilon. \tag{5.2}$$

Lemma 5.1. *Let the sequence $\{w^k\}$ be generated by PID-SADMM (1.9) and $\{\tilde{w}^k\}$ be defined in (3.1). Then for any $\tau \in [\frac{\alpha^2-\alpha+4}{\alpha^2-2\alpha+5}, 1)$, we have*

$$\begin{aligned}
&\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^{\mathrm{T}} F(w) \\
&\ge \Big(\frac{1}{2}\|\upsilon - \upsilon^{k+1}\|_H^2 + \frac{1-\alpha}{2(1+\alpha)^2}\big[\tau\|y^k - y^{k+1}\|_D^2 + (1-\tau)\|B(y^k - y^{k+1})\|^2\big]\Big) \\
&- \Big(\frac{1}{2}\|\upsilon - \upsilon^k\|_H^2 + \frac{1-\alpha}{2(1+\alpha)^2}\big[\|y^k - y^{k-1}\|_D^2 + (1-\tau)\|B(y^k - y^{k-1})\|^2\big]\Big).
\end{aligned} \tag{5.3}$$

Proof. The lemma can be direct obtained by combining (4.1), (4.9) and (4.20). □

Theorem 5.2. *Let $\{w^k\}$ be generated by PID-SADMM (1.9) and $\{\tilde{w}^k\}$ be defined in (3.1). Then for any $\tau \in [\frac{\alpha^2-\alpha+4}{\alpha^2-2\alpha+5}, 1)$, we have*

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^{\mathrm{T}} F(w) \le \frac{1}{2t}\Big(\|\upsilon - \upsilon^1\|_H^2 + \frac{1-\alpha}{(1+\alpha)^2}\big[\tau\|y^0 - y^1\|_D^2 + (1-\tau)\beta\|B(y^0 - y^1)\|^2\big]\Big), \ \forall w \in \Omega, \tag{5.4}$$

where

$$\tilde{w}_t = \frac{1}{t}\big(\sum_{k=1}^{t} \tilde{w}^k\big). \tag{5.5}$$

Proof. Summing (5.3) over $k = 1, 2\cdots, t$ gives

$$\begin{aligned}
&\sum_{k=1}^{t} \theta(\tilde{u}^k) - t\theta(u) + \big(\sum_{k=1}^{t} \tilde{w}^k - tw\big)^{\mathrm{T}} F(w) \\
&\le \Big(\frac{1}{2}\|\upsilon - \upsilon^1\|_H^2 + \frac{1-\alpha}{2(1+\alpha)^2}\big[\tau\|y^0 - y^1\|_D^2 + (1-\tau)\|B(y^0 - y^1)\|^2\big]\Big).
\end{aligned} \tag{5.6}$$

Since $\theta(\cdot)$ is convex and

$$\tilde{u}_t = \frac{1}{t}\big(\sum_{k=1}^{t} \tilde{u}^k\big), \tag{5.7}$$

we have

$$\theta(\tilde{u}_t) \le \frac{1}{t}\big(\sum_{k=1}^{t} \theta(\tilde{u}^k)\big), \tag{5.8}$$

and the assertion of the theorem is obtained. □

For a given compact set $D_{(\tilde{w})} \subset \Omega$, let

$$d := \sup\{\frac{1}{2}\|\upsilon - \upsilon^1\|_H^2 + \frac{1-\alpha}{2(1+\alpha)^2}\big[\tau\|y^0 - y^1\|_D^2 + (1-\tau)\|B(y^0 - y^1)\|^2\big] \;\big|\; w \in D_{(\tilde{w})}\}. \tag{5.9}$$

Theorem 5.2 shows that, after t iterations of the PID-SADMM (1.9), the point $\tilde{w}_t$ defined in (5.5) satisfies

$$\tilde{w}_t \in \Omega \quad \text{and} \quad \sup_{w\in D_{(\tilde{w})}}\{\theta(\tilde{u}-\theta(u)+(\tilde{w}_t-w)^{\mathrm{T}}F(w)\} \leq \frac{d}{t} = \mathcal{O}(1/t). \tag{5.10}$$

which means $\tilde{w}_t$ is an approximate solution of $\mathrm{VI}(\Omega, F, \theta)$ (2.4a) with an accuracy of $\mathcal{O}(1/t)$. Hence, the $\mathcal{O}(1/t)$ convergence rate of proposed PID-SADMM is estabilished in an ergodic sense.

