

Generalized ADMM with Optimal Indefinite Proximal Term for Linearly Constrained Convex Optimization

Fan Jiang¹ Zhongming Wu² Xingju Cai³

Abstract. We consider the generalized alternating direction method of multipliers (ADMM) for linearly constrained convex optimization. Many problems derived from practical applications have showed that usually one of the subproblems in the generalized ADMM is hard to solve, thus a special proximal term is added. In the literature, the proximal term can be indefinite which plays a vital role in accelerating numerical performance. In this paper, we are devoted to deriving the optimal lower bound of the proximal parameter and result in the generalized ADMM with optimal indefinite proximal term. The global convergence and the $O(1/t)$ convergence rate measured by the iteration complexity of the proposed method are proved. Moreover, some preliminary numerical experiments on LASSO and total variation-based denoising problems are presented to demonstrate the efficiency of the proposed method and the advantage of the optimal lower bound.

Keywords. Alternating direction method of multipliers, generalized, convex programming, convergence rate, indefinite proximal term, optimal range.

1 Introduction

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

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

where $A \in \mathbb{R}^{m \times n_1}$, $B \in \mathbb{R}^{m \times n_2}$, $b \in \mathbb{R}^m$, $\mathcal{X} \subset \mathbb{R}^{n_1}$ and $\mathcal{Y} \subset \mathbb{R}^{n_2}$ are closed convex sets, $\theta_1 : \mathbb{R}^{n_1} \rightarrow \mathbb{R}$ and $\theta_2 : \mathbb{R}^{n_2} \rightarrow \mathbb{R}$ are convex (not necessarily smooth) functions. The model (1.1) with separable structure captures a variety of applications in signal processing, computer vision and statistical learning, see e.g., [2, 3, 4, 31, 34, 35, 37] and references therein for more details. Throughout this paper, the solution set of (1.1) is assumed to be nonempty and the matrix B is full column rank. The augmented Lagrangian function of (1.1) is defined as

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

where $\lambda \in \mathbb{R}^m$ is the Lagrangian multiplier and $\beta > 0$ is a penalty parameter.

It is well known that alternating direction method of multipliers (ADMM) is an efficient approach for solving the problem (1.1), which originally proposed in [13, 16]. The iterative scheme of the classical ADMM can be read as

$$\begin{cases} x^{k+1} = \arg \min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & (1.2a) \\ y^{k+1} = \arg \min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, & (1.2b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & (1.2c) \end{cases}$$

¹School of Mathematical Sciences, Key Laboratory for NSLSCS of Jiangsu Province, Nanjing Normal University, Nanjing 210023, Peoples Republic of China. Email: 15905154902@163.com

²School of Economics and Management, Southeast University, Nanjing 210096, Peoples Republic of China. Email: wuzm1991@126.com

³School of Mathematical Sciences, Key Laboratory for NSLSCS of Jiangsu Province, Nanjing Normal University, Nanjing 210023, Peoples Republic of China. Email: caixingju@njnu.edu.cn

This method treats the functions θ_1 and θ_2 individually and makes the subproblems easier to solve than the original one (1.1). Recently, due to its high efficiency, the ADMM has been applied in a broad spectrum of areas such as signal processing [5, 6, 7], image restoration and denoising [12, 33], and sparse signal recovery [11]. We refer to [1, 9, 15, 17] for some reviews on ADMM.

In order to further accelerate the numerical performance of the original ADMM (1.2), a generalized ADMM which applied an acceleration scheme in y -subproblem (1.2b) was proposed in [8] and extensively studied in [8, 10, 27]. The scheme of the generalized ADMM is

$$\begin{cases} x^{k+1} = \arg \min\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2} \|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, & (1.3a) \\ y^{k+1} = \arg \min\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2} \|[\alpha Ax^{k+1} - (1 - \alpha)(By^k - b)] + By - b\|^2 \mid y \in \mathcal{Y}\}, & (1.3b) \\ \lambda^{k+1} = \lambda^k - \beta\{[\alpha Ax^{k+1} - (1 - \alpha)(By^k - b)] + By^{k+1} - b\}, & (1.3c) \end{cases}$$

where the parameter $\alpha \in (0, 2)$ is a relaxation factor. Apparently, the generalized ADMM (1.3) reduces to the original ADMM (1.2) when $\alpha = 1$. The generalized ADMM inherits the whole advantages of the original ADMM and usually performs better in numerical experiments with some values $\alpha \in (1, 2)$, see e.g., [8, 10]. Indeed, the scheme (1.3) is equivalent to

$$\begin{cases} x^{k+1} = \arg \min\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2} \|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, & (1.4a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), & (1.4b) \\ y^{k+1} = \arg \min\{\theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2} \|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\}, & (1.4c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b), & (1.4d) \end{cases}$$

where $r = \alpha - 1$.

However, the efficiency of the generalized ADMM (1.4) largely relies on the solving difficulty of the subproblems (1.4a) and (1.4c). Usually, when applying the generalized ADMM for solving some concrete applications such as low-rank optimization problem [36] and the models considered in [38], one of the subproblems is relatively easy to solve and the other one is difficult because the corresponding coefficient matrix is somewhat complicated. Therefore, we assume that the subproblem (1.4a) is easy to solve throughout this paper. While for another subproblem (1.4c), instead of solving itself, a proximal term $\|y - y^k\|_D^2$ is added, thus a new scheme is produced as follows

$$\begin{cases} x^{k+1} = \arg \min\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2} \|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, & (1.5a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), & (1.5b) \\ y^{k+1} = \arg \min\{\theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2} \|Ax^{k+1} + By - b\|^2 + \frac{1}{2} \|y - y^k\|_D^2 \mid y \in \mathcal{Y}\}, & (1.5c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b), & (1.5d) \end{cases}$$

where D is a positive definite matrix. If we choose a special form of D , i.e.,

$$D = \rho I - \beta B^T B \quad \text{with} \quad \rho > \beta \|B^T B\|, \quad (1.6)$$

the subproblem (1.5c) is the linearized version of (1.4c). The solution of (1.5c) can be obtained via

$$y^{k+1} = \arg \min\{\theta_2(y) + \frac{\rho}{2} \|y - (y^k + q_k)\|^2 \mid y \in \mathcal{Y}\}, \quad (1.7)$$

where

$$q_k = \frac{1}{\rho} B^T \left(\lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^k - b) \right). \quad (1.8)$$

We name (1.5) with special D defined in (1.6) proximal generalized ADMM (PG-ADMM for short). Compared with (1.4c), the problem (1.7) is usually easier to solve if the proximal mapping of θ_2 , such as the soft-thresholding operator for l_1 -norm minimization [35], is easy to evaluate. The PG-ADMM (1.5) is similar to the method proposed in [10] while they added the proximal term to the x -subproblem.

For the PG-ADMM (1.5) and the linearized ADMM in [29, 30, 36], the positive definiteness of the proximal matrix D plays a central role in ensuring the convergence. However, these methods with small values of ρ in (1.6) or with indefinite proximal term [28] are preferred in practical implementation. Along this line, He et al. [21, 22] proposed the indefinite proximal linearized ADMM (IP-LADMM) which was devoted to shrinking the value of ρ while still guaranteed the convergence. The method can be read as

$$\begin{cases} x^{k+1} = \arg \min \{ \mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X} \}, & (1.9a) \\ y^{k+1} = \arg \min \{ \mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2} \|y - y^k\|_{D_0}^2 \mid y \in \mathcal{Y} \}, & (1.9b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & (1.9c) \end{cases}$$

where

$$D_0 = \tau \rho I - \beta B^T B, \quad 0.75 < \tau < 1 \quad \text{and} \quad \rho > \beta \|B^T B\|. \quad (1.10)$$

It is easy to see that the matrix D_0 defined in (1.10) is not necessarily positive definite. They also gave a simple counterexample to illustrate that 0.75 is the optimal lower bound of τ .

For the PG-ADMM (1.5), Gao and Ma [14] also relaxed the requirement the positive definiteness of D in (1.5c) and proposed the positive-indefinite proximal symmetric ADMM (PID-SADMM)

$$\begin{cases} x^{k+1} = \arg \min \{ \theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2} \|Ax + By^k - b\|^2 \mid x \in \mathcal{X} \}, & (1.11a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), & (1.11b) \\ y^{k+1} = \arg \min \{ \theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2} \|Ax^{k+1} + By - b\|^2 + \frac{1}{2} \|y - y^k\|_{D_0}^2 \mid y \in \mathcal{Y} \}, & (1.11c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b), & (1.11d) \end{cases}$$

where

$$D_0 = \tau \rho I - \beta B^T B, \quad \text{with} \quad \rho > \beta \|B^T B\| \quad \text{and} \quad \tau \in \left[\frac{r^2 - r + 4}{r^2 - 2r + 5}, 1 \right). \quad (1.12)$$

Note that the proximal matrix D_0 in (1.11c) is not necessarily positive definite. When $r = 0$, the PID-SADMM reduces to the frame of the IP-LADMM (1.9) and $\tau \in [0.8, 1)$ while the range of τ is $(0.75, 1)$ in (1.9). So there is a gap between the optimal lower bound and $\frac{r^2 - r + 4}{r^2 - 2r + 5}$.

Motivated by the very recent work [22], in this paper, we further relax the requirement of the proximal matrix D_0 and focus on finding the optimal lower bound of τ in (1.12). The main contributions of this paper are presented as follows. We propose an indefinite proximal generalized ADMM (IPG-ADMM) and prove the global convergence and a worst $O(1/t)$ convergence rate of the proposed method by specifying a better domain $\tau \in (\frac{3+r}{4}, 1)$. Besides, we present an example to show that the proposed method is not necessarily convergent when $\tau \in (0, \frac{3+r}{4})$, which means that the optimal

lower bound of the parameter τ in (1.12) should be $\frac{3+r}{4}$. We also apply the proposed method to solve the LASSO and total variation denoising problems to illustrate its feasibility and efficiency.

