# CONVERGENCE ANALYSIS ON A DATA-DERIVEN INEXACT PROXIMAL-INDEFINITE STOCHASTIC ADMM *

JIANCHAO BAI†, YANG CHEN‡, XINGJU CAI§, AND XUE YU¶

**Abstract.** In this paper, we propose an Inexact Proximal-indefinite Stochastic ADMM (abbreviated as IPS-ADMM) to solve a class of separable convex optimization problems whose objective functions consist of two parts: one is an average of many smooth convex functions and the other is a convex but potentially nonsmooth function. The involved smooth subproblem is tackled by an inexact accelerated stochastic gradient method based on an adaptive expansion step to avoid the scenario that the sample size can be extremely huge so that computing the objective function value or its gradient is much expensive. The involved nonsmooth subproblem is solved inexactly under a relative error criterion to avoid the case that the proximal operator is potentially unavailable. In contrast to most of deterministic and stochastic ADMM, our dual variable updates twice and allows a more flexible and larger stepsize region. By a variational analysis, we characterize the generated iterates as a variational inequality and finally establish the sublinear convergence rate of this IPS-ADMM in terms of the objective function gap and constraint violation. Experiments on solving the 3D CT reconstruction problem in medical imaging and the graph-guided fused lasso problem in machine learning show that our IPS-ADMM is very promising.

**Key words.** Convex optimization; Stochastic ADMM; Proximal-indefinite term; Larger stepsize; Convergence complexity; Machine learning

**AMS subject classifications.** 94A08; 90C25; 65Y20; 65K10

## 1. Introduction.

One of the important tasks in machine learning is to design efficient and reliable methods for structural empirical risk minimization problem, that is, to minimize a finite-sum of loss functions and an empirical regularizer subject to linear constraints:

$$\min_{x,y}\Big\{f(x)+g(y)\mid Ax+By=b,\ x\in\mathbb{R}^m,\ y\in\mathbb{R}^n\Big\},\tag{1.1}$$

where $f$ is an average of $N$ real-valued convex functions, that is, $f(x)=\frac{1}{N}\sum_{i=1}^{N}f_i(x)$; $g:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is a convex but possibly nonsmooth function; $A\in\mathbb{R}^{l\times m}, B\in\mathbb{R}^{l\times n}, b\in\mathbb{R}^l$ are given. Hereafter, the symbols $\mathbb{R},\mathbb{R}^m$, and $\mathbb{R}^{l\times m}$ denote the sets of real numbers, $m$ dimensional real column vectors, and $l\times m$ real matrices, respectively. Problems in the form of (1.1) are prevalent in 3D CT image reconstruction, graph-guided fused lasso and 6G multi-access edge computing networks, see [1, 2, 3, 5, 23].

Let the Lagrangian function of (1.1) be

$$\mathcal{L}(x,y,\lambda)=f(x)+g(y)-\lambda^\top(Ax+By-b)$$

and the augmented Lagrangian function with penalty $\beta>0$ be

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

* This research was partially supported by the National Natural Science Foundation of China (12471298, 12171479), the Shaanxi Fundamental Science Research Project for Mathematics and Physics (23JSQ031), and the National Social Science Fund of China (22BGL118).

† School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, P.R. China (jianchaobai@nwpu.edu.cn).

‡ School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, P.R. China (cy1202208@163.com).

§Corresponding author. School of Mathematical Sciences, Nanjing Normal University, Nanjing 210023, P.R. China (caixingju@njnu.edu.cn).

¶Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, P.R. China (xueyu_2019@ruc.edu.cn).

A popular Lagrangian-based method for solving (1.1) is the Alternating Direction Method of Multipliers (ADMM, [13, 15]). Starting with given $(y^k, \lambda^k)$, ADMM generates $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ by the following scheme

$$\begin{cases} x^{k+1} = \arg\min\limits_{x} \mathcal{L}_\beta(x, y^k, \lambda^k), \\ y^{k+1} = \arg\min\limits_{y} \mathcal{L}_\beta(x^{k+1}, y, \lambda^k), \\ \lambda^{k+1} = \lambda^k - s\beta(Ax^{k+1} + By^{k+1} - b), \end{cases}$$

where $s \in (0, \frac{1+\sqrt{5}}{2})$ denotes the stepsize of the dual variable $\lambda$. ADMM has been extensively investigated due to its simplicity and wide applications in machine learning, statistic learning, image processing, signal processing and so forth, and it has certain relationships with some first-order algorithms. As explained in [21], the augmented Lagrangian method, ADMM, proximal point method, proximal-like contraction method, primal-dual hybrid gradient algorithm, Douglas-Rachford/Peaceman-Rachford splitting method can be viewed as variational-inequality-based optimization algorithms. Especially, if the Peaceman-Rachford splitting method [27] is applied to the dual form of (1.1) and the dual variable updates twice with suitable stepsizes, then we can obtain the following symmetric ADMM

$$\begin{cases} x^{k+1} = \arg\min\limits_{x} \mathcal{L}_\beta(x, y^k, \lambda^k), \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \tau\beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\limits_{y} \mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}), \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(Ax^{k+1} + By^{k+1} - b). \end{cases}$$

Its global convergence and sublinear rate had been successfully established by He, et al. [20]. Moreover, the above symmetric ADMM was extended to a multi-block version [4] that enjoys the following relatively flexible stepsize region

$$\Delta_0 = \Big\{(\tau, s) \mid \ \tau + s > 0, \ \tau \le 1, \ \ -\tau^2 - s^2 - \tau s + \tau + s + 1 > 0\Big\}.$$

By adding a quadratic proximal term $\frac{1}{2}\|y - y^k\|_D^2$ with $D = r\mathbf{I} - \beta B^\top B$ to the involved $y$-subproblem of the standard ADMM, it is easy to obtain the following linearized ADMM:

$$\begin{cases} x^{k+1} = \arg\min\limits_{x} \mathcal{L}_\beta(x, y^k, \lambda^k), \\ y^{k+1} = \arg\min\limits_{y} \Big\{g(y) + \frac{r}{2}\big\|y - y^k - \frac{1}{r}B^\top\big[\lambda^k - \beta(Ax^{k+1} + By^k - b)\big]\big\|^2\Big\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{cases}$$

Generally speaking, the proximal parameter $r$ needs to satisfy $r > \beta\big\|B^\top B\big\|$. However, in this case $r$ might be very large while a smaller value is preferred from the viewpoint of numerical computing. In 2020, He, et al. [22] developed an optimal linearized ADMM as follows

$$\begin{cases} x^{k+1} = \arg\min\limits_{x} \mathcal{L}_\beta(x, y^k, \lambda^k), \\ y^{k+1} = \arg\min\limits_{y} \Big\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\big\|y - y^k\big\|_{D_0}^2\Big\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), \end{cases}$$

where

$$D_0 = \tau r\mathbf{I} - \beta B^\top B \quad \text{ with } \quad r > \beta\big\|B^\top B\big\| \text{ and } \tau \in (0.75, 1).$$

The above parametric conditions imply that the matrix $D$ might be positive indefinite. For more details on using the indefinite proximal matrix, we refer to [3, 9, 16, 18, 25].

In practice, it is difficult to solve the core subproblems exactly when applying ADMM-type methods to some application problems in image restoration [17] and medical imaging [3]. Consequently, inexact techniques are introduced to obtain an approximate solution to improve the efficiency of ADMM and to simplify the solving difficulty of subproblems. In 1992, Eckstein and Bertsekas [10] proposed an inexact ADMM which tackles the subproblems approximately and gradually increases its precision without disrupting the convergence properties. Later, He, et al. [19] extended the inexact method [10] to solve a class of variational inequality problems. Recently, some useful relative error criteria were proposed in [8, 11, 24, 25, 28]. Inspired by these inexact techniques, we will propose an Inexact Proximal-indefinite Stochastic ADMM (IPS-ADMM) framework which enjoys an indefinite proximal term and an adaptive expansion step with an Armijo-type linesearch approach to improve the algorithm performance as well as reduce the sensitivity of the parameter choice. Our proposed algorithm and its main features are organized in the forthcoming section.