6 Numerical experiments

In this section, we present numerical results to demonstrate the computational performance of PID-SADMM. Note that when $\tau = 1$ in PID-SADMM (1.9), D_0 is positive definite, the resulting algorithm is called linearized SADMM with positive-definite proximal regularization (PD-SADMM). Since the comparable performance of PD-SADMM relative to other algorithms is well demonstrated in the literature, see, e.g., [23, 11], here we focus on comparing PID-SADMM and its positive part PD-SADMM. We concentrate on two representative problems: LASSO model [2] and total variation minimization [28] . All numerical experiments were conducted using MATLAB R2014a on a PC with an Intel I7 processor, and 8Gb RAM.[3]

6.1 LASSO model

The LASSO model in [29, 2] is

$$\min_{y}\{\frac{1}{2}\|Ay-b\|^2+\sigma\|y\|_1\}, \tag{6.1}$$

where $\|y\|_1 := \sum_{i=1}^{n}|y_i|$, $A \in \Re^{m\times n}$ is a design matrix usually with $m \ll n$, m is the number of data points, n is the number of features, $b \in \Re^m$ is the response vector and $\sigma > 0$ is a regularization parameter. The LASSO model provides a sparse estimation of y when there are more features than data points. It can also be explained as a model for finding a sparse solution of the under-determined system of linear equations $Ay = b$.

Introducing a new variable x, we can write (6.1) into the form of

$$\min_{x,y}\{\frac{1}{2}\|x-b\|^2+\sigma\|y\|_1 \mid x-Ay=0,\, x\in\Re^m, y\in\Re^n\}, \tag{6.2}$$

which is a special case of (1.1). Then applying the proposed PID-SADMM (1.9) to (6.2), we obtain the scheme

$$\begin{cases} x^{k+1} = \operatorname{argmin}\{\frac{1}{2}\|x-b\|^2 - \lambda^k(x-Ay^k) + \frac{\beta}{2}\|x-Ay^k\|^2\}, & (6.3a)\\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(x^{k+1}-Ay^k), & (6.3b)\\ y^{k+1} = \operatorname*{argmin}_{y}\{\sigma\|y\|_1 + \frac{\tau r}{2}\|y-(y^k+\frac{1}{\tau r}q_k)\|^2\}, & (6.3c)\\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(x^{k+1}-Ay^{k+1}), & (6.3d) \end{cases}$$

where $q_k = -A^T\big(\lambda^{k+\frac{1}{2}} - \beta(x^{k+1}-Ay^k)\big)$. The solution of the x-subproblem (6.3a) is given by

$$x^{k+1} = \frac{1}{1+\beta}(b+\lambda^k+\beta Ay^k). \tag{6.4}$$

[3]The PID-SADMM code is available online at https://github.com/gaobingaobingaobin/PIDSADMM.

Then, for the y-subproblem (6.3c), its closed-form solution is given by

$$y^{k+1} = S_{\sigma/\tau r}(y^k + q_k/\tau r), \tag{6.5}$$

where $S_\delta(t)$ is the soft thresholding operator [8] defined as

$$(S_\delta(t))_i = (1 - \delta/|t_i|)_+ \cdot t_i, \quad i = 1, 2, \cdots, m. \tag{6.6}$$