The rest of this paper is organized as follows. In Section 2, we summarize some preliminaries that are useful for further analysis. Then, we propose an indefinite proximal generalized ADMM and prove some lemmas for the main convergence analysis in Section 3. The global convergence and the $O(1/t)$ convergence rate are conducted when $\tau \in (\frac{3+r}{4}, 1)$ in Section 4. In Section 5, we give a simple linear programming example to show the optimality of the parameter τ . Moreover, some numerical experiments are presented to show the efficiency of the IPG-ADMM in Section 6. Finally, we draw some conclusions in Section 7.

2 Preliminaries

In this section, we provide some preliminaries which are useful in later analysis.

2.1 Basic properties and definitions

In this paper, we use $\|\cdot\|$ to represent the Euclidean norm. If D is a $n \times n$ symmetric positive definite matrix and y is a vector in \mathfrak{R}^n , we use the notation $\|y\|_D^2$ to denote $y^T D y$. Also, we slightly abuse the notation $\|y\|_{D_0}^2 := y^T D_0 y$ when D_0 is not positive definite. With a little abuse of notation, the columnwise adhesion of column vectors x , y , and λ , i.e., $(x^T, y^T, \lambda^T)^T$, is often denoted by (x, y, λ) whenever it will not incur any confusion.

We give a simple but useful lemma in the following.

Lemma 2.1. *For the vectors a, b, c, d in \mathfrak{R}^n and a symmetric positive definite matrix $H \in \mathfrak{R}^{n \times n}$, we have the identity*

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

2.2 Variational Inequality Characterization of (1.1)

The Lagrangian function of the problem (1.1) is

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

which is defined on $\mathcal{X} \times \mathcal{Y} \times \mathfrak{R}^m$. In (2.1), (x, y) and λ are primal and dual variables, respectively. We call $((x^*, y^*), \lambda^*) \in \mathcal{X} \times \mathcal{Y} \times \mathfrak{R}^m$ a saddle point of $\mathcal{L}(x, y, \lambda)$ if the following inequalities are satisfied:

$$\mathcal{L}_{\lambda \in \mathfrak{R}^m}(x^*, y^*, \lambda) \leq \mathcal{L}(x^*, y^*, \lambda^*) \leq \mathcal{L}_{x \in \mathcal{X}, y \in \mathcal{Y}}(x, y, \lambda^*),$$

which is equivalent to the following variational inequality system:

$$\begin{cases} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x - x^*)^T (-A^T \lambda^*) \geq 0, \quad \forall x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) + (y - y^*)^T (-B^T \lambda^*) \geq 0, \quad \forall y \in \mathcal{Y}, \\ \lambda^* \in \mathfrak{R}^m, & (\lambda - \lambda^*)^T (Ax^* + By^* - b) \geq 0, \quad \forall \lambda \in \mathfrak{R}^m. \end{cases} \quad (2.2)$$

This can be written in a compact form:

$$\text{VI}(\Omega, F, \theta) \quad w^* \in \Omega, \quad \theta(w) - \theta(w^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega, \quad (2.3a)$$

where

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T \lambda \\ -B^T \lambda \\ Ax + By - b \end{pmatrix}, \quad (2.3b)$$

$$\theta(u) = \theta_1(x) + \theta_2(y), \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \mathfrak{R}^m.$$

Throughout this paper, we use Ω^* to denote the solution set of $\text{VI}(\Omega, F, \theta)$, and Ω^* is nonempty under the previous assumption.

3 Indefinite proximal generalized ADMM

In the PID-SADMM (1.11), the parameter τ should be larger than $\frac{r^2-r+4}{r^2-2r+5}$ so that the convergence can be guaranteed. However, a large value of τ in (1.11c) usually leads to slow convergence in practical computation. In this paper, we will reduce the lower bound of τ to $\frac{3+r}{4}$. Specifically, The indefinite proximal generalized ADMM (IPG-ADMM) can be read as follows

$$\begin{cases} x^{k+1} = \arg \min \{ \theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2} \|Ax + By^k - b\|^2 \mid x \in \mathcal{X} \}, & (3.1a) \end{cases}$$

$$\begin{cases} \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), & (3.1b) \end{cases}$$

$$\begin{cases} y^{k+1} = \arg \min \{ \theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2} \|Ax^{k+1} + By - b\|^2 + \frac{1}{2} \|y - y^k\|_{D_0}^2 \mid y \in \mathcal{Y} \}, & (3.1c) \end{cases}$$

$$\begin{cases} \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b), & (3.1d) \end{cases}$$

where

$$D_0 = \tau\rho I - \beta B^T B, \quad \rho > \beta \|B^T B\| \quad \text{and} \quad \tau \in \left(\frac{3+r}{4}, 1 \right). \quad (3.2)$$

It is easy to verify that D_0 in (3.1c) is not necessarily positive definite, which is the reason why we call (3.1) indefinite proximal generalized ADMM. From the definition of D in (1.6), it also holds that

$$D_0 = \tau D - (1 - \tau)\beta B^T B, \quad (3.3)$$

Note that the IPG-ADMM (3.1) reduces to the IP-LADMM (1.9) when $r = 0$.

3.1 A prediction-correction interpretation

In this subsection, we show that the scheme (3.1) can be reformulated to a prediction-correction framework by introducing some auxiliary variables, which plays a vital role in the convergence analysis for (3.1).

The variable x plays an intermediate role in the sense that x^k is not involved in the iteration in (3.1). Therefore, as [1], we still call x an intermediate variable and (y, λ) essential variables because they are essentially needed in the iteration. To distinguish their roles, accompanied with the notation in (2.3b), we additionally define the notations

$$v = \begin{pmatrix} y \\ \lambda \end{pmatrix}, \quad \mathcal{V} = \mathcal{Y} \times \mathfrak{R}^m \quad \text{and} \quad \mathcal{V}^* = \{(y^*, \lambda^*) \mid (x^*, y^*, \lambda^*) \in \Omega^*\}. \quad (3.4)$$

Moreover, for the iterate $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ generated by the indefinite proximal generalized ADMM (3.1), we define an auxiliary vector $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ as

$$\tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1}, \quad (3.5a)$$

and

$$\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b). \quad (3.5b)$$

In the following, we show some properties for the auxiliary vector \tilde{w}^k defined in (3.5).

Lemma 3.1. *For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by IPG-ADMM (3.1) and \tilde{w}^k be defined by (3.5). Then, we have*

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \quad (3.6a)$$

where

$$Q = \begin{pmatrix} \tau\rho I & -rB^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix}. \quad (3.6b)$$

Proof. Using the notation defined in (3.5), we can rewrite $\lambda^{k+\frac{1}{2}}$ in (3.1b) as

$$\lambda^{k+\frac{1}{2}} = \lambda^k - r(\lambda^k - \tilde{\lambda}^k) = \tilde{\lambda}^k + (r-1)(\tilde{\lambda}^k - \lambda^k). \quad (3.7)$$

The optimality condition of the x -subproblem of (3.1) is

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

It follows from the definition of \tilde{w}^k in (3.5) that

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T (-A^T \tilde{\lambda}^k) \geq 0, \quad \forall x \in \mathcal{X}. \quad (3.8a)$$

Substituting the relation (3.7) into the y -subproblem (3.1c), we have

$$\tilde{y}^k = \arg \min \{ \theta_2(y) - (\tilde{\lambda}^k + (r-1)(\tilde{\lambda}^k - \lambda^k))^T B y + \frac{\beta}{2} \|A\tilde{x}^k + B y - b\|^2 + \frac{1}{2} \|y - y^k\|_{D_0}^2 \mid y \in \mathcal{Y} \}.$$

Consequently, by using (3.2), we obtain

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \left\{ \begin{array}{l} -B^T (\tilde{\lambda}^k + (r-1)(\tilde{\lambda}^k - \lambda^k)) + \\ \beta B^T (A\tilde{x}^k + B\tilde{y}^k - b) + (\tau\rho I - \beta B^T B)(\tilde{y}^k - y^k) \end{array} \right\} \geq 0, \quad \forall y \in \mathcal{Y}.$$

Now, we treat the $\{\cdot\}$ term in the last inequality. Using $\beta(A\tilde{x}^k + B\tilde{y}^k - b) = -(\tilde{\lambda}^k - \lambda^k)$ (see (3.5b)), we obtain

$$\begin{aligned} & -B^T (\tilde{\lambda}^k + (r-1)(\tilde{\lambda}^k - \lambda^k)) + \beta B^T (A\tilde{x}^k + B\tilde{y}^k - b) + (\tau\rho I - \beta B^T B)(\tilde{y}^k - y^k) \\ &= -B^T (\tilde{\lambda}^k + (r-1)(\tilde{\lambda}^k - \lambda^k)) + \beta B^T B(\tilde{y}^k - y^k) \\ & \quad + \beta B^T (A\tilde{x}^k + B y^k - b) + (\tau\rho I - \beta B^T B)(\tilde{y}^k - y^k) \\ &= -B^T \tilde{\lambda}^k - (r-1)B^T (\tilde{\lambda}^k - \lambda^k) + \tau\rho(\tilde{y}^k - y^k) - B^T (\tilde{\lambda}^k - \lambda^k) \\ &= -B^T \tilde{\lambda}^k + \tau\rho(\tilde{y}^k - y^k) - rB^T (\tilde{\lambda}^k - \lambda^k). \end{aligned}$$

As a result, with the notations w^k and \tilde{w}^k in (3.5), the optimality condition of the y -subproblem (3.1c) can be written as

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

According to the definition of \tilde{w}^k in (3.5), we have

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

and it can be written as

$$\tilde{\lambda}^k \in \mathfrak{R}^m, \quad (\lambda - \tilde{\lambda}^k)^T \{(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \quad \forall \lambda \in \mathfrak{R}^m. \quad (3.8c)$$

Combining (3.8a), (3.8b) and (3.8c), and using the notations in (2.3), we obtain the assertion (3.6) immediately. This completes the proof. \square

Using the auxiliary variable $\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b)$ in (3.5b), the right-hand side of (3.6a) only involves the essential variables v , which makes us able to merely focus on the essential variables for establishing convergence for the scheme (3.1).