**Notations**. We follow the same notations as introduced in [6]. The bold $\mathbf{I}$ denotes the identity matrix and $\mathbf{0}$ denotes the zero matrix/vector. For any symmetric matrices $A$ and $B$ having the same dimension, $A \succ B (A \succeq B)$ represents $A - B$ is a positive definite(semidefinite) matrix. For any symmetric matrix $G$, we simply denote $\|x\|_G^2 := x^\top G x$ and specially $\|x\|_G = \sqrt{x^\top G x}$ means a weighted norm if $G \succeq \mathbf{0}$, where the superscript $^\top$ represents the transpose operator. The subdifferential of a convex function $f$ is denoted as $\partial f(\cdot)$ and it reduces to $\nabla f(\cdot)$ if $f$ is differentiable. The mathematical expectation of a random variable is denoted as $\mathbb{E}[\cdot]$. For convenience of analysis, we denote $F(w) = f(x) + g(y)$ and define

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad w^k = \begin{pmatrix} x^k \\ y^k \\ \lambda^k \end{pmatrix}, \quad \mathcal{J}(w) = \begin{pmatrix} -A^\top \lambda \\ -B^\top \lambda \\ Ax + By - b \end{pmatrix}. \tag{1.3}$$

**2. Development of IPS-ADMM.** In this section, we describe the development of our IPS-ADMM based on the following preliminary assumption.

Assumption 2.1. *For any symmetric positive definite matrix $\mathcal{H}$, there exists a constant $v > 0$ such that the gradients $\nabla f_i$ satisfy the Lipschitz condition*

$$\left\|\nabla f_i(x_1) - \nabla f_i(x_2)\right\|_{\mathcal{H}^{-1}} \leq v \left\|x_1 - x_2\right\|_{\mathcal{H}}$$

*for every $x_1, x_2 \in \mathbb{R}^m$ and $i = 1, 2, \cdots, N$.*

The above condition will reduce to the regular Lipschitz continuity when $\mathcal{H}$ is an identity matrix. By the well-known Taylor expansion, Assumption 2.1 suggests that the function $f$ is $v$-bounded in the sense of

$$f(x_1) \leq f(x_2) + \langle \nabla f(x_2), x_1 - x_2 \rangle + \frac{v}{2}\|x_1 - x_2\|_{\mathcal{H}}^2. \tag{2.1}$$

Now, we present our inexact stochastic ADMM as shown in Algorithm 2.1 which enjoys an adaptive expansion step with the Armijo-type linesearch for the $x$-iterate and shares the same routine `xsub` proposed in [5] to obtain an inexact solution. The routine `xsub` is actually deduced by solving the proximal problem inexactly:

$$\min_{x \in \mathbb{R}^m} \mathcal{L}_\beta(x, y^k, \lambda^k) + \frac{1}{2}\left\|x - x^k\right\|_{\mathcal{D}_k}^2, \quad \mathcal{D}_k = \mathcal{M}_k - \beta A^\top A, \tag{2.2}$$

ALG. 2.1. Inexact Proximal-indefinite Stochastic ADMM (IPS-ADMM)

**Parameters:** $\beta > 0$, $\mathcal{H} \succ \mathbf{0}$, $\varpi > 1$, $\gamma \in (0.5, 1]$, $\sigma_3 \in [0, 1)$, $\sigma_1, \sigma_2 \geq 0$ satisfying $\sigma_1 + \sigma_2 \in [0, 1)$, $(\tau, s) \in \Delta$ is given by (2.8).
**Initialization:** $(x^0, y^0, \lambda^0, v^0) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$, $\breve{x}^0 = x^0$, $d^0 = \mathbf{0}$.
**For** $k = 0, 1, \cdots$
1. Choose $m_k > 0$, $\vartheta_k > 0$, and $\mathcal{M}_k \succeq \beta A^\top A$.
2. $h^k = -A^\top \big[\lambda^k - \beta(Ax^k + By^k - b)\big]$.
3. $(\overline{x}^k, \breve{x}^{k+1}) = \mathbf{xsub}(x^k, \breve{x}^k, h^k, m_k, \vartheta_k, \mathcal{M}_k, \mathcal{H})$.
4. Update $x^{k+1}$ by one of the following strategies:
   - S1: $x^{k+1} = \overline{x}^k$; [when $N$ is large]
   - S2: $x^{k+1} = x^k + \alpha_k \overline{d}^k$, where $\overline{d}^k = \overline{x}^k - x^k$, [when $N$ is small]
     and $\alpha_k = \varpi^j$ with $j > 0$ being the largest integer such that
     $\mathcal{L}_\beta(\mathbf{x}^{k+1}, \mathbf{y}^k, \lambda^k) \leq \mathcal{L}_\beta(\overline{\mathbf{x}}^k, \mathbf{y}^k, \lambda^k) - \frac{1}{\alpha_k - 1}\big\|x^{k+1} - \overline{x}^k\big\|^2_{\mathcal{M}_k - \frac{1+\alpha_k}{2}\beta A^\top A}$
     and $\frac{1}{\alpha_k}\mathcal{M}_k - \beta A^\top A \succeq \mathbf{0}$;
5. $\lambda^{k+\frac{1}{2}} = \lambda^k - \tau\beta(Ax^{k+1} + By^k - b)$.
6. Update $y^{k+1}$ by (2.4) such that the condition in (2.6) holds with $d^{k+1}$ satisfying (2.7).
7. Update the auxiliary variable $v^{k+1} = v^k - d^{k+1}$.
8. $\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(Ax^{k+1} + By^{k+1} - b)$.

**end**

---

$(x^+, \breve{x}^+) = \mathbf{xsub}(x_1, \breve{x}_1, h, m_k, \vartheta_k, \mathcal{M}_k, \mathcal{H})$.
**For** $t = 1, 2, \ldots, m_k$
1. Randomly select $\xi_t \in \{1, 2, \ldots, N\}$ with uniform probability.
2. $\beta_t = 2/(t+1)$, $\gamma_t = 2/(t\vartheta_k)$, $\widehat{x}_t = \beta_t \breve{x}_t + (1 - \beta_t)x_t$.
3. $d_t = \widehat{g}_t + e_t$, where $\widehat{g}_t = \nabla f_{\xi_t}(\widehat{x}_t)$ and $e_t$ is a random vector satisfying $\mathbb{E}\big[e_t\big] = \mathbf{0}$.
4. $\breve{x}_{t+1} = \arg\min\limits_{x \in \mathbb{R}^m} \Big\{\big\langle d_t + h, x\big\rangle + \frac{\gamma_t}{2}\big\|x - \breve{x}_t\big\|^2_{\mathcal{H}} + \frac{1}{2}\big\|x - x^k\big\|^2_{\mathcal{M}_k}\Big\}$.
5. $x_{t+1} = \beta_t \breve{x}_{t+1} + (1 - \beta_t)x_t$.