We generate the data in the following manner: We first choose $A_{ij} \sim \mathcal{N}(0,1)$ and then scaled the columns to have unit norm. We use the script 'sprandn' to generate a sparse vector y^* which have approximately density $= 100/n$ non-zeros entries taken from the normal distribution with zero mean and unit variance. We generate b via $b = Ay^* + e$, where e is a small white noise taken from $e \sim \mathcal{N}(0, 10^{-3}I)$. We test five cases of the dimension of A ranging from 900×3000 to 1500×5000, and set $r = \beta\|A^TA\|$. We set the regularization parameter σ to 0.1 and the ADMM parameter β to 1. We set $\alpha = -0.3$ and $\alpha = 0.3$, respectively. Since small proximal term can allow for larger steps, here we always let $\tau = \frac{\alpha^2-\alpha+4}{\alpha^2-2\alpha+5}$. The termination tolerance is defined by

$$\|x^{k+1} - Ay^{k+1}\|_2 \le \epsilon^{\text{pri}} \quad \text{and} \quad \|\beta \mathrm{A}(\mathrm{y}^{\mathrm{k}+1} - \mathrm{y}^{\mathrm{k}})\|_2 \le \epsilon^{\text{dual}}, \tag{6.7}$$

where $\epsilon^{\text{pri}} = \sqrt{n}\epsilon^{\text{abs}} + \epsilon^{\text{rel}} \max\{\|x^{k+1}\|_2, \|Ay^{k+1}\|_2\}$, and $\epsilon^{\text{dual}} = \sqrt{n}\epsilon^{\text{abs}} + \epsilon^{\text{rel}}\|y^{k+1}\|_2$. with ϵ^{abs} and ϵ^{rel} set to be 10^{-4} and 10^{-2} respectively for all of the methods. The initial points (y^0, λ^0) were chosen to be zero.

Table 1 shows the the number of iterations and runtime in seconds of the PD-SADMM, PID-SADMM schemes applied to problem (6.2) for different dimensions of A. To better visualize the performance improvement, we plot the results in terms of iterations in Fig. 1. From the table and the figure, we see that the PID-SADMM performs better than the PD-SADMM, and can achieve an improvement of about 10-15% reduction both in the number of iterations and runtime over PD-SADMM.

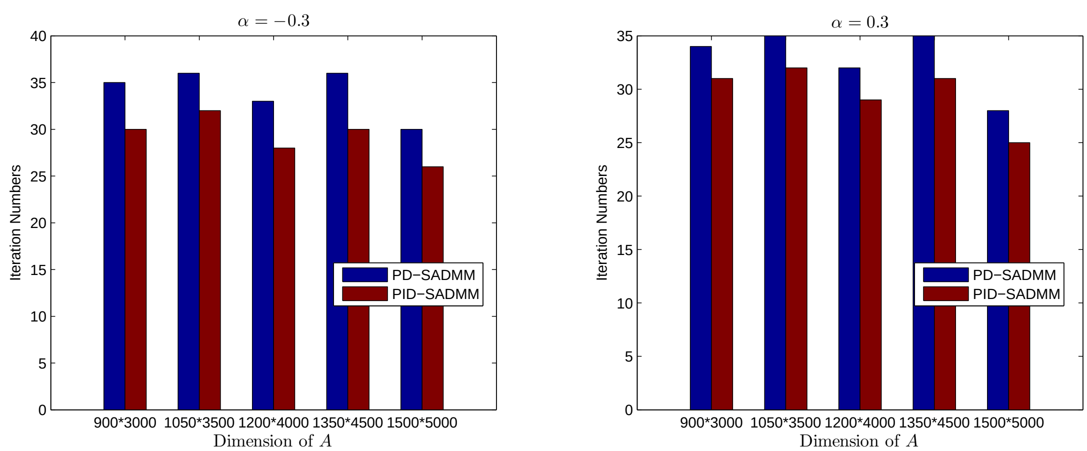

Figure 1. Comparison results on $\alpha = -0.3$ and $\alpha = 0.3$ respectively for LASSO model.