Lemma 3.2. *For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by IPG-ADMM (3.1) and \tilde{w}^k be defined by (3.5). Then, we have*

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \quad (3.9a)$$

where

$$M = \begin{pmatrix} I & 0 \\ -\beta B & (r+1)I_m \end{pmatrix}. \quad (3.9b)$$

Proof. It follows from (3.1) that

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

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

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

The proof is completed. \square

Therefore, based on the previous two lemmas, the scheme (3.1) can be interpreted as a prediction-correction reformulation which consists of the prediction step (3.6) and the correction step (3.9). We thus also frequently call \tilde{w}^k and w^{k+1} the predictor and corrector respectively in our proof. However, the prediction-correction reformulation only serves for the convergence analysis and it is unnecessary to divide two stages in practical implementation for the IPG-ADMM (3.1).

3.2 Basic properties

Let us define a matrix as

$$H = \begin{pmatrix} \tau\rho I - \frac{r}{r+1}\beta B^T B & -\frac{r}{r+1}B^T \\ -\frac{r}{r+1}B & \frac{1}{(r+1)\beta}I_m \end{pmatrix}. \quad (3.10)$$

For the matrices Q and M defined in (3.6b) and (3.9b) respectively, it holds that $Q = HM$. In the following, we show that the matrix H is positive definite under some mild assumptions.

Lemma 3.3. *The matrix H defined in (3.10) is positive definite if*

$$\tau > \frac{3+r}{4} \quad \text{with } r \in (-1, 1) \quad (3.11)$$

and the matrix B in (1.1) is full column rank.

Proof. For any $\tau > \frac{3+r}{4}$ with $r \in (-1, 1)$, we have $\tau > r$. Then with the definition of H in (3.10), we know that

$$\begin{aligned} H &= \begin{pmatrix} \tau\rho I - \frac{r}{r+1}\beta B^T B & -\frac{r}{r+1}B^T \\ -\frac{r}{r+1}B & \frac{1}{(r+1)\beta}I_m \end{pmatrix} \\ &\succ \begin{pmatrix} r\beta B^T B - \frac{r}{r+1}\beta B^T B & -\frac{r}{r+1}B^T \\ -\frac{r}{r+1}B & \frac{1}{(r+1)\beta}I_m \end{pmatrix} = \begin{pmatrix} \frac{r^2}{r+1}\beta B^T B & -\frac{r}{r+1}B^T \\ -\frac{r}{r+1}B & \frac{1}{(r+1)\beta}I_m \end{pmatrix} \\ &= \frac{1}{r+1} \begin{pmatrix} \sqrt{\beta}B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}}I \end{pmatrix} \begin{pmatrix} r^2 I & -rI \\ -rI & I \end{pmatrix} \begin{pmatrix} \sqrt{\beta}B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}}I \end{pmatrix}. \end{aligned} \quad (3.12)$$

Since the matrix B in (1.1) is full column rank and

$$\begin{pmatrix} r^2 & -r \\ -r & 1 \end{pmatrix}$$

is positive semi-definite, the positive definiteness of H follows immediately. \square

In the following, we will handle the right-hand side of (3.6a) and it makes us better understand the convergence proof for the scheme (3.1).

Lemma 3.4. *For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by IPG-ADMM (3.1) and \tilde{w}^k be defined by (3.5). Then, we have*

$$\begin{aligned} &\theta(u) - \theta(\tilde{w}^k) + (w - \tilde{w}^k)^T F(w) \\ &\geq \frac{1}{2} \left(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2 \right) + \frac{1}{2} \|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega, \end{aligned} \quad (3.13)$$

where H is defined in (3.10) and

$$G = Q^T + Q - M^T H M. \quad (3.14)$$

Proof. First, it follows from the definition of F in (2.3b) that

$$(w - \tilde{w})^T F(\tilde{w}) = (w - \tilde{w})^T F(w). \quad (3.15)$$

Substituting it into (3.6a) and using the relation $Q = HM$, we get

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \quad \forall w \in \Omega. \quad (3.16)$$

Applying the Lemma 2.1 to the right-hand side in (3.16) with

$$a = v, \quad b = \tilde{v}^k, \quad c = v^k, \quad \text{and} \quad d = v^{k+1},$$

we obtain

$$(v - \tilde{v}^k)^T H(v^k - v^{k+1}) = \frac{1}{2} \left(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2 \right) + \frac{1}{2} \left(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \right). \quad (3.17)$$

Substituting (3.17) into the right-hand side of (3.16), we obtain

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \\ & \geq \frac{1}{2} \left(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2 \right) + \frac{1}{2} \left(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \right), \quad \forall w \in \Omega. \end{aligned} \quad (3.18)$$

For the last term of the right-hand side of (3.18), we have

$$\begin{aligned} & \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ & = \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ & \stackrel{(3.9a)}{=} \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ & = 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ & \stackrel{(3.10)}{=} (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k). \end{aligned}$$

Substituting this equation into (3.18) and using the definition of G in (3.14), the assertion of this theorem is proved. \square

Theorem 3.5. For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by IPG-ADMM (3.1) and \tilde{w}^k be defined by (3.5). Then, we have

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2. \quad (3.19)$$

Proof. By setting $w = w^* \in \Omega^*$ in (3.13) and using (2.3a), we get

$$\begin{aligned} \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 & \geq \|v^k - \tilde{v}^k\|_G^2 + 2[\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*)] \\ & \geq \|v^k - \tilde{v}^k\|_G^2. \end{aligned}$$

The assertion (3.19) follows immediately. \square

Now we investigate the matrix G defined in (3.14). Since $HM = Q$ and $M^T HM = M^T Q$, we have

$$M^T HM = \begin{pmatrix} I & -\beta B^T \\ 0 & (r+1)I_m \end{pmatrix} \begin{pmatrix} \tau\rho I & -rB^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix} = \begin{pmatrix} \tau\rho I + \beta B^T B & -(r+1)B^T \\ -(r+1)B & \frac{1}{\beta}(r+1)I_m \end{pmatrix}.$$

Using (3.6b) and the above equation, we have

$$\begin{aligned}
G &= (Q^T + Q) - M^T H M \\
&= \begin{pmatrix} 2\tau\rho I & -(r+1)B^T \\ -(r+1)B & \frac{2}{\beta}I_m \end{pmatrix} - \begin{pmatrix} \tau\rho I + \beta B^T B & -(r+1)B^T \\ -(r+1)B & \frac{1}{\beta}(r+1)I_m \end{pmatrix} \\
&= \begin{pmatrix} D_0 & 0 \\ 0 & \frac{1-r}{\beta}I_m \end{pmatrix}. \tag{3.20}
\end{aligned}$$

Since the matrix D_0 is indefinite, G is not necessarily positive definite.

When G in (3.14) is positive definite, together with the positive definiteness of H , it is relatively easy to prove the global convergence and $O(1/t)$ convergence rate as that in [19, 20, 26]. Since G is not necessarily positive definite, we need further to analyze the properties of $\|w^k - \tilde{w}^k\|_G^2$. In the following, motivated by the techniques in [21, 22, 23, 24, 25], we obtain a lower bound of $\|w^k - \tilde{w}^k\|_G^2$ which are useful for proving the convergence and the convergence rate of IPG-ADMM (3.1).

Lemma 3.6. *For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by IPG-ADMM (3.1) and \tilde{w}^k be defined by (3.5). Then, we have*

$$\begin{aligned}
&(v^k - \tilde{v}^k)^T G (v^k - \tilde{v}^k) \\
&= \tau \|y^k - y^{k+1}\|_D^2 + (\tau - r)\beta \|B(y^k - y^{k+1})\|^2 + (1 - r)\beta \|Ax^{k+1} + By^{k+1} - b\|^2 \\
&\quad + 2(1 - r)\beta (Ax^{k+1} + By^{k+1} - b)^T B (y^k - y^{k+1}). \tag{3.21}
\end{aligned}$$

Proof. According to (3.20), $v = (y, \lambda)$ and $\tilde{v}^k = y^{k+1}$, we have

$$(v^k - \tilde{v}^k)^T G (v^k - \tilde{v}^k) = \tau \|y^k - y^{k+1}\|_D^2 - (1 - \tau)\beta \|B(y^k - y^{k+1})\|^2 + \frac{1 - r}{\beta} \|\lambda^k - \tilde{\lambda}^k\|^2. \tag{3.22}$$

Notice that(see (3.5b))

$$\lambda^k - \tilde{\lambda}^k = \beta (Ax^{k+1} + By^k - b) = \beta \{ (Ax^{k+1} + By^{k+1} - b) + B(y^k - y^{k+1}) \}.$$

Thus, we have

$$\begin{aligned}
\|\lambda^k - \tilde{\lambda}^k\|^2 &= \beta^2 \|(Ax^{k+1} + By^{k+1} - b) + B(y^k - y^{k+1})\|^2 \\
&= \beta^2 \|Ax^{k+1} + By^{k+1} - b\|^2 + 2\beta^2 (Ax^{k+1} + By^{k+1} - b)^T B (y^k - y^{k+1}) \\
&\quad + \beta^2 \|B(y^k - y^{k+1})\|^2. \tag{3.23}
\end{aligned}$$

Substituting (3.23) into the right-hand side of (3.22), we obtain (3.21) immediately. \square

For the right-hand side of (3.21), there are some quadratic terms involving two consecutive iteration points and it is easy to manipulate them in the convergence proof. Now, we handle the crossing term in the right-hand side of (3.21).