**end**
**Return** $(x^+, \breve{x}^+) = (x_{m_k+1}, \breve{x}_{m_k+1})$.

---

that is,

$$\min_{x \in \mathbb{R}^m} \frac{1}{N}\sum_{i=1}^N f_i(x) + \frac{\beta}{2}\Big(\big\|Ax + By^k - b - \frac{\lambda^k}{\beta}\big\|^2 - \big\|A(x - x^k)\big\|^2\Big) + \frac{1}{2}\big\|x - x^k\big\|^2_{\mathcal{M}_k}. \quad (2.3)$$

More precisely, we linearize the bracketed terms $\frac{\beta}{2}(\cdot)$ in (2.3) as $\big\langle -A^\top[\lambda^k - \beta(Ax^k + By^k - b)], x\big\rangle$, apply (2.1) to the first summable term in (2.3) by replacing the full gradient by a stochastic gradient, and exploit the popular momentum acceleration technique to obtain the solver `xsub`. For an inexact solution of the $y$-subproblem, we update it by employing an indefinite proximal term:

$$y^{k+1} \approx \arg\min_{y \in \mathbb{R}^n} \Big\{g(y) + \frac{\beta}{2}\big\|Ax^{k+1} + By - b - \lambda^{k+\frac{1}{2}}/\beta\big\|^2 + \frac{1}{2}\big\|y - y^k\big\|^2_{L_0}\Big\}, \quad (2.4)$$

where

$$L_0 = L - (1-\gamma)\beta B^\top B \quad \text{with} \quad \gamma \in (0.5, 1] \tag{2.5}$$

for any arbitrarily positive semidefinite matrix $L$. When $L$ is properly chosen and the proximal operator of $g(y)$ is available, the difficulty of solving the subproblem (2.4) could be greatly alleviated. However, for the more general case, we will compute $y^{k+1}$ such that the following relative error criterion

$$\begin{aligned} &2\left|\left\langle v^k - y^{k+1} + \frac{1-\tau}{1+\tau} d^k, d^{k+1}\right\rangle\right| + \frac{3-\tau}{1+\tau}\left\|d^{k+1}\right\|^2 \\ \le &\sigma_1\left\|y^{k+1} - y^k\right\|_L^2 + \sigma_2\left\|y^k - y^{k-1}\right\|_L^2 + (2-\tau-s)\sigma_3\beta\left\|Ax^{k+1} + By^{k+1} - b\right\|^2 \end{aligned} \tag{2.6}$$

holds, where $v^{k+1} = v^k - d^{k+1}$ and

$$d^{k+1} \in \partial g(y^{k+1}) - B^\top \lambda^{k+\frac{1}{2}} + \beta B^\top (Ax^{k+1} + By^{k+1} - b) + L_0(y^{k+1} - y^k). \tag{2.7}$$

In (2.6), $(\tau, s)$ are the stepsize parameters of the dual variable belonging to

$$\Delta = \left\{ (\tau, s) \left| \begin{array}{l} 0 < \tau + s < 2, \quad \tau > -1, \quad (1-\tau)^2(1-s)^2 \\ -(1+\tau)(2-\tau-s)(1-\sigma_3)[(5-3\tau)\gamma + \tau^2 + \tau - 4] \le 0 \end{array} \right. \right\}. \tag{2.8}$$

The step 4 of the main algorithm implies $\mathcal{M}_k - \frac{1+\alpha_k}{2}\beta A^\top A \succ \mathbf{0}$ since $\alpha_k > 1$ for any $k > 0$. The relative error criteria in (2.6) implies that the iterate $d^{k+1}$ involved in the optimality condition of $y$-subproblem can be controlled by the $y$-residual and/or the equality residual. Main features and contributions of the proposed IPS-ADMM are summarized in the following three aspects:

(i) **Flexibility of the dual stepsize**. Unlike the classical ADMM, the proposed IPS-ADMM updates the dual variable twice with a relatively flexible stepsize region as shown in (2.8). Especially, IPS-ADMM reduces to the existing SAS-ADMM [6] with an extra expansion linesearch step, see Remark 3.1 for more details. By special choices for the parameters $(\sigma_3, \gamma)$, two instances of the region $\Delta$ are depicted in yellow area of Figure 2.1:
   - If $(\sigma_3, \gamma) = (0, 1)$, the domain of $(\tau, s)$ is depicted in Figure 2.1 (a), where the top boundary curve is actually obtained by $\tau^2 + s^2 + \tau s - \tau - s - 1 = 0$.
   - If $(\sigma_3, \gamma) = (0.5, 0.8)$, the domain of $(\tau, s)$ is depicted in Figure 2.1 (b), where the top boundary curve is actually obtained by $(1-\tau)^2(1-s)^2 - \frac{1}{2}(1+\tau)(2-\tau-s)(\tau^2 - \frac{7}{5}\tau) = 0$.

   The first case shows that $\Delta$ contains the classical region $(0, \frac{1+\sqrt{5}}{2})$. Moreover, in this case, the stepsize $s$ could be $5/3$ which is larger than $\frac{1+\sqrt{5}}{2}$. These instances suggest that our dual variable allows more flexible stepsizes than some in the literature.

(ii) **Generality of the algorithm**. First and foremost, our IPS-ADMM has low memory requirement, since the $x$-subproblem is solved inexactly by an accelerated stochastic gradient method and there is no need to save previous stochastic gradients and iterates. The relative error criterion (2.6) is provided to solve the $y$-subproblem inexactly and ensure the convergence of IPS-ADMM. When the parameters $(\sigma_1, \sigma_2, \sigma_3)$ are zero, i.e., $d^{k+1} = \mathbf{0}$, the $y$-subproblem is solved exactly. However, unlike most of ADMM-type methods [4, 7, 26] that employ a positive definite/semidefinite matrix, our proximal matrix $L_0$ is possibly positive-indefinite according to (2.5) and the lower

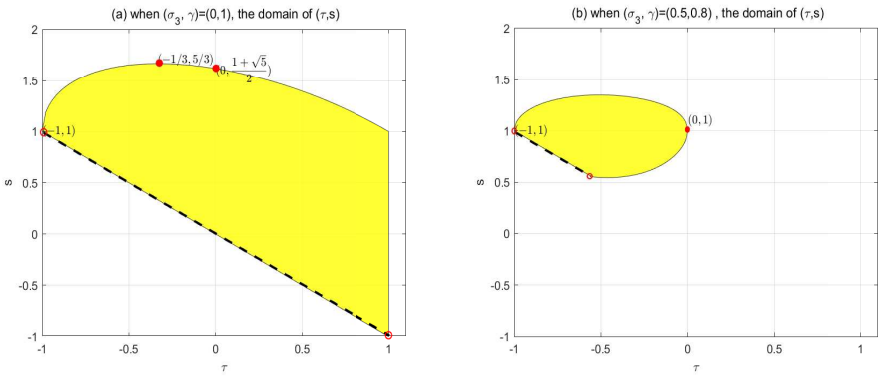


Fig. 2.1: Two instances of the stepsize region $\Delta$

bound of the parameter involved in $L_0$ is 0.5. Similar indefinite proximal matrix to $L_0$ has been used in [25]. When $(\sigma_1, \sigma_2) = (0, 0)$ and $(\tau, s) = (1, 0)$, our (2.6) immediately reduces to the existing criterion in [28]. Note that both the existing algorithms in [25, 28] are deterministic methods, while the proposed method in this paper is a stochastic method. These cases implies the generality and flexibility of IPS-ADMM.

(iii) **Convergence and performance guaranteed**. By a unified variational characterization for both the saddle-point of the problem and the generated iterates, we show that IPS-ADMM has the worst-case sublinear ergodic convergence rate in terms of the expectation of both the objective value gap and the constraint violation, and similar results can be found in [5, 26]. A key step to prove the convergence of IPS-ADMM is to estimate the lower bound of $\left\|w^k - \widetilde{w}^k\right\|_{G_k}^2$, where $G_k$ is a given block symmetric but positive indefinite matrix. Although $G_k$ is positive indefinite, convergence of IPS-ADMM is still established by making full use of the optimality condition of $y$-subproblem and several variants of the Cauchy-Schwarz inequality. Numerical results on testing a realistic 3D CT reconstruction problem in medical imaging and the graph-guided fused lasso problem with public dataset demonstrate the feasibility and effectiveness of the proposed IPS-ADMM, showing numerical improvements over several existing state-of-the-art methods.