6.2 Total variation denoising model

In this subsection, we test the PID-SADMM on the total variation (TV) denoising problem in [28, 1]:

$$\frac{1}{2}\|y - b\|^2 + \eta\|Dy\|_1 \tag{6.8}$$

Table 1: Comparison between the number of iterations (time in seconds) taken by PD-SADMM and PID-SADMM for LASSO model.

Dim. of A $m \times n$	PD-SADMM ($\alpha = -0.3$)	PID-SADMM ($\alpha = -0.3$)	Ratio(%)	PD-SADMM ($\alpha = 0.3$)	PID-SADMM ($\alpha = 0.3$)	Ratio(%)
900×3000	35(3.21)	30(2.71)	0.86(0.84)	34(3.11)	31(2.82)	0.91(0.91)
1050×3500	36(4.85)	32(4.03)	0.89(0.83)	35(4.33)	32(4.19)	0.91(0.97)
1200×4000	33(5.28)	28(4.52)	0.85(0.86)	32(5.10)	29(4.64)	0.91(0.91)
1350×4500	36(7.34)	30(6.09)	0.83(0.83)	35(7.42)	31(6.30)	0.89(0.85)
1500×5000	30(7.54)	26(6.55)	0.87(0.87)	28(7.05)	25(6.27)	0.89(0.89)

where $\|y\|_1 = \sum_{i=1}^{2n} |y_i|$, $D^T = [D_1^T, D_2^T]^T$ is a discrete gradient operator with $D_1 : \Re^n \to \Re^n$, $D_2 : \Re^n \to \Re^n$ being the finite-difference operators in the horizontal and vertical directions, respectively.

Similar with (6.2), by introducing a new variable x, we reformulate (6.8) as

$$\min_{x,y}\{\frac{1}{2}\|y-b\|^2 + \eta\|x\|_1 \mid x - Dy = 0, x \in \Re^{2n}, y \in \Re^{2n}\}. \tag{6.9}$$

which is a special case of (1.1). Applying PID-SADMM to (6.9), the resulting iterative scheme is

$$\begin{cases} x^{k+1} = \text{argmin}\,\{\eta\|x\|_1 - \lambda(x - Dy) + \frac{\beta}{2}\|x - Dy\|^2\}, & (6.10a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(x^{k+1} - Dy^k), & (6.10b) \\ y^{k+1} = \underset{y}{\text{argmin}}\,\{\frac{1}{2}\|y-b\|^2 + \frac{\tau r}{2}\|y - (y^k + \frac{1}{\tau r}q_k)\|^2\}, & (6.10c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(x^{k+1} - Dy^{k+1}), & (6.10d) \end{cases}$$

where $q_k = -D^T\big(\lambda^{k+\frac{1}{2}} - \beta(x^{k+1} - Dy^k)\big)$. The x-subproblem (6.10a) is also a $l_1 + l_2$ minimization, thus can be solved exactly by invoking the soft thresholding operator (see (6.6)) given by

$$x^{k+1} = S_{\eta/\beta}(Dy^k + \lambda/\beta). \tag{6.11}$$

The y-subproblem (6.10c) amounts to solving a system of linear equations that is given by

$$y^{k+1} = \frac{1}{1+\tau r}(b + y^k + \frac{1}{\tau r}q_k). \tag{6.12}$$

We test the algorithm on the random data of six different $n \times n$ sizes of D. For each instance, we use two different step sizes of α, which are $\{-0.1, 0.1\}$. The associated parameters for the algorithms are chosen as $\eta = 5$, $\beta = 1$, $r = \beta\|D^T D\|$, and $\tau = \frac{\alpha^2-\alpha+4}{\alpha^2-2\alpha+5}$. In addition, we choose the termination criterion

$$\|x^{k+1} - Dy^{k+1}\|_2 \le \epsilon^{\text{pri}} \quad \text{and} \quad \|\beta \mathrm{D}(\mathrm{y}^{k+1} - \mathrm{y}^k)\|_2 \le \epsilon^{\text{dual}}, \tag{6.13}$$