Lemma 3.7. *For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by IPG-ADMM (3.1) and \tilde{w}^k be defined by (3.5). Then, we have*

$$\begin{aligned}
&(y^k - y^{k+1})^T B^T (Ax^{k+1} + By^{k+1} - b) \\
&\geq \frac{\tau}{2(1+r)\beta} \left(\|y^k - y^{k+1}\|_D^2 - \|y^{k-1} - y^k\|_D^2 \right) + \frac{1 - \tau}{2(1+r)} \left(\|B(y^k - y^{k+1})\|^2 \right. \\
&\quad \left. - \|B(y^{k-1} - y^k)\|^2 \right) - \frac{2(1 - \tau) + r}{1 + r} \|B(y^k - y^{k+1})\|^2. \tag{3.24}
\end{aligned}$$

Proof. From the optimality condition of the y -subproblem (3.1c), for all $y \in \mathcal{Y}$, it holds that $y^{k+1} \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \begin{pmatrix} -B^T \lambda^{k+\frac{1}{2}} + \beta B^T (Ax^{k+1} + By^{k+1} - b) \\ + D_0(y^{k+1} - y^k) \end{pmatrix} \geq 0. \quad (3.25)$$

Analogously, for the previous iteration, for any $y \in \mathcal{Y}$ and $y^k \in \mathcal{Y}$, it holds that

$$\theta_2(y) - \theta_2(y^k) + (y - y^k)^T \begin{pmatrix} -B^T \lambda^{k-\frac{1}{2}} + \beta B^T (Ax^k + By^k - b) \\ + D_0(y^k - y^{k-1}) \end{pmatrix} \geq 0. \quad (3.26)$$

Setting $y = y^k$ and $y = y^{k+1}$ in (3.25) and (3.26), respectively, and then adding them, we get

$$(y^k - y^{k+1})^T \begin{pmatrix} B^T (\lambda^{k-\frac{1}{2}} - \lambda^{k+\frac{1}{2}}) - \beta B^T (Ax^k + By^k - b) + \beta B^T (Ax^{k+1} + By^{k+1} - b) \\ + D_0(y^{k+1} - y^k) - D_0(y^k - y^{k-1}) \end{pmatrix} \geq 0. \quad (3.27)$$

Note that in the k -th iteration (see (3.1b)), we have

$$\lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), \quad (3.28)$$

and in the previous iteration (see (3.1d)),

$$\lambda^k = \lambda^{k-\frac{1}{2}} - \beta(Ax^k + By^k - b). \quad (3.29)$$

It follows from (3.28) and (3.29) that

$$\begin{aligned} \lambda^{k-\frac{1}{2}} - \lambda^{k+\frac{1}{2}} &= r\beta(Ax^{k+1} + By^k - b) + \beta(Ax^k + By^k - b) \\ &= r\beta(Ax^{k+1} + By^{k+1} - b) + r\beta B(y^k - y^{k+1}) + \beta(Ax^k + By^k - b). \end{aligned} \quad (3.30)$$

Substituting (3.30) into (3.27), with a simple manipulation, we have

$$(y^k - y^{k+1})^T \begin{pmatrix} (1+r)\beta B^T (Ax^{k+1} + By^{k+1} - b) + r\beta B^T B(y^k - y^{k+1}) \\ + D_0(y^{k+1} - y^k) - D_0(y^k - y^{k-1}) \end{pmatrix} \geq 0.$$

Thus,

$$\begin{aligned} &(y^k - y^{k+1})^T B^T (Ax^{k+1} + By^{k+1} - b) \\ &\geq \frac{1}{(1+r)\beta} (y^k - y^{k+1})^T D_0 [(y^k - y^{k+1}) - (y^{k-1} - y^k)] - \frac{r}{1+r} \|B(y^k - y^{k+1})\|^2. \end{aligned} \quad (3.31)$$

Besides, note that $D_0 = \tau D - (1-\tau)\beta B^T B$, and using the Cauchy-Schwarz inequality, we get

$$\begin{aligned} &(y^k - y^{k+1})^T D_0 [(y^k - y^{k+1}) - (y^{k-1} - y^k)] \\ &= (y^k - y^{k+1})^T (\tau D - (1-\tau)\beta B^T B) [(y^k - y^{k+1}) - (y^{k-1} - y^k)] \\ &\geq \tau \|y^k - y^{k+1}\|_D^2 - (1-\tau)\beta \|B(y^k - y^{k+1})\|^2 - \frac{\tau}{2} \|y^{k-1} - y^k\|_D^2 - \frac{\tau}{2} \|y^k - y^{k+1}\|_D^2 \\ &\quad - \frac{1-\tau}{2} \beta \|B(y^k - y^{k+1})\|^2 - \frac{1-\tau}{2} \beta \|B(y^{k-1} - y^k)\|^2 \\ &= \frac{\tau}{2} \left(\|y^k - y^{k+1}\|_D^2 - \|y^{k-1} - y^k\|_D^2 \right) - \frac{3(1-\tau)}{2} \beta \|B(y^k - y^{k+1})\|^2 - \frac{1-\tau}{2} \beta \|B(y^{k-1} - y^k)\|^2. \end{aligned}$$

Substituting the above inequality into (3.31), we obtain the assertion (3.24) immediately. This completes the proof. \square

Applying the Cauchy-Schwarz inequality, we can easily derive another evaluation for the crossing term in the right-hand side of (3.21) in the following lemma, whose proof is similar to that in [22, Lemma 4.4].

Lemma 3.8. *For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by IPG-ADMM (3.1) and \tilde{w}^k be defined by (3.5). Then, for a constant $\mu > \frac{1}{4}$ and any $\delta \in (0, \frac{4\mu-1}{4\mu})$, we have*

$$\begin{aligned} & (y^k - y^{k+1})^T B^T (Ax^{k+1} + By^{k+1} - b) \\ & \geq - \left(\mu\delta + \frac{1}{4} \right) \|B(y^k - y^{k+1})\|^2 - (1 - \delta) \|Ax^{k+1} + By^{k+1} - b\|^2. \end{aligned} \quad (3.32)$$

Based on the previous lemmas, we obtain a lower bound of the term $(v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k)$ summarized in the following.

Lemma 3.9. *For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by IPG-ADMM (3.1) and \tilde{w}^k be defined by (3.5). Then, for $\delta_0 = \frac{(3-r)(4\tau-r-3)}{2(1-r)(5+r)}$, we have*

$$\begin{aligned} & (v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k) \\ & \geq \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^k - y^{k+1}\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^k - y^{k+1})\|^2 \right) \\ & \quad - \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^{k-1} - y^k)\|^2 \right) \\ & \quad + \tau \|y^k - y^{k+1}\|_D^2 + (1-r)\delta_0\beta (\|B(y^k - y^{k+1})\|^2 + \|Ax^{k+1} + By^{k+1} - b\|^2). \end{aligned} \quad (3.33)$$

Proof. Since $\tau \in (\frac{3+r}{4}, 1)$ and $r \in (-1, 1)$, we can take $\mu = \frac{3-r}{2+2r}$ in Lemma 3.8. It follows that $\mu > \frac{1}{4}$ and

$$\delta_0 = \frac{(3-r)(4\tau-r-3)}{2(1-r)(5+r)} \in \left(0, \frac{4\mu-1}{4\mu} \right).$$

Then, adding the two inequalities (3.24) and (3.32) with $\delta = \delta_0$, we get

$$\begin{aligned} & 2(y^k - y^{k+1})^T B^T (Ax^{k+1} + By^{k+1} - b) \\ & \geq \frac{\tau}{2(1+r)\beta} \left(\|y^k - y^{k+1}\|_D^2 - \|y^{k-1} - y^k\|_D^2 \right) \\ & \quad + \frac{1-\tau}{2(1+r)} \left(\|B(y^k - y^{k+1})\|^2 - \|B(y^{k-1} - y^k)\|^2 \right) \\ & \quad - \left(\frac{2(1-\tau)+r}{1+r} + \frac{2\delta_0+1}{4} \right) \|B(y^k - y^{k+1})\|^2 - (1 - \delta_0) \|Ax^{k+1} + By^{k+1} - b\|^2. \end{aligned}$$

Substituting the above inequality into (3.21), with some manipulations, we have

$$\begin{aligned} & (v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k) \\ & \geq \tau \|y^k - y^{k+1}\|_D^2 + \frac{\tau(1-r)}{2(1+r)} (\|y^k - y^{k+1}\|_D^2 - \|y^{k-1} - y^k\|_D^2) \\ & \quad + \frac{(1-\tau)(1-r)}{2(1+r)} \beta (\|B(y^k - y^{k+1})\|^2 - \|B(y^{k-1} - y^k)\|^2) \\ & \quad + (1-r)\delta_0\beta (\|B(y^k - y^{k+1})\|^2 + \|Ax^{k+1} + By^{k+1} - b\|^2). \end{aligned}$$

It follows that the assertion (3.33) holds immediately. This completes the proof. \square

4 Convergence analysis

In this section, we analyze the global convergence and convergence rate for IPG-ADMM (3.1).