**3. Convergence analysis.** In this section, we first provide a variational characterization for the saddle-point of (1.1) and a similar characterization for the iterates generated by IPS-ADMM. Then, we analyze convergence of IPS-ADMM based on some technical results from the first-order optimality condition of each subproblem.

**3.1. Variational characterization.** Let $\Omega := \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^l$. Since any saddle-point of (1.2) is the primal-dual solution of the convex optimization problem (1.1), we thus focus on the saddle-point of (1.1) in the following discussions. We call $w^* = (x^*; y^*; \lambda^*) \in \Omega$ the saddle-point of (1.1) if

$$\mathcal{L}(x^*, y^*, \lambda) \leq \mathcal{L}(x^*, y^*, \lambda^*) \leq \mathcal{L}(x, y, \lambda^*)$$

holds for any $w \in \Omega$, i.e.,

$$\left\{\begin{array}{r} f(x)-f(x^*)+\left\langle x-x^*, -A^\top\lambda^*\right\rangle \geq 0, \\ g(y)-g(y^*)+\left\langle y-y^*, -B^\top\lambda^*\right\rangle \geq 0, \\ Ax^*+By^*-b=\mathbf{0}. \end{array}\right.$$

Write these inequalities in a compact form to obtain

$$F(w)-F(w^*)+\left\langle w-w^*, \mathcal{J}(w^*)\right\rangle \geq 0, \tag{3.1}$$

namely,

$$F(w)-F(w^*)+\left\langle w-w^*, \mathcal{J}(w)\right\rangle \geq 0 \tag{3.2}$$

because the affine mapping $\mathcal{J}(w)$ defined by (1.3) satisfies

$$\left\langle w-\overline{w}, \mathcal{J}(w)-\mathcal{J}(\overline{w})\right\rangle = 0, \quad \forall w, \overline{w} \in \Omega. \tag{3.3}$$

Motivated by these discussions, a natural conjecture is that the iterative sequence generated by IPS-ADMM shall converge to the primal-dual solution $w^*$ if it can be characterized as a similar variational inequality to (3.2). To verify this conjecture, we first provide a property about the iterates generated by the routine `xsub` with the aid of the following notations

$$\widetilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1}+By^k-b) \ \text{ and } \ \delta_t = \nabla f(\widehat{x}_t) - d_t.$$

Lemma 3.1. *Suppose $\vartheta_k \in (0, 1/v)$ and Assumption 2.1 holds. Then, the iterates generated by IPS-ADMM satisfy*

$$f(x)-f(x^{k+1})+\left\langle x-x^{k+1}, -A^\top\widetilde{\lambda}^k\right\rangle \geq \left\langle x^{k+1}-x, \widetilde{\mathcal{D}}_k(x^{k+1}-x^k)\right\rangle + \zeta^k, \tag{3.4}$$

*where $\widetilde{\mathcal{D}}_k = \frac{1}{\alpha_k}\mathcal{M}_k - \beta A^\top A$ and*

$$\zeta^k = \frac{2}{m_k(m_k+1)}\left\{\begin{array}{l} \frac{1}{\vartheta_k}\left(\left\|x-\breve{x}^{k+1}\right\|_{\mathcal{H}}^2 - \left\|x-\breve{x}^k\right\|_{\mathcal{H}}^2\right) \\ -\sum_{t=1}^{m_k} t\left\langle \delta_t, \breve{x}_t - x\right\rangle - \frac{\vartheta_k}{4(1-\vartheta_k v)}\sum_{t=1}^{m_k} t^2\left\|\delta_t\right\|_{\mathcal{H}^{-1}}^2 \end{array}\right\}. \tag{3.5}$$

*Proof.* First of all, we have by [5, Lemma 3.2] that

$$f(x)-f(\overline{x}^k)+\left\langle x-\overline{x}^k, -A^\top\left[\lambda^k-\beta(A\overline{x}^k+By^k-b)\right]+\mathcal{D}_k(\overline{x}^k-x^k)\right\rangle \geq \zeta^k, \tag{3.6}$$

where $\mathcal{D}_k$ and $\zeta^k$ are given by (2.2) and (3.5) respectively. Then, the inequality in (3.4) will be proved in two cases:

- If the strategy S1 is adopted, then (3.6) reduces to (3.4) by using $x^{k+1} = \overline{x}^k$ by forcing $\alpha_k = 1$.
- If the strategy S2 is adopted, then it follows from the inequality in S2 [and $\alpha_k > 1$ if $j > 0$] that

$$\begin{aligned} & f(\overline{x}^k)-f(x^{k+1})+\left\langle \overline{x}^k-x^{k+1}, -A^\top\left[\lambda^k-\beta(Ax^{k+1}+By^k-b)\right]\right\rangle \\ \geq & \frac{1}{\alpha_k-1}\left\|x^{k+1}-\overline{x}^k\right\|_{\mathcal{M}_k-\frac{1+\alpha_k}{2}\beta A^\top A}^2 - \frac{\beta}{2}\left\|A(\overline{x}^k-x^{k+1})\right\|^2. \end{aligned} \tag{3.7}$$

Sum up the inequalities (3.6) and (3.7) together with the notations $\mathcal{D}_k$ in (2.2) and $\widetilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b)$ to obtain

$$\begin{aligned} & f(x) - f(x^{k+1}) + \langle x - x^{k+1}, -A^\top \widetilde{\lambda}^k \rangle \\ & + \langle x - \overline{x}^k, \beta A^\top A(\overline{x}^k - x^{k+1}) + \mathcal{D}_k(\overline{x}^k - x^k) \rangle + \frac{\beta}{2}\|A(\overline{x}^k - x^{k+1})\|^2 \\ \geq & \frac{1}{\alpha_k - 1}\|x^{k+1} - \overline{x}^k\|^2_{\mathcal{M}_k - \frac{1+\alpha_k}{2}\beta A^\top A} + \zeta^k. \end{aligned} \tag{3.8}$$

Notice that the expansion step in S2 indicates

$$\overline{x}^k - x^k = \frac{1}{\alpha_k}(x^{k+1} - x^k) \quad \text{and} \quad x^{k+1} - x^k = \frac{\alpha_k}{\alpha_k - 1}(x^{k+1} - \overline{x}^k),$$

which, by $\mathcal{D}_k = \mathcal{M}_k - \beta A^\top A$, shows

$$\begin{aligned} & \langle x - \overline{x}^k, \beta A^\top A(\overline{x}^k - x^{k+1}) + \mathcal{D}_k(\overline{x}^k - x^k) \rangle + \frac{\beta}{2}\|A(\overline{x}^k - x^{k+1})\|^2 \\ = & \langle x - x^{k+1} + x^{k+1} - \overline{x}^k, \beta A^\top A(x^k - x^{k+1}) \rangle \\ & + \langle x - \overline{x}^k, \mathcal{M}_k(\overline{x}^k - x^k) \rangle + \frac{\beta}{2}\|A(\overline{x}^k - x^{k+1})\|^2 \\ = & \langle x - x^{k+1}, \beta A^\top A(x^k - x^{k+1} \rangle + \langle x - \overline{x}^k, \mathcal{M}_k(\overline{x}^k - x^k) \rangle \\ & + \Big\langle \overline{x}^k - x^{k+1}, \frac{\beta}{2} A^\top A\big(\overline{x}^k - x^k + x^{k+1} - x^k\big) \Big\rangle \\ = & \langle x - x^{k+1}, \beta A^\top A(x^k - x^{k+1} \rangle + \Big\langle x - x^{k+1} + x^{k+1} - \overline{x}^k, \frac{1}{\alpha_k}\mathcal{M}_k(x^{k+1} - x^k) \Big\rangle \\ & + \Big\langle \overline{x}^k - x^{k+1}, \frac{\beta}{2} A^\top A\Big(1 + \frac{1}{\alpha_k}\Big)(x^{k+1} - x^k) \Big\rangle \\ = & \Big\langle x - x^{k+1}, \Big(\frac{1}{\alpha_k}\mathcal{M}_k - \beta A^\top A\Big)(x^{k+1} - x^k) \Big\rangle + \frac{1}{\alpha_k - 1}\|x^{k+1} - \overline{x}^k\|^2_{\mathcal{M}_k - \frac{1+\alpha_k}{2}\beta A^\top A}. \end{aligned}$$