where $\epsilon^{\text{pri}} = \sqrt{n}\epsilon^{\text{abs}} + \epsilon^{\text{rel}}\max\{\|x^{k+1}\|_2, \|Dy^{k+1}\|_2\}$, and $\epsilon^{\text{dual}} = \sqrt{n}\epsilon^{\text{abs}} + \epsilon^{\text{rel}}\|y^{k+1}\|_2$. The initial points for all methods were $(y^0, \lambda^0) = (\mathbf{0}, \mathbf{0})$. For a given dimension $n \times n$, we generate the data for (6.8) randomly as the following ways:

```
for j = 1:3
   idx = randsample(n,1);
   k = randsample(1:10,1);
   y(ceil(idx/2):idx) = k*y(ceil(idx/2):idx);
end
b = y + randn(n,1);
e = ones(n,1);
D = spdiags([e -e], 0:1, n,n);
```

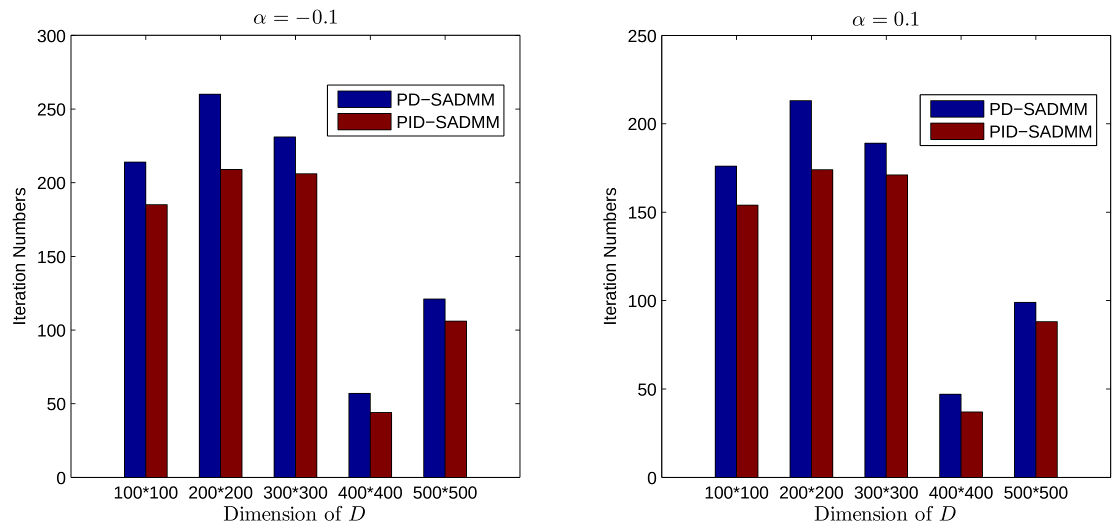

Figure 2. Comparison results on $\alpha = -0.1$ and $\alpha = 0.1$ respectively for TV model.

Table 2 and Fig. 2 report the comparison between the PD-SADMM and the PID-SADMM for TV model. We observe that the PID-SADMM scheme uses fewer iterations for all experiments in comparison with the PD-SADMM with the same values of r and β. For instance, for 400×400 the PID-SADMM scheme cost almost 23% reduction in the number of iterations as compared to the PD-SADMM.

Table 2: Comparison between the number of iterations (time in seconds) taken by PD-SADMM and PID-SADMM for TV model.

Dim. of D	PD-SADMM ($\alpha = -0.1$)	PID-SADMM ($\alpha = -0.1$)	Ratio(%)	PD-SADMM ($\alpha = 0.1$)	PID-SADMM ($\alpha = 0.1$)	Ratio(%)
100×100	214(0.02)	185(0.01)	0.86(0.65)	176(0.01)	154(0.01)	0.88(0.92)
200×200	260(0.02)	209(0.02)	0.80(0.84)	213(0.02)	174(0.02)	0.82(0.84)
300×300	231(0.02)	206(0.02)	0.89(0.91)	189(0.02)	171(0.02)	0.90(0.88)
400×400	57(0.01)	44(0.00)	0.77(0.80)	47(0.01)	37(0.00)	0.79(0.76)
500×500	121(0.01)	106(0.01)	0.88(0.88)	99(0.01)	88(0.01)	0.89(0.89)