4.1 Global convergence

The following corollaries are obtained directly from Lemma 3.4, Lemma 3.9 and Theorem 3.5.

Corollary 4.1. *For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by IPG-ADMM (3.1) and \tilde{w}^k be defined by (3.5). Then, for $\delta_0 = \frac{(3-r)(4\tau-r-3)}{2(1-r)(5+r)}$, we have*

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \\ & \geq \frac{1}{2} \left[\|v - v^{k+1}\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^k - y^{k+1}\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^k - y^{k+1})\|^2 \right) \right] \\ & \quad - \frac{1}{2} \left[\|v - v^k\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^{k-1} - y^k)\|^2 \right) \right] \\ & \quad + \frac{\tau}{2} \|y^k - y^{k+1}\|_D^2 + \frac{(1-r)\delta_0}{2} \beta \left(\|B(y^k - y^{k+1})\|^2 + \|Ax^{k+1} + By^{k+1} - b\|^2 \right). \end{aligned} \quad (4.1)$$

Corollary 4.2. *For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by IPG-ADMM (3.1) and \tilde{w}^k be defined by (3.5). Then, for $\delta_0 = \frac{(3-r)(4\tau-r-3)}{2(1-r)(5+r)}$, we have*

$$\begin{aligned} & \|v^{k+1} - v^*\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^k - y^{k+1}\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^k - y^{k+1})\|^2 \right) \\ & \leq \|v^k - v^*\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^{k-1} - y^k)\|^2 \right) \\ & \quad - \left[\tau \|y^k - y^{k+1}\|_D^2 + (1-r)\delta_0 \beta \left(\|B(y^k - y^{k+1})\|^2 + \|Ax^{k+1} + By^{k+1} - b\|^2 \right) \right]. \end{aligned} \quad (4.2)$$

Now we present the convergence result in the following theorem.

Theorem 4.3. *For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by IPG-ADMM (3.1) and \tilde{w}^k be defined by (3.5). Then the sequence $\{v^k\}$ converges to $v^\infty \in \mathcal{V}^*$.*

Proof. For an arbitrary fixed $v^* \in \mathcal{V}^*$, it follows from the (4.2) and (3.11) that

$$\begin{aligned} & \|v^k - v^*\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^{k-1} - y^k)\|^2 \right) \\ & \leq \|v^1 - v^*\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^0 - y^1\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^0 - y^1)\|^2 \right), \quad \forall k \geq 1, \end{aligned}$$

and thus the sequence $\{v^k\}$ is bounded. It follows from (3.9) that $\{\tilde{v}^k\}$ is also bounded. Now, we rewrite (4.2) as

$$\begin{aligned} & \tau \|y^k - y^{k+1}\|_D^2 + (1-r)\delta_0 \beta \left(\|B(y^k - y^{k+1})\|^2 + \|Ax^{k+1} + By^{k+1} - b\|^2 \right) \\ & \leq \left[\|v^k - v^*\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^{k-1} - y^k)\|^2 \right) \right] \\ & \quad - \left[\|v^{k+1} - v^*\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^k - y^{k+1}\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^k - y^{k+1})\|^2 \right) \right]. \end{aligned} \quad (4.3)$$

Summarizing the inequality (4.3) over $k = 1, 2, 3, \dots$, we obtain

$$\begin{aligned} & \sum_{k=1}^{\infty} \left(\tau \|y^k - y^{k+1}\|_D^2 + (1-r)\delta_0\beta(\|B(y^k - y^{k+1})\|^2 + \|Ax^{k+1} + By^{k+1} - b\|^2) \right) \\ & \leq \|v^1 - v^*\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^0 - y^1\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^0 - y^1)\|^2 \right). \end{aligned}$$

Because D is positive definite, $r \in (-1, 1)$ and $\delta_0 > 0$, it follows from the above inequality that

$$\lim_{k \rightarrow \infty} \|y^k - y^{k+1}\|^2 = 0, \quad \lim_{k \rightarrow \infty} \|B(y^k - y^{k+1})\|^2 = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} \|Ax^{k+1} + By^{k+1} - b\|^2 = 0.$$

Applying the above relations to (3.1b) and (3.1d), we obtain

$$\begin{aligned} \lim_{k \rightarrow \infty} \|\lambda^k - \lambda^{k+1}\|^2 &= \lim_{k \rightarrow \infty} \|r\beta(Ax^{k+1} + By^k - b) + \beta(Ax^{k+1} + By^{k+1} - b)\|^2 \\ &= \lim_{k \rightarrow \infty} \|(r+1)\beta(Ax^{k+1} + By^{k+1} - b) + r\beta B(y^k - y^{k+1})\|^2 \\ &= 0. \end{aligned}$$

Thus, we have

$$\lim_{k \rightarrow \infty} \|v^k - v^{k+1}\|^2 = 0. \quad (4.4)$$

For an arbitrarily fixed $v^* \in \mathcal{V}^*$ and $\forall k \geq 1$, it follows that

$$\begin{aligned} \|v^{k+1} - v^*\|_H^2 &\leq \|v^k - v^*\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^{k-1} - y^k)\|^2 \right) \\ &\leq \|v^1 - v^*\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^0 - y^1\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^0 - y^1)\|^2 \right), \end{aligned}$$

and thus the sequence $\{v^k\}$ is bounded. Because M is non-singular, according to (3.9), $\{\tilde{v}^k\}$ is also bounded. Let v^∞ be a cluster point of $\{\tilde{v}^k\}$. Then $\{\tilde{v}^k\}$ has a subsequence $\{\tilde{v}^{k_j}\}$ converging to v^∞ . Let x^∞ be the vector accompanied with $(y^\infty, \lambda^\infty) \in \mathcal{V}$. Recalling the matrix B is assumed to be full column rank, it follows from (3.16) that

$$w^\infty \in \Omega, \quad \theta(u) - \theta(u^\infty) + (w - w^\infty)^T F(w^\infty) \geq 0, \quad \forall w \in \Omega,$$

which means w^∞ is a solution point of (2.3) and its essential part $v^\infty \in \mathcal{V}^*$. Since $v^\infty \in \mathcal{V}^*$, it follows from (4.5) that

$$\|v^{k+1} - v^\infty\|_H^2 \leq \|v^k - v^\infty\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^{k-1} - y^k)\|^2 \right).$$

Notice that v^∞ is also the limit point of $\{v^{k_j}\}$. Together with (4.4), it is impossible that the sequence $\{v^k\}$ has more than one cluster point. Thus $\{v^k\}$ converges to v^∞ and the theorem is proved. \square

4.2 Convergence rate

According to (2.3), if we find a \tilde{w} which satisfies

$$\tilde{w} \in \Omega, \quad \theta(u) - \theta(\tilde{w}) + (w - \tilde{w})^T F(\tilde{w}) \geq 0, \quad \forall w \in \Omega,$$

then \tilde{w} is a solution of (2.3). By using (3.15), the solution \tilde{w} can be also stated as

$$\tilde{w} \in \Omega, \quad \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq 0, \quad \forall w \in \Omega.$$

We use the latter equality to measure the approximation of (2.3), which means that for given $\epsilon > 0$, $\tilde{w} \in \Omega$ is an ϵ -approximate solution of VI(Ω, F, θ) if

$$\tilde{w} \in \Omega, \quad \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq -\epsilon, \quad \forall w \in \mathcal{D}(\tilde{w}),$$

where

$$\mathcal{D}(\tilde{w}) = \{w \in \Omega \mid \|w - \tilde{w}\| \leq 1\}.$$

We will show that for given $\epsilon > 0$, after t iterations, it can offer a $\tilde{w} \in \Omega$, such that

$$\tilde{w} \in \Omega \quad \text{and} \quad \sup_{w \in \mathcal{D}(\tilde{w})} \{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\} \leq \epsilon. \quad (4.5)$$

Corollary 4.1 is the base for the convergence rate proof.

Theorem 4.4. *For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by IPG-ADMM (3.1) and \tilde{w}^k be defined by (3.5). Then for any integer t , we have*

$$\begin{aligned} & \theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \\ & \leq \frac{1}{2t} \left[\|v - v^1\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^0 - y^1\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^0 - y^1)\|^2 \right) \right], \quad \forall w \in \Omega, \end{aligned} \quad (4.6a)$$

where

$$\tilde{w}_t = \frac{1}{t} \sum_{k=1}^t \tilde{w}^k. \quad (4.6b)$$

Proof. First, it follows from (4.1) that

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \\ & \geq \frac{1}{2} \left[\|v - v^{k+1}\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^k - y^{k+1}\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^k - y^{k+1})\|^2 \right) \right] \\ & \quad - \frac{1}{2} \left[\|v - v^k\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^{k-1} - y^k)\|^2 \right) \right]. \end{aligned}$$

Subsequently, we have

$$\begin{aligned} & \theta(\tilde{u}^k) - \theta(u) + (\tilde{w}^k - w)^T F(w) \\ & \quad + \frac{1}{2} \left[\|v - v^{k+1}\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^k - y^{k+1}\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^k - y^{k+1})\|^2 \right) \right] \\ & \leq \frac{1}{2} \left[\|v - v^k\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^{k-1} - y^k)\|^2 \right) \right]. \end{aligned} \quad (4.7)$$

Summing the inequality (4.7) over $k = 1, 2, \dots, t$, we obtain

$$\begin{aligned} & \sum_{k=1}^t \theta(\tilde{u}^k) - t\theta(u) + \left(\sum_{k=1}^t \tilde{w}^k - tw \right)^T F(w) \\ & \leq \frac{1}{2} \left[\|v - v^1\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^0 - y^1\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^0 - y^1)\|^2 \right) \right]. \end{aligned}$$

and consequently

$$\begin{aligned} & \frac{1}{t} \sum_{k=1}^t \theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t^k - w)^T F(w) \\ & \leq \frac{1}{2t} \left[\|v - v^1\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^0 - y^1\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^0 - y^1)\|^2 \right) \right]. \end{aligned} \quad (4.8)$$