Inserting this relation into (3.8) with simple algebra confirms (3.4). This completes the proof. □

In the above Lemma 3.1, $\vartheta_k \in (0, 1/v)$ is a necessary and sufficient condition to ensure that the last term of $\zeta^k$ is positive. Next, we provide a variational characterization for the iterates generated by IPS-ADMM.

LEMMA 3.2. *Suppose $\vartheta_k \in (0, 1/v)$. Then, the iterates generated by IPS-ADMM satisfy*

$$F(w) - F(\widetilde{w}^k) + \langle w - \widetilde{w}^k, \mathcal{J}(\widetilde{w}^k) \rangle - \langle y - \widetilde{y}^k, d^{k+1} \rangle \geq \langle w - \widetilde{w}^k, Q_k(w^k - \widetilde{w}^k) \rangle + \zeta^k \tag{3.9}$$

*for all $w \in \Omega$, where*

$$\widetilde{w}^k = \begin{pmatrix} \widetilde{x}^k \\ \widetilde{y}^k \\ \widetilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \widetilde{\lambda}^k \end{pmatrix} \quad \textit{and} \quad Q_k = \begin{bmatrix} \widetilde{\mathcal{D}}_k & & \\ & L_0 + \beta B^\top B & -\tau B^\top \\ & -B & \frac{1}{\beta}\mathbf{I} \end{bmatrix}. \tag{3.10}$$

*Proof.* By the updates of $\lambda^{k+\frac{1}{2}}$ and $\widetilde{\lambda}^k$, we have

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \tau(\lambda^k - \widetilde{\lambda}^k) = \widetilde{\lambda}^k + (\tau - 1)(\widetilde{\lambda}^k - \lambda^k). \tag{3.11}$$

Then, combine the convexity of $g$ and (2.7) to obtain

$$g(y)-g(\widetilde{y}^k)+\left\langle y-\widetilde{y}^k,\left\{\begin{array}{l}-B^\top\big[\widetilde{\lambda}^k+(\tau-1)(\widetilde{\lambda}^k-\lambda^k)\big]-d^{k+1}\\+\beta B^\top(A\widetilde{x}^k+B\widetilde{y}^k-b)+L_0(\widetilde{y}^k-y^k)\end{array}\right\}\right\rangle\geq 0 \quad (3.12)$$

for any $y\in\mathbb{R}^n$. Next, we focus on the $\{\cdot\}$ term in (3.12). Since $\beta(A\widetilde{x}^k+By^k-b)=-(\widetilde{\lambda}^k-\lambda^k)$,

$$\begin{aligned}&-B^\top\big[\widetilde{\lambda}^k+(\tau-1)(\widetilde{\lambda}^k-\lambda^k)\big]+\beta B^\top(A\widetilde{x}^k+B\widetilde{y}^k-b)+L_0(\widetilde{y}^k-y^k)-d^{k+1}\\=&-B^\top\big[\widetilde{\lambda}^k+(\tau-1)(\widetilde{\lambda}^k-\lambda^k)\big]+\beta B^\top B(\widetilde{y}^k-y^k)-d^{k+1}\\&+\beta B^\top(A\widetilde{x}^k+By^k-b)+L_0(\widetilde{y}^k-y^k)\\=&-B^\top\widetilde{\lambda}^k-\tau B^\top(\widetilde{\lambda}^k-\lambda^k)+(\beta B^\top B+L_0)(\widetilde{y}^k-y^k)-d^{k+1}.\end{aligned}$$

Substituting the above relation into (3.12) gives

$$g(y)-g(\widetilde{y}^k)+\left\langle y-\widetilde{y}^k,\left\{\begin{array}{l}-B^\top\widetilde{\lambda}^k-\tau B^\top(\widetilde{\lambda}^k-\lambda^k)-d^{k+1}\\+(L_0+\beta B^\top B)(\widetilde{y}^k-y^k)\end{array}\right\}\right\rangle\geq 0. \quad (3.13)$$

Besides, the way of generating $\widetilde{\lambda}^k$ implies

$$\left\langle\lambda-\widetilde{\lambda}^k,(A\widetilde{x}^k+B\widetilde{y}^k-b)-B(\widetilde{y}^k-y^k)+\frac{1}{\beta}(\widetilde{\lambda}^k-\lambda^k)\right\rangle\geq 0,\quad \forall\lambda\in\mathbb{R}^l. \quad (3.14)$$

So, the result (3.9) is confirmed by the last (3.4), (3.13) and (3.14). □

Similar to the techniques in [20, 26], we need to reformulate the term $Q_k(w^k-\widetilde{w}^k)$ and ensure the positive definiteness of a new matrix related to $Q$ for the convergence of IPS-ADMM. In fact, by the updates of $\lambda^{k+1},\widetilde{\lambda}^k$ and (3.11), we have

$$\begin{aligned}\lambda^{k+1}=&\lambda^{k+\frac{1}{2}}-s\beta(Ax^{k+1}+By^k-b)+s\beta B(y^k-y^{k+1})\\=&\lambda^k-\tau(\lambda^k-\widetilde{\lambda}^k)-s(\lambda^k-\widetilde{\lambda}^k)+s\beta B(y^k-\widetilde{y}^k),\end{aligned}$$

that is,

$$\lambda^k-\lambda^{k+1}=-s\beta B(y^k-\widetilde{y}^k)+(\tau+s)(\lambda^k-\widetilde{\lambda}^k).$$

Combine it with the notations $\widetilde{x}^k=x^{k+1}$ and $\widetilde{y}^k=y^{k+1}$ in (3.10) to get

$$w^k-w^{k+1}=P(w^k-\widetilde{w}^k),\qquad \text{where } P=\begin{bmatrix}\mathbf{I}&&\\&\mathbf{I}&\\&-s\beta B&(\tau+s)\mathbf{I}\end{bmatrix} \quad (3.15)$$

is clearly nonsingular for any $\tau+s>0$. By the first equality in (3.15), we obtain

$$Q_k(w^k-\widetilde{w}^k)=H(w^k-w^{k+1}) \quad (3.16)$$

and

$$H_k=Q_kP^{-1}=\begin{bmatrix}\widetilde{\mathcal{D}}_k&&\\&L_0+(1-\frac{\tau s}{\tau+s})\beta B^\top B&-\frac{\tau}{\tau+s}B^\top\\&-\frac{\tau}{\tau+s}B&\frac{1}{\beta(\tau+s)}\mathbf{I}\end{bmatrix}. \quad (3.17)$$

In the following, we provide a sufficient condition to ensure the positive definiteness of $H_k$. In particular, the condition with $\gamma = 1$ is the same as that in [6, Lemma 4.3]. Besides, combine (2.5) and this condition to have $L_0 \succeq (\tau - 1)\beta B^\top B$, which indicates that $L_0$ could be positive indefinite.