7 Conclusions

Proximal regularization is an useful technology to design implementable algorithms in optimization. For alternating direction method of multipliers (ADMM) and its variants, using proper regularization, the resulting subproblems are usually easier to solve or have closed-form solutions. In this paper, we restrict our discussion to the prox-linear regularization case, and inspired by He et al's work, we revisit the linearized symmetric ADMM for solving linearly constrained separable convex optimization problems. Without any additional assumptions, we confirm that the symmetric ADMM can also be regularized with positive-indefinite proximal term. The global convergence and worst-case $\mathcal{O}(1/t)$ convergence rate measured by the iteration complexity are established for the newly linearized SADMM. We demonstrate the improvements via using positive-indefinite proximal term by some numerical experiments. Note that the choices of the parameters are highly problem dependent, it will be interesting to investigate strategies to dynamically adjust the step sizes (α, β and τ) to accelerate the convergence. We leave this topic to be our primary work in the near future.

Finally, we would mention that the symmetric ADMM (1.8) discussed in this paper is equivalent to the generalized ADMM [9] proposed by Eckstein and Bertsekas, which takes

$$\begin{cases} x^{k+1} = \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, & (7.1a) \\ y^{k+1} = \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|[rAx^{k+1} - (1-r)(By^k - b)] + By - b\|^2 | y \in \mathcal{Y}\}, & (7.1b) \\ \lambda^{k+1} = \lambda^k - \beta\{[rAx^{k+1} - (1-r)(By^k - b)] + By^{k+1} - b\}, & (7.1c) \end{cases}$$

where the parameter $r \in (0, 2)$. See [18] for elaboration on the detail. Thus our discussion on proximal version of symmetric ADMM in this paper can directly extend to the generalized ADMM (7.1), and gives the following proximal iterates

$$\begin{cases} x^{k+1} = \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, & (7.2a) \\ y^{k+1} = \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|[rAx^{k+1} - (1-r)(By^k - b)] + By - b\|^2 & \\ \qquad\qquad + \frac{1}{2}\|y - y^k\|_{D_0}^2 \,|\, y \in \mathcal{Y}\}, & (7.2b) \\ \lambda^{k+1} = \lambda^k - \beta\{[rAx^{k+1} - (1-r)(By^k - b)] + By^{k+1} - b\}, & (7.2c) \end{cases}$$

where D_0 is defined in (1.9e).

References

[1] A. Beck and M. Teboulle , *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Trans. Image Process., 18(2009), pp. 2419-2434.

[2] 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. Learn., 3 (2011), pp. 1-122.

[3] X. J. Cai, D. R. Han and X. M. Yuan, *On the convergence of the direct extension of ADMM for three-block separable convex minimization models with one strongly convex function*, Comput. Optim. Appl., (2016), pp. 1-35.

[4] T. F. Chan and R. Glowinski, *Finite Element Approximation and Iterative Solution of a Class of Mildly Nonlinear Elliptic Equations*, Technical Report STAN-CS-78-674, Computer Science Department, Stanford University, 1978.

[5] R. H. Chan, M. Tao and X. M. Yuan, *Linearized alternating direction method of multipliers for constrained linear least-squares problem*, East Asian J. Appl. Math., 2(2012), pp. 326-341.

[6] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, Math. Program., Ser. A, 155 (2016), pp. 57-79.

[7] W. Deng and W. T. Yin, *On the global and linear convergence of the generalized alternating direction method of multipliers*, J. Sci. Comput., 66 (2016), pp. 1-28.

[8] D. L. Donoho and Y. Tsaig, *Fast solution of l1-norm minimization problems when the solution may be sparse*, IEEE Trans. Inform. Theory, 54 (2008), pp. 4789-4812.