It follows from the definition of \tilde{w}_t in (4.6b) that

$$\tilde{w}_t = \frac{1}{t} \sum_{k=1}^t \tilde{u}^k.$$

Since $\theta(u)$ is convex, it follows that

$$\theta(\tilde{w}_t) \leq \frac{1}{t} \sum_{k=1}^t \theta(\tilde{u}^k).$$

Substituting it into (4.8), the assertion (4.6a) follows directly. \square

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

$$d = \sup \left\{ \|v - v^1\|_H^2 + \frac{1-r}{1+r} \left(\frac{\tau}{2} \|y^0 - y^1\|_D^2 + \frac{1-\tau}{2} \beta \|B(y^0 - y^1)\|^2 \right) \right\},$$

where $v^0 = (y^0, \lambda^0)$ and v^1 are the initial point and the first computed iterate, respectively. Then, after t iterations of the indefinite proximal generalized ADMM (3.1), the point $\tilde{w}_t \in \Omega$ defined in (4.6b) satisfies

$$\tilde{w} \in \Omega \quad \text{and} \quad \sup_{w \in \mathcal{D}(\tilde{w})} \{ \theta(\tilde{w}) - \theta(u) + (\tilde{w} - w)^T F(w) \} \leq \frac{d}{2t},$$

which means \tilde{w}_t is an approximate solution of $\text{VI}(\Omega, F, \theta)$ with the accuracy $O(1/t)$ (recall (4.5)). That is, the convergence rate $O(1/t)$ of the IPG-ADMM (3.1) is established in an ergodic sense.

5 Optimality of τ

In Section 4, we have showed that the IPG-ADMM (3.1) is convergent when $\tau \in (\frac{3+r}{4}, 1)$. In this section, we show that the IPG-ADMM (3.1) is not necessarily convergent for any $\tau \in (0, \frac{3+r}{4})$ by a simple linear programming. Thus, $\frac{3+r}{4}$ is the optimal lower bound with $r \in (-1, 1)$ for the proximal parameter τ of the IPG-ADMM (3.1).

Let us consider the following extremely simple linear programming in [22]:

$$\min \{ 0 \cdot x + 0 \cdot y \mid 0 \cdot x + y = 0, x \in \{0\}, y \in \mathfrak{R} \}, \quad (5.1)$$

which is a special case of the model (1.1). Without loss of generality, we set $\beta = 1$ and hence the augmented Lagrangian function of (5.1) is

$$\mathcal{L}_1(x, y, \lambda) = -\lambda^T y + \frac{1}{2} y^2.$$

The scheme of the IPG-ADMM (3.1) for (5.1) is

$$\begin{cases} x^{k+1} = \arg \min \{ \mathcal{L}_1(x, y^k, \lambda^k) \mid x \in \{0\} \}, & (5.2a) \end{cases}$$

$$\begin{cases} \lambda^{k+\frac{1}{2}} = \lambda^k - ry^k, & (5.2b) \end{cases}$$

$$\begin{cases} y^{k+1} = \arg \min \{ -y^T \lambda^{k+\frac{1}{2}} + \frac{1}{2} y^2 + \frac{1}{2} \|y - y^k\|_{D_0}^2 \mid y \in \mathfrak{R} \}, & (5.2c) \end{cases}$$

$$\begin{cases} \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - y^{k+1}. & (5.2d) \end{cases}$$

Since $\beta = 1$ and $B^T B = 1$, it follows from (3.2) and (3.3) that

$$D_0 = \tau\rho - 1 \quad \text{and} \quad D = \rho - 1 \quad \text{with} \quad \rho > 1.$$

Thus the recursion (5.2) can be specified

$$\begin{cases} x^{k+1} = 0, & (5.3a) \end{cases}$$

$$\begin{cases} \lambda^{k+\frac{1}{2}} = \lambda^k - ry^k, & (5.3b) \end{cases}$$

$$\begin{cases} -\lambda^{k+\frac{1}{2}} + y^{k+1} + (\tau\rho - 1)(y^{k+1} - y^k) = 0, & (5.3c) \end{cases}$$

$$\begin{cases} \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - y^{k+1}. & (5.3d) \end{cases}$$

For any $k > 0$, we have $x^k = 0$. Thus, we only need to study the essential sequence $v^k = (y^k, \lambda^k)$. For any given $\tau \in (0, \frac{3+r}{4})$, there exists $\rho > 1$ such that $\tau\rho \leq \frac{3+r}{4}$. We set $\alpha = \tau\rho$ and the recursion (5.3) can be written as

$$\begin{cases} y^{k+1} = \frac{\alpha - 1 - r}{\alpha} y^k + \frac{1}{\alpha} \lambda^k, & (5.4a) \end{cases}$$

$$\begin{cases} \lambda^{k+1} = \frac{(r+1)(1-\alpha)}{\alpha} y^k + \frac{\alpha-1}{\alpha} \lambda^k. & (5.4b) \end{cases}$$

Its compact form is

$$v^{k+1} = P(\alpha)v^k \quad \text{with} \quad P(\alpha) = \frac{1}{\alpha} \begin{pmatrix} \alpha - 1 - r & 1 \\ (r+1)(1-\alpha) & \alpha - 1 \end{pmatrix}. \quad (5.5)$$

Let $f_1(\alpha)$ and $f_2(\alpha)$ be the two eigenvalues of matrix $P(\alpha)$. Then, we have

$$f_1(\alpha) = \frac{(2\alpha - 2 - r) + \sqrt{(r+2)^2 - 4\alpha(r+1)}}{2\alpha} \quad \text{and} \quad f_2(\alpha) = \frac{(2\alpha - 2 - r) - \sqrt{(r+2)^2 - 4\alpha(r+1)}}{2\alpha}.$$

It is easy to estimate that $f_2(\frac{3+r}{4}) = -1$ and

$$f_2'(\alpha) = \frac{1}{2\alpha^2} \left(2 + r + \frac{(r+2)^2 - 2\alpha(r+1)}{(r+2)^2 - 4\alpha(r+1)} \right) > 0, \quad \forall \alpha \in (0, \frac{3+r}{4}).$$

Thus, we have

$$f_2(\alpha) = \frac{(2\alpha - 2 - r) - \sqrt{(r+2)^2 - 4\alpha(r+1)}}{2\alpha} < -1, \quad \forall \alpha \in (0, \frac{3+r}{4}).$$

That is to say, for any $\alpha \in (0, \frac{3+r}{4})$, the matrix $P(\alpha)$ (5.5) has the eigenvalue which is less than -1 . Hence, the recursion (5.3) is not necessarily convergent for any $\tau \in (0, \frac{3+r}{4})$.

6 Numerical results

In this section, we verify the theoretical assertions analyzed in previous sections by some numerical experiments. We apply the proposed method to solve the least absolute shrinkage and selection operator (LASSO) and the total variation (TV) denoising problems. We shall verify the following assertions.

- (1) The proposed IPG-ADMM (3.1) converges quickly for a wide range of applications. The theoretical significance of the underdetermined relaxation parameter r is thus verified.
- (2) For the proposed IPG-ADMM (3.1), the convergence speed is relative to the parameter r .
- (3) The proposed IPG-ADMM (3.1) is faster than the PID-SADMM in [14] and PG-ADMM (1.5) when we fix a value of r .

We should point out that our proposed method may be not the best method for solving these models. We use merely these models to illustrate the efficiency and stability for the proposed method. All codes were written by MATLAB R2013a and all the numerical experiments were conducted on a personal computer with a 3.6GHz i7 processor and a 4GB memory.

6.1 LASSO

The LASSO model proposed in [35] can be characterized as

$$\min \left\{ \varrho \|y\|_1 + \frac{1}{2} \|By - b\|_2^2 \mid y \in \mathfrak{R}^m \right\}. \quad (6.1)$$

where $b \in \mathfrak{R}^n$ is the response vector, $B \in \mathfrak{R}^{n \times m}$ is the design matrix, n is the number of data points, m is the number of features, $\varrho > 0$ is a regularization parameter and $\|y\|_1 := \sum_{i=1}^n |y_i|$. The LASSO model provides a sparse estimation of y when there are more features than data points (i.e., $m > n$).

Now, by introducing an auxiliary variable $x \in \mathfrak{R}^n$, the model (6.1) can be reformulated as

$$\min \left\{ \frac{1}{2} \|x\|_2^2 + \varrho \|y\|_1 \mid x - By + b = 0, x \in \mathfrak{R}^n, y \in \mathfrak{R}^m \right\}, \quad (6.2)$$

which is a special case of (1.1). Thus PG-ADMM, PID-SADMM and IPG-ADMM can be applied to solve (6.2). In order to obtain the solution of y -subproblem, we use the soft-thresholding operator defined as

$$(S_\kappa(a))_i = (1 - \kappa/|a_i|)_+ \cdot a_i, \quad i = 1, \dots, m,$$

where $\kappa > 0$ and $a \in \mathfrak{R}^m$ (see [1]).