LEMMA 3.3. *Let $L \succeq (\tau - \gamma)\beta B^\top B$. Then, the matrix $H_k$ given by (3.17) is symmetric positive semidefinite for any $(\tau, s) \in \Delta$.*

*Proof.* The 4th step in IPS-ADMM ensures the positive definiteness of the matrix $\widetilde{\mathcal{D}}_k$ involved in Lemma 3.1. Recalling the structure of $H_k$, we only needs to demonstrate the positive semi-definiteness of its lower-upper 2-by-2 block, i.e.,

$$\begin{bmatrix} L_0 + (1 - \frac{\tau s}{\tau+s})\beta B^\top B & -\frac{\tau}{\tau+s}B^\top \\ -\frac{\tau}{\tau+s}B & \frac{1}{\beta(\tau+s)}\mathbf{I} \end{bmatrix} = \begin{bmatrix} L + (\gamma - \frac{\tau s}{\tau+s})\beta B^\top B & -\frac{\tau}{\tau+s}B^\top \\ -\frac{\tau}{\tau+s}B & \frac{1}{\beta(\tau+s)}\mathbf{I} \end{bmatrix} := H_k^L.$$

For any $L \succeq (\tau - \gamma)\beta B^\top B$ and $\beta > 0$, we have

$$H_k^L \succeq \begin{bmatrix} (\tau - \frac{\tau s}{\tau+s})\beta B^\top B & -\frac{\tau}{\tau+s}B^\top \\ -\frac{\tau}{\tau+s}B & \frac{1}{\beta(\tau+s)}\mathbf{I} \end{bmatrix} = \frac{1}{\tau+s}\begin{pmatrix} \sqrt{\beta}\tau B^\top \\ -\frac{1}{\sqrt{\beta}}\mathbf{I} \end{pmatrix}^\top \begin{pmatrix} \sqrt{\beta}\tau B & -\frac{1}{\sqrt{\beta}}\mathbf{I} \end{pmatrix}.$$

Clearly, the matrix in the right-hand-side of the last equality is positive semidefinite for any $\tau + s > 0$. Consequently, the matrix $H_k^L$ is positive semidefinite. □

COROLLARY 3.4. *Suppose $\vartheta_k \in (0, 1/v)$ and the conditions in Lemma 3.3 hold. Then, we have*

$$\begin{aligned} & F(w) - F(\widetilde{w}^k) + \langle w - \widetilde{w}^k, \mathcal{J}(w)\rangle - \langle y - \widetilde{y}^k, d^{k+1}\rangle \\ \geq & \frac{1}{2}\Big(\|w - w^{k+1}\|_{H_k}^2 - \|w - w^k\|_{H_k}^2\Big) + \frac{1}{2}\|w^k - \widetilde{w}^k\|_{G_k}^2 + \zeta^k \end{aligned} \tag{3.18}$$

*for all $w \in \Omega$, where $H_k$ is given by (3.17) and*

$$G_k = \begin{bmatrix} \widetilde{\mathcal{D}}_k & \\ & \begin{matrix} L_0 + (1-s)\beta B^\top B & (s-1)B^\top \\ (s-1)B & \frac{2-\tau-s}{\beta}\mathbf{I} \end{matrix} \end{bmatrix}.$$

*Proof.* It follows from (3.3), (3.9) together with (3.16) that

$$\begin{aligned} & F(w) - F(\widetilde{w}^k) + \langle w - \widetilde{w}^k, \mathcal{J}(\widetilde{w}^k)\rangle - \langle y - \widetilde{y}^k, d^{k+1}\rangle \\ = & F(w) - F(\widetilde{w}^k) + \langle w - \widetilde{w}^k, \mathcal{J}(w)\rangle - \langle y - \widetilde{y}^k, d^{k+1}\rangle \\ \geq & \zeta^k + \langle w - \widetilde{w}^k, H_k(w^k - w^{k+1})\rangle \\ = & \zeta^k + \frac{1}{2}\Big\{\|w - w^{k+1}\|_{H_k}^2 - \|w - w^k\|_{H_k}^2\Big\} \\ & + \frac{1}{2}\Big\{\|w^k - \widetilde{w}^k\|_{H_k}^2 - \|w^{k+1} - \widetilde{w}^k\|_{H_k}^2\Big\}, \end{aligned} \tag{3.19}$$

where the equality uses the following identity

$$\langle u - v, H_k(p - q)\rangle = \frac{1}{2}\Big\{\|u - q\|_{H_k}^2 - \|u - p\|_{H_k}^2 + \|v - p\|_{H_k}^2 - \|v - q\|_{H_k}^2\Big\} \tag{3.20}$$

with specifications $u := w, v := \widetilde{w}^k, p := w^k, q := w^{k+1}$.

Now, we focus on the last $\{\cdot\}$ term in (3.19). It holds by (3.15) that

$$\begin{aligned}
&\|w^k-\widetilde{w}^k\|_{H_k}^2-\|w^{k+1}-\widetilde{w}^k\|_{H_k}^2\\
=&\|w^k-\widetilde{w}^k\|_{H_k}^2-\|w^{k+1}-w^k+w^k-\widetilde{w}^k\|_{H_k}^2\\
=&\|w^k-\widetilde{w}^k\|_{H_k}^2-\|w^k-\widetilde{w}^k-P(w^k-\widetilde{w}^k)\|_{H_k}^2\\
=&\|w^k-\widetilde{w}^k\|_{G_k}^2,
\end{aligned}$$

where

$$\begin{aligned}
G_k&=P^\top H_k+H_k^\top P-P^\top H_kP=Q_k^\top+Q_k-P^\top H_kP\\
&=\begin{bmatrix}\widetilde{\mathcal{D}}_k & & \\ & L_0+(1-s)\beta B^\top B & (s-1)B^\top\\ & (s-1)B & \frac{2-\tau-s}{\beta}\mathbf{I}\end{bmatrix}.
\end{aligned}$$

Finally, (3.18) follows from (3.19) and the last two relationships. □

**3.2. More technical results.** If the matrix $G_k$ is also positive semidefinite, then convergence of IPS-ADMM can be proved by Corollary 3.4. However, it is not always positive semidefinite for any $(\tau,s)\in\Delta$, which brings a new challenge in the convergence analysis of IPS-ADMM. To proceed, a feasible scheme is to estimate a lower bound of $\|w^k-\widetilde{w}^k\|_{G_k}^2$ with the aid of an evaluation on the crossing term $(Ax^k+By^k-b)^\top B(y^k-y^{k+1})$.

LEMMA 3.5. *For any $\varsigma>0$, the iterates generated by IPS-ADMM satisfy*

$$\begin{aligned}
&\langle Ax^k+By^k-b,B(y^k-y^{k+1})\rangle\\
\geq&-\frac{\varsigma}{2}\|B(y^k-y^{k+1})\|^2-\frac{1}{2\varsigma}\|Ax^k+By^k-b\|^2.
\end{aligned}\tag{3.21}$$

*Proof.* It follows directly from the Cauchy-Schwarz inequlity. □

LEMMA 3.6. *For $\gamma\leq 1$, under the conditions in Lemma 3.3 we have*

$$\begin{aligned}
&\|w^k-\widetilde{w}^k\|_{G_k}^2\\
\geq&\|x^k-x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2+\|y^k-y^{k+1}\|_L^2+\omega_0\|Ax^{k+1}+By^{k+1}-b\|^2\\
&+\omega_1\Big(\|Ax^{k+1}+By^{k+1}-b\|^2-\|Ax^k+By^k-b\|^2\Big)+\omega_2\Big[\|y^k-y^{k+1}\|_L^2\\
&-\|y^{k-1}-y^k\|_L^2+(1-\gamma)\beta\Big(\|B(y^k-y^{k+1})\|^2-\|B(y^{k-1}-y^k)\|^2\Big)\Big]\\
&+\omega_3\|B(y^k-y^{k+1})\|^2+\frac{2(1-\tau)}{1+\tau}\langle y^k-y^{k+1},d^{k+1}-d^k\rangle,
\end{aligned}\tag{3.22}$$