[9] J. Eckstein and 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] J. Eckstein and W. Yao, *Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives*, Pacific J. Optim., 11(2015), pp. 619-644.

[11] E. X. Fang, B. S. He, H. Liu and X. M. Yuan, *The generalized alternating direction method of multipliers: New theoretical insights and applications*, Math. Program. Comput., 7(2015), pp. 149-187.

[12] R. Glowinski and 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*, Revue Fr. Autom. Inform. Rech. Opér., Anal. Numér. 2 (1975), pp. 41–76.

[13] R. Glowinski, *On alternating direction methods of multipliers: A historical perspective*, In Modeling, Simulation and Optimization for Science and Technology, W. Fitzgibbon, Y.A. Kuznetsov, P. Neittaanmki & O. Pironneau, eds., Computational Methods in Applied Sciences, Vol. 34, Springer, Dordrecht, 2014, 59-82.

[14] R. Glowinski, T. W. Pan and X. C. Tai, *Some facts about operator-splitting and alternating direction methods*, CAM report 16-10, UCLA, (2016).

[15] D. R. Han and X. M. Yuan, *A note on the alternating direction method of multipliers*, J. Optim. Theory Appl., 155(2012), pp. 227-238.

[16] Y. Gu, B. Jiang and D. R. Han, *A semi-proximal-based strictly contractive Peaceman-Rachford splitting method*, Optimization Online (2015), http://www.optimization-online.org/DB_HTML/2015/06/4950.html.

[17] B. S. He, F. Ma and X. M. Yuan, *Linearized alternating direction method of multipliers via positive-indefinite proximal regularization for convex programming*, Optimization Online (2016), http://www.optimization-online.org/DB_HTML/2016/07/5569.html.

[18] B. S. He, F. Ma and X. M. Yuan, *Convergence study on the symmetric version of ADMM with larger step sizes*, SIAM J. Imag. Sci., (2016), doi:10.1137/15M1044448.

[19] B. S. He and X. M. Yuan, *On the $O(1/t)$ convergence rate of the Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700-709.

[20] B. S. He and X. M. Yuan, *On non-ergodic convergence rate of the Douglas-Rachford alternating direction method*, Numer. Math., 130 (2015), pp. 567-577.

[21] M. Y. Hong and Z. Q. Luo, *On the linear convergence of the alternating direction method of multipliers*, Math. Program., (2016), pp. 1-35.

[22] M. Li, D. F. Sun and K. C. Toh, *A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization*, SIAM J. Opt., 26 (2016), pp. 922-950.

[23] X. X. Li and X. M. Yuan, *A proximal strictly contractive Peaceman-Rachford splitting method for convex programming with applications to imaging*, SIAM J. Imag. Sci., 8(2) (2015), pp. 1332-1365.

[24] T. Y. Lin, S. Q. Ma and S. Z. Zhang, *On the global linear convergence of the ADMM with multi-block variables*, SIAM J. Opt., 25 (2015), pp. 1478-1497.

[25] S. Q. Ma, *Alternating proximal gradient method for convex minimization*, J. Sci. Comput., 68 (2016), pp. 546–572.

[26] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16(1979), pp. 964-979.

[27] D. H. Peaceman and H. H. Rachford, *The numerical solution of parabolic elliptic differential equations*, SIAM J. Appl. Math. 3 (1955), pp. 28-41.

[28] L. Rudin, S. J. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp.259-268.

[29] R. Tibshirani, *Regression shrinkage and selection via the LASSO*, J. Roy. Statist. Soc. Ser. B, 58 (1996), pp. 267-288.

[30] X. F. Wang and X. M. Yuan, *The linearized alternating direction method of multipliers for Dantzig selector*, SIAM J. Sci. Comput., 34 (2012), pp. A2792–A2811.

[31] J. F. Yang and X. M. Yuan, *Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization*, Math. Comp., 82 (2013), pp. 301-329.