Now we specify the setting for the LASSO model to be tested. We first choose $B_{ij} \sim \mathcal{N}(0, 1)$ and then normalize the columns. We generate a random sparse vector in \mathfrak{R}^m with the density 0.02 as y , the noise vector $\epsilon \sim \mathcal{N}(0, 10^{-3}I)$ and the vector $b = By + \epsilon$, the regularization parameter is set as $\varrho = 0.1 \|B^T b\|_\infty$. Here, we use the residuals of y -subproblem and λ -subproblem to construct the stopping criteria which is described as follows:

$$tol := \max \left\{ \|\tau \rho (y^k - \tilde{y}^k) - r B^T (\lambda^k - \tilde{\lambda}^k)\|_\infty, \left\| -B(y^k - \tilde{y}^k) + \frac{1}{\beta} (\lambda^k - \tilde{\lambda}^k) \right\|_\infty \right\} \leq 10^{-3}. \quad (6.3)$$

Now we specify the choices of parameters to implement these algorithms. First, we choose $\beta = 1$ and $\rho = \beta \|B^T B\| + 0.01$ for the tested algorithms. Moreover, the initial points are $y^0 =$

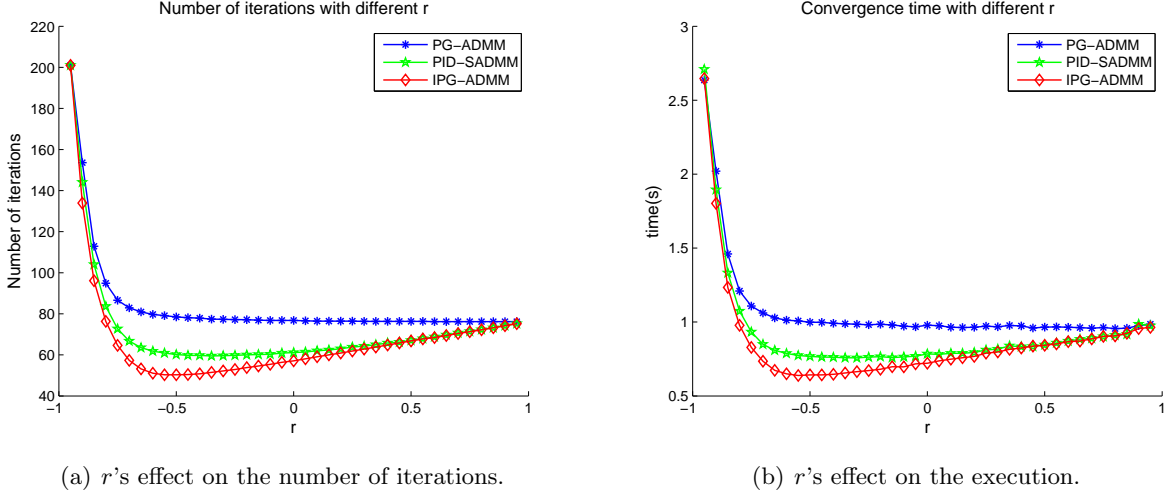


Figure 1: Sensitivity test on the factor r when $\beta = 1$.

$\mathbf{0}_{m \times 1}$ and $\lambda^0 = \mathbf{0}_{n \times 1}$. In order to investigate the stability and efficiency of our algorithm, we test 8 groups of problems with different n and m . For each group, we set $r = 0.3$ and $r = -0.3$ respectively and test 10 times. Since small values of τ are preferred in practical computation, we set $\tau = \frac{3+r}{4} + 0.01$ in the IPG-ADMM and $\tau = \frac{r^2-r+4}{r^2-2r+5}$ in the PID-SADMM. In Table 1, we report the average performances including the number of iterations (Iter.) and the CPU time (CPU(s)). We also test the sensitivity of r of three methods for the LASSO model (6.1). We first fix $n = 400, m = 1200, \beta = 1$ and choose different values of r in the interval $[-0.95, 0.95]$. More specifically, we choose $r = \{-0.95, 0.90, 0.85, \dots, 0.85, 0.90, 0.95\}$. Then, We set the maximal number of iteration is 200. The computing time and number of iterations required by the three methods are reported for each choice of r and then we plot them in Figure 1.

Table 1: Numerical results for LASSO.

| n | m | $r = 0.3$ | | | | | | $r = -0.3$ | | | | | |
|------|------|-----------|--------|-----------|--------|----------|--------|------------|--------|-----------|--------|----------|--------|
| | | PG-ADMM | | PID-SADMM | | IPG-ADMM | | PG-ADMM | | PID-SADMM | | IPG-ADMM | |
| | | Iter. | CPU(s) | Iter. | CPU(s) | Iter. | CPU(s) | Iter. | CPU(s) | Iter. | CPU(s) | Iter. | CPU(s) |
| 200 | 500 | 65.5 | 0.04 | 55.1 | 0.03 | 53.9 | 0.03 | 66.5 | 0.04 | 51.3 | 0.03 | 45.0 | 0.03 |
| 300 | 800 | 68.0 | 0.35 | 57.2 | 0.29 | 55.7 | 0.28 | 69.0 | 0.35 | 53.3 | 0.27 | 46.6 | 0.24 |
| 300 | 1000 | 91.1 | 0.66 | 76.9 | 0.55 | 75.0 | 0.54 | 92.0 | 0.66 | 71.0 | 0.51 | 62.0 | 0.46 |
| 500 | 1500 | 80.1 | 1.79 | 67.3 | 1.49 | 65.9 | 1.46 | 81.3 | 1.80 | 62.8 | 1.39 | 55.0 | 1.22 |
| 500 | 2000 | 108.0 | 3.34 | 90.7 | 2.82 | 88.6 | 2.77 | 108.3 | 3.38 | 83.6 | 2.60 | 73.1 | 2.27 |
| 800 | 2500 | 85.8 | 5.40 | 72.1 | 4.56 | 70.5 | 4.43 | 87.0 | 5.47 | 67.0 | 4.21 | 58.7 | 3.73 |
| 1000 | 3000 | 79.0 | 7.42 | 66.3 | 6.22 | 64.8 | 6.17 | 80.2 | 7.54 | 61.9 | 5.84 | 54.1 | 5.12 |
| 1500 | 5000 | 92.8 | 22.22 | 77.9 | 18.61 | 76.2 | 18.24 | 93.9 | 22.51 | 72.3 | 17.37 | 63.4 | 15.22 |

According to the results presented in Table 1, we see that IPG-ADMM can solve the problems efficiently. Moreover, the IPG-ADMM is stable and faster than the other methods for slightly larger scale problems in the experiment. From Figure 1, it is obvious to see that the factor r functions well for a wide range of values. In particular, based on our experiments, some aggressive values (e.g., $[-0.6, -0.4]$) are preferred for the LASSO model.

6.2 Total variation denoising model

In this subsection, we test the IPG-ADMM on the total variation (TV) denoising problem in [1, 32]:

$$\min \left\{ \frac{1}{2} \|y - b\|_2^2 + \eta \|Dy\|_1 \mid y \in \mathfrak{R}^m \right\}, \quad (6.4)$$

where $D : \mathfrak{R}^n \rightarrow \mathfrak{R}^n$ is the finite-difference operator in the vertical direction. By introducing a new variable x , we can reformulate (6.4) as

$$\min \left\{ \eta \|x\|_1 + \frac{1}{2} \|y - b\|_2^2 \mid x - Dy = 0, x \in \mathfrak{R}^n, y \in \mathfrak{R}^n \right\}, \quad (6.5)$$

which is also a special case of (1.1). Then the PG-ADMM, PID-SADMM and IPG-ADMM can be applied to (6.5).

Now we specify the setting for the TV model (6.4) to be tested. For a given dimension $n \times n$, We generate the data randomly as follows [1]

```

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);

```

The regularization parameter is set as $\eta = 5$. Here, we still use the stopping criteria (6.3). Now we specify the choices of parameters to implement these algorithms. First, we choose $\beta = 5$ and $\rho = \beta \|D^T D\| + 0.01$ for the tested algorithms. Moreover, the initial points are $y^0 = \mathbf{0}_{n \times 1}$ and $\lambda^0 = \mathbf{0}_{n \times 1}$. In order to investigate the stability and efficiency of our algorithm, we test 8 groups of problems with different n . For each group, we set $r = 0.3$ and $r = -0.3$ respectively and test 10 times. We also set $\tau = \frac{3+r}{4} + 0.01$ in the IPG-ADMM and $\tau = \frac{r^2-r+4}{r^2-2r+5}$ in the PID-SADMM.

In Table 2, we report the average performances including the number of iterations (Iter.) and the CPU time (CPU(s)). We also test the sensitivity of r of three methods for the TV denoising model (6.4). We first fix $n = 2000$, $\beta = 5$ and choose different values of r in the interval $[-0.95, 0.95]$ as that in last subsection. Then, we set the maximal number of iteration is 1000. The computing time and number of iterations required by the three methods are reported for each choice of r and then we plot them in Figure 2.

Table 2 shows that the IPG-ADMM is more efficient than the PG-ADMM and PID-SADMM for the TV denoising model (6.4). Moreover, the speed of algorithms are revelent to the dimension of the variables. From Figure 2, the IPG-ADMM is stable in our test, and faster than the PG-ADMM and PID-SADMM for slightly larger problems.