*where $\varsigma>0$ is a constant and*

$$\begin{cases}
\omega_0=\left(2-\tau-s-\dfrac{1-\tau}{\varsigma(1+\tau)}\right)\beta,\\
\omega_1=\dfrac{1-\tau}{\varsigma(1+\tau)}\beta,\quad \omega_2=\dfrac{1-\tau}{1+\tau},\\
\omega_3=\left(\gamma-\tau-\dfrac{1-\tau}{1+\tau}\big[2(2-2\gamma+\tau)+\varsigma(1-s)^2\big]\right)\beta.
\end{cases}\tag{3.23}$$

*Proof.* Combine the definition of $\widetilde{\lambda}^k$ and the notation $\widetilde{w}^k$ in (3.10) to obtain

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

which, by the structure of $G_k$, further shows

$$\begin{aligned}\|w^k - \widetilde{w}^k\|_{G_k}^2 =& \|x^k - x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2 + \|y^k - y^{k+1}\|_{L+(\gamma-s)\beta B^\top B}^2 \\ &+ 2(s-1)\langle \lambda^k - \widetilde{\lambda}^k, B(y^k - y^{k+1})\rangle + \frac{2-\tau-s}{\beta}\|\lambda^k - \widetilde{\lambda}^k\|^2 \\ =& \|x^k - x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2 + \|y^k - y^{k+1}\|_L^2 + (\gamma-\tau)\beta\|B(y^k - y^{k+1})\|^2 \\ &+ (2-\tau-s)\beta\|Ax^{k+1} + By^{k+1} - b\|^2 \\ &+ 2(1-\tau)\beta\langle Ax^{k+1} + By^{k+1} - b, B(y^k - y^{k+1})\rangle.\end{aligned} \tag{3.24}$$

Next, we analyze the crossing term in the second equality of (3.24). Setting $y = y^k$ in (3.12) and $y = y^{k+1}$ in (3.12) at the $(k-1)$-th iteration respectively, we have

$$g(y^k) - g(y^{k+1}) + \left\langle y^k - y^{k+1}, \left\{ \begin{array}{l} -B^\top\lambda^{k+\frac{1}{2}} + \beta B^\top(Ax^{k+1} + By^{k+1} - b) \\ +L_0(y^{k+1} - y^k) - d^{k+1} \end{array} \right\} \right\rangle \geq 0$$

and

$$g(y^{k+1}) - g(y^k) + \langle y^{k+1} - y^k, -B^\top\lambda^{k-\frac{1}{2}} + \beta B^\top(Ax^k + By^k - b) + L_0(y^k - y^{k-1}) - d^k\rangle \geq 0$$

respectively. Adding these two inequality and using the following relation

$$\lambda^{k-\frac{1}{2}} - \lambda^{k+\frac{1}{2}} = \tau\beta(Ax^{k+1} + By^{k+1} - b) + s\beta(Ax^k + By^k - b) + \tau\beta B(y^k - y^{k+1}),$$

we have (since $1 + \tau > 0$ and $\gamma \leq 1$)

$$\begin{aligned}&\langle Ax^{k+1} + By^{k+1} - b, B(y^k - y^{k+1})\rangle \\ \geq& \frac{1-s}{1+\tau}\langle Ax^k + By^k - b, B(y^k - y^{k+1})\rangle - \frac{\tau}{1+\tau}\|B(y^k - y^{k+1})\|^2 \\ &+ \frac{1}{\beta(1+\tau)}\Big\langle y^k - y^{k+1}, (L - (1-\gamma)\beta B^\top B)\big[(y^k - y^{k+1}) - (y^{k-1} - y^k)\big]\Big\rangle \\ &+ \frac{1}{\beta(1+\tau)}\langle y^k - y^{k+1}, d^{k+1} - d^k\rangle \\ =& \frac{1-s}{1+\tau}\langle Ax^k + By^k - b, B(y^k - y^{k+1})\rangle - \frac{\tau}{1+\tau}\|B(y^k - y^{k+1})\|^2 \\ &+ \frac{1}{\beta(1+\tau)}\Big[\|y^k - y^{k+1}\|_L^2 - \langle y^k - y^{k+1}, L(y^{k-1} - y^k)\rangle - (1-\gamma)\beta\|B(y^k - y^{k+1})\|^2 \\ &+ (1-\gamma)\beta\langle y^k - y^{k+1}, B^\top B(y^{k-1} - y^k)\rangle\Big] + \frac{1}{\beta(1+\tau)}\langle y^k - y^{k+1}, d^{k+1} - d^k\rangle \\ \geq& \frac{1-s}{1+\tau}\langle Ax^k + By^k - b, B(y^k - y^{k+1})\rangle - \frac{2(1-\gamma)+\tau}{1+\tau}\|B(y^k - y^{k+1})\|^2 \\ &+ \frac{1}{2\beta(1+\tau)}\Big(\|y^k - y^{k+1}\|_L^2 - \|y^{k-1} - y^k\|_L^2 + (1-\gamma)\beta\big(\|B(y^k - y^{k+1})\|^2 \\ &- \|B(y^{k-1} - y^k)\|^2\big)\Big) + \frac{1}{\beta(1+\tau)}\langle y^k - y^{k+1}, d^{k+1} - d^k\rangle,\end{aligned} \tag{3.25}$$

where the last inequality uses the Cauchy-Schwarz inequality for the terms $\langle y^k - y^{k+1}, L(y^{k-1}-y^k)\rangle$ and $\langle y^k - y^{k+1}, B^\top B(y^{k-1}-y^k)\rangle$. Combine (3.24) and (3.25) to get

$$\begin{aligned}
&\|w^k-\widetilde{w}^k\|_{G_k}^2\\
\geq&\|x^k-x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2+\|y^k-y^{k+1}\|_L^2+(2-\tau-s)\beta\|Ax^{k+1}+By^{k+1}-b\|^2\\
&+(\gamma-\tau)\beta\|B(y^k-y^{k+1})\|^2+\frac{2\beta(1-\tau)(1-s)}{1+\tau}\langle Ax^k+By^k-b,B(y^k-y^{k+1})\rangle\\
&-\frac{2(2-2\gamma+\tau)(1-\tau)}{1+\tau}\beta\|B(y^k-y^{k+1})\|^2+\frac{2(1-\tau)}{1+\tau}\langle y^k-y^{k+1},d^{k+1}-d^k\rangle\\
&+\frac{1-\tau}{1+\tau}\left[\|y^k-y^{k+1}\|_L^2-\|y^{k-1}-y^k\|_L^2+(1-\gamma)\beta\begin{pmatrix}\|B(y^k-y^{k+1})\|^2\\-\|B(y^{k-1}-y^k)\|^2\end{pmatrix}\right]\\
\geq&\|x^k-x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2+\|y^k-y^{k+1}\|_L^2+\frac{2(1-\tau)}{1+\tau}\langle y^k-y^{k+1},d^{k+1}-d^k\rangle\\
&+\left(2-\tau-s-\frac{1-\tau}{\varsigma(1+\tau)}\right)\beta\|Ax^{k+1}+By^{k+1}-b\|^2\\
&+\frac{1-\tau}{\varsigma(1+\tau)}\beta\left(\|Ax^{k+1}+By^{k+1}-b\|^2-\|Ax^k+By^k-b\|^2\right)\\
&+\frac{1-\tau}{1+\tau}\left[\|y^k-y^{k+1}\|_L^2-\|y^{k-1}-y^k\|_L^2+(1-\gamma)\beta\begin{pmatrix}\|B(y^k-y^{k+1})\|^2\\-\|B(y^{k-1}-y^k)\|^2\end{pmatrix}\right]\\
&+\left(\gamma-\tau-\frac{1-\tau}{1+\tau}\left[2(2-2\gamma+\tau)+\varsigma(1-s)^2\right]\right)\beta\|B(y^k-y^{k+1})\|^2,
\end{aligned}$$

where the second inequality follows from (3.21) for the term $\frac{2\beta(1-\tau)(1-s)}{1+\tau}\langle Ax^k+By^k-b,B(y^k-y^{k+1})\rangle$. As a result, (3.22) holds with $\omega_i(i=0,1,2,3)$ given in (3.23). □