7 Conclusions

In this paper, we proposed a new indefinite proximal generalized ADMM, which combined the linearized generalized ADMM with the indefinite proximal term. We analyzed the convergence and the worst-case $O(1/t)$ convergence rate of the proposed method by specifying a better lower bound for

Table 2: Numerical results for TV denoising model.

| n | $r = 0.3$ | | | | | | $r = -0.3$ | | | | | |
|-------|-----------|--------|-----------|--------|----------|--------|------------|--------|-----------|--------|----------|--------|
| | PG-ADMM | | PID-SADMM | | IPG-ADMM | | PG-ADMM | | PID-SADMM | | IPG-ADMM | |
| | Iter. | CPU(s) | Iter. | CPU(s) | Iter. | CPU(s) | Iter. | CPU(s) | Iter. | CPU(s) | Iter. | CPU(s) |
| 500 | 278.0 | 0.03 | 260.7 | 0.03 | 258.7 | 0.03 | 401.5 | 0.05 | 356.6 | 0.04 | 339.9 | 0.04 |
| 1000 | 325.9 | 0.06 | 297.4 | 0.05 | 294.6 | 0.05 | 464.8 | 0.08 | 422.0 | 0.08 | 402.3 | 0.07 |
| 2000 | 414.9 | 0.12 | 380.8 | 0.11 | 376.7 | 0.11 | 585.3 | 0.17 | 533.2 | 0.16 | 497.9 | 0.15 |
| 3000 | 449.5 | 0.20 | 423.9 | 0.19 | 418.9 | 0.19 | 683.8 | 0.31 | 602.7 | 0.27 | 573.6 | 0.26 |
| 5000 | 488.7 | 0.33 | 456.0 | 0.30 | 452.3 | 0.30 | 730.0 | 0.48 | 662.7 | 0.44 | 621.7 | 0.41 |
| 6000 | 488.1 | 0.38 | 456.0 | 0.35 | 452.1 | 0.35 | 720.7 | 0.56 | 642.0 | 0.50 | 612.7 | 0.47 |
| 8000 | 598.3 | 0.59 | 558.2 | 0.54 | 553.9 | 0.54 | 869.5 | 0.85 | 774.7 | 0.75 | 729.9 | 0.72 |
| 10000 | 609.6 | 0.72 | 561.9 | 0.66 | 556.7 | 0.66 | 877.0 | 1.03 | 791.0 | 0.93 | 748.7 | 0.88 |

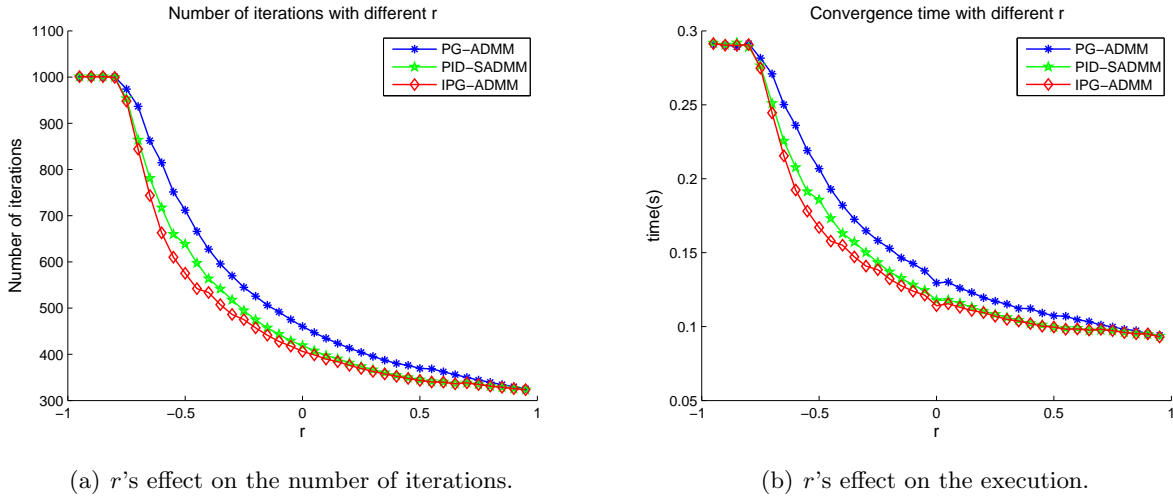


Figure 2: Sensitivity test on the factor r when $\beta = 5$.

the proximal parameter. Moreover, by showing a special example of linear programming, we illustrated that the proximal parameter of the proposed method is optimal. We also gave some primal numerical experiments to demonstrate the efficiency and the advantage of the proposed method.

Acknowledgement. The authors are deeply grateful to Professor Bingsheng He for his help during the completion of this work.

References

- [1] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends® Mach. Learn.* **3**(1), 1–122(2011)
- [2] Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202(2009)
- [3] Candès, E.J., Recht, B.: Exact matrix completion via convex optimization. *Found. Comput. Math.* **9**(6), 717–772(2009)

- [4] Chen, S., Donoho, D., Saunders, M.: Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20**(1), 33–61(1998)
- [5] Combettes, P.L., Pesquet, J.C.: A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE J-STSP.* **1**(4), 564–574(2008)
- [6] Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*. Springer New York. 185–212(2011)
- [7] Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forwardbackward splitting. *SIAM J. Multiscale Model. Simul.* **4**(4), 1168–1200(2005)
- [8] Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* **55**(1), 293–318(1992)
- [9] Eckstein, J., Wang, Y.: Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. *Pac. J. Optim.* **11**(4), 619–644(2015)
- [10] Fang, E.X., He, B.S., Liu, H., Yuan, X.M.: Generalized alternating direction method of multipliers: new theoretical insights and applications. *Math. Prog. Comp.* **7**(2), 149–187(2015)
- [11] Fadili, M.J., Starck, J.L.: Monotone operator splitting for optimization problems in sparse recovery. *IEEE ICIP.* 1461–1464(2009)
- [12] Figueiredo, M.A.T., Bioucas-Dias, J.M.: Restoration of poissonian images using alternating direction optimization. *IEEE T. Image Process.* **19**(12), 3133C3145(2010)
- [13] Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.* **2**(1), 17–40(1976)
- [14] Gao, B., Ma, F.: Symmetric ADMM with positive-indefinite proximal regularization for linearly constrained convex optimization Available on <http://www.optimization-online.org>(2016)
- [15] Glowinski, R.: On alternating direction methods of multipliers: a historical perspective. *Model. Simul. Optim. Sci. Technol. Comput. Methods Appl. Sci.* **34**, 59–82(2014)
- [16] Glowinski, R., Marrocco, A.: 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. Inf. Rech. Opér., Anal. Numér.* **2**, 41–76(1975)
- [17] Gu, Y., Jiang, B., Han, D.R.: A semi-proximal-based strictly contractive Peaceman-Rachford splitting method. Available on <http://www.optimization-online.org>(2015)
- [18] He, B.S.: A new method for a class of linear variational inequalities. *Math. Program.* **66**(1), 137–144(1994)
- [19] He, B.S.: PPA-like contraction methods for convex optimization: a framework using variational inequality approach. *J. Oper. Res. Soc. China* **3**(4), 391–420(2015)
- [20] He, B.S., Liu, H., Wang, Z.R., Yuan, X.M.: A strictly contractive Peaceman-Rachford splitting method for convex programming. *SIAM J. Optim.* **24**(3), 1011–1040(2014)

- [21] He, B.S., Ma, F., Yuan, X.M.: Linearized alternating direction method of multipliers via positive-indefinite proximal regularization for convex programming. Available on <http://www.optimization-online.org>(2016)
- [22] He, B.S., Ma, F., Yuan, X.M.: Optimal linearized alternating direction method of multipliers for convex programming Available on <http://www.optimization-online.org>(2017)
- [23] He, B.S., Yuan, X.M.: Improving an ADMM-like splitting method via positive-indefinite proximal regularization for three-block separable convex minimization. Available on <http://www.optimization-online.org>(2016)
- [24] He, B.S., Ma, F., Yuan, X.M.: Positive-indefinite proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems. Available on <http://www.optimization-online.org>(2016)
- [25] He, B.S., Yang, H.: Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities. *Oper. Res. Lett.* **23**(3), 151–161(1998)
- [26] He, B.S., Yuan, X.M.: On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. *Numer. Math.* **130**(3), 567–577(2015)
- [27] Li, M., Li, X.X., Yuan, X.M.: Convergence analysis of the generalized alternating direction method of multipliers with logarithmic Quadratic proximal regularization. *J. Optim. Theory Appl.* **164**, 218–233(2015)
- [28] Li, M., Sun, D.F., Toh, K.C.: A majorized ADMM with indefinite proximal term for linearly constrained convex composite optimization. *SIAM J. Optim.* **26**(2), 922–950(2016)
- [29] Li, X.X., Mo, L.L., Yuan, X.M., Zhang, J.Z.: Linearized alternating direction method of multipliers for sparse group and fused LASSO models. *Comput. Statist. Data Anal.* **79**(79), 203–221(2014)
- [30] Li, X.X., Yuan, X.M.: A proximal strictly contractive Peaceman-Rachford splitting method for convex programming with applications to imaging *SIAM J. Imaging Sci.* **8**(2), 1332–1365(2015)
- [31] Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **52**(3), 471–501(2010)
- [32] Rudin, L.I., Osher, S., Fatemi, E: Nonlinear total variation based noise removal algorithms. *Phys. D* **60**(1–4), 259–268(1992)
- [33] Steidl, G., Teuber, T.: Removing multiplicative noise by Douglas-Rachford splitting methods. *J. Math. Imaging Vis.* **36**(2), 168–184(2010)
- [34] Tao, M., Yuan, X.M.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM J. Optim.* **21**(1), 57–81(2011)
- [35] Tibshirani, R.J.: Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58**, 267–288(1996)
- [36] Yang, J.F., Yuan, X.M.: Linearized augmented lagrangian and alternating direction methods for nuclear norm minimization. *Math. Comp.* **82**(281), 301–329(2013)

- [37] Yuan, X.M.: Alternating direction method for covariance selection models. *J. Sci. Comput.* **51**(2), 261–273(2012)
- [38] Zhang, W.X., Cai, X.J., Jia, Z.H.: A proximal alternating linearization method for minimizing the sum of two convex functions. *Sci. China Math.* **58**(10), 1–20(2015)