LEMMA 3.7. *Suppose $\vartheta_k\in(0,1/v)$, $\widetilde{\mathcal{D}}_k\succeq\widetilde{\mathcal{D}}_{k+1}\succeq\mathbf{0}$ and the conditions in lemma 3.6 hold. Then, the iterates generated by IPS-ADMM satisfy*

$$2\left[F(w)-F(\widetilde{w}^k)+\langle w-\widetilde{w}^k,\mathcal{J}(w)\rangle\right]\geq\Phi_{k+1}(w)-\Phi_k(w)+\Psi_{k+1}+2\zeta^k, \tag{3.26}$$

*where*

$$\begin{aligned}
\Phi_k(w)=&\|w-w^k\|_{H_k}^2+\|y-v^k\|^2+\omega_1\|Ax^k+By^k-b\|^2\\
&+(\omega_2+\sigma_2)\|y^{k-1}-y^k\|_L^2+\omega_2(1-\gamma)\beta\|B(y^k-y^{k+1})\|^2+\frac{2(1-\tau)}{1+\tau}\|d^k\|^2
\end{aligned}\tag{3.27}$$

*and*

$$\begin{aligned}
\Psi_k=&\|x^k-x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2+\|y^k-y^{k+1}\|_{(1-\sigma_1-\sigma_2)L-\frac{1-\tau}{1+\tau}I}^2\\
&+\overline{\omega}_0\|Ax^{k+1}+By^{k+1}-b\|^2+\omega_3\|B(y^{k+1}-y^k)\|^2
\end{aligned}\tag{3.28}$$

*with $\overline{\omega}_0=\left[(2-\tau-s)(1-\sigma_3)-\frac{1-\tau}{\varsigma(1+\tau)}\right]\beta$ and $\omega_i(i=1,2,3)$ given by (3.23).*

*Proof.* First of all, we have

$$\begin{aligned}
&2\langle y^k-y^{k+1},d^{k+1}-d^k\rangle\\
=&\|y^k-y^{k+1}+d^{k+1}-d^k\|^2-\|d^{k+1}-d^k\|^2-\|y^k-y^{k+1}\|^2\\
\geq&(\|d^{k+1}\|^2-\|d^k\|^2)+2\langle d^k,d^{k+1}\rangle-2\|d^{k+1}\|^2-\|y^k-y^{k+1}\|^2
\end{aligned}\tag{3.29}$$

Rearrange (3.18) with $\widetilde{y}^k = y^{k+1}$ to obtain

$$\begin{aligned}
&2\big[F(w) - F(\widetilde{w}^k) + \big\langle w - \widetilde{w}^k, \mathcal{J}(w)\big\rangle\big] \\
&\quad - \big\|w^k - \widetilde{w}^k\big\|_{G_k}^2 - 2\big\langle v^k - y^{k+1}, d^{k+1}\big\rangle + \big\|d^{k+1}\big\|^2 - 2\zeta^k \\
&\geq \Big(\big\|w - w^{k+1}\big\|_{H_k}^2 - \big\|w - w^k\big\|_{H_k}^2\Big) - 2\big\langle v^k - y, d^{k+1}\big\rangle + \big\|d^{k+1}\big\|^2 \\
&= \Big(\big\|w - w^{k+1}\big\|_{H_k}^2 - \big\|w - w^k\big\|_{H_k}^2\Big) + 2\big\langle y - v^k, v^k - v^{k+1}\big\rangle + \big\|v^k - v^{k+1}\big\|^2 \\
&= \Big(\big\|w - w^{k+1}\big\|_{H_k}^2 - \big\|w - w^k\big\|_{H_k}^2\Big) + \big\|y - v^{k+1}\big\|^2 - \big\|y - v^k\big\|^2,
\end{aligned} \tag{3.30}$$

where the first equality follows from the 7th step of IPS-ADMM and the last equality uses the aforementioned identity in (3.20).

Based on the previous inequalities (2.6) and the assumption that $\widetilde{\mathcal{D}}_k \succeq \widetilde{\mathcal{D}}_{k+1} \succeq \mathbf{0}$ (showing that $H_k \succeq H_{k+1}$), it holds by (3.29)-(3.30) and (3.22) that

$$\begin{aligned}
&2\big[F(w) - F(\widetilde{w}^k) + \big\langle w - \widetilde{w}^k, \mathcal{J}(w)\big\rangle\big] \\
\geq&\Big(\big\|w - w^{k+1}\big\|_{H_{k+1}}^2 + \big\|y - v^{k+1}\big\|^2\Big) - \Big(\big\|w - w^k\big\|_{H_k}^2 + \big\|y - v^k\big\|^2\Big) \\
&+ \big\|w^k - \widetilde{w}^k\big\|_{G_k}^2 + 2\big\langle v^k - y^{k+1}, d^{k+1}\big\rangle - \big\|d^{k+1}\big\|^2 + 2\zeta^k \\
\geq&2\zeta^k + 2\Big\langle v^k - y^{k+1} + \frac{1-\tau}{1+\tau}d^k, d^{k+1}\Big\rangle - \frac{3-\tau}{1+\tau}\big\|d^{k+1}\big\|^2 - \frac{1-\tau}{1+\tau}\big\|y^k - y^{k+1}\big\|^2 \\
&+ \big\|x^k - x^{k+1}\big\|_{\widetilde{\mathcal{D}}_k}^2 + \big\|y^k - y^{k+1}\big\|_L^2 + \omega_0\big\|Ax^{k+1} + By^{k+1} - b\big\|^2 + \omega_3\big\|B(y^{k+1} - y^k)\big\|^2 \\
&+ \Phi_{k+1}(w) - \Phi_k(w) + \sigma_2\Big(\big\|y^k - y^{k-1}\big\|_L^2 - \big\|y^{k+1} - y^k\big\|_L^2\Big) \\
\geq&2\zeta^k + \Phi_{k+1}(w) - \Phi_k(w) + \sigma_2\Big(\big\|y^k - y^{k-1}\big\|_L^2 - \big\|y^{k+1} - y^k\big\|_L^2\Big) - \frac{1-\tau}{1+\tau}\big\|y^k - y^{k+1}\big\|^2 \\
&+ \big\|x^k - x^{k+1}\big\|_{\widetilde{\mathcal{D}}_k}^2 + \big\|y^k - y^{k+1}\big\|_L^2 + \omega_0\big\|Ax^{k+1} + By^{k+1} - b\big\|^2 + \omega_3\big\|B(y^{k+1} - y^k)\big\|^2 \\
&- \sigma_1\big\|y^{k+1} - y^k\big\|_L^2 - \sigma_2\big\|y^k - y^{k-1}\big\|_L^2 - (2-\tau-s)\sigma_3\beta\big\|Ax^{k+1} + By^{k+1} - b\big\|^2 \\
=&2\zeta^k + \Phi_{k+1}(w) - \Phi_k(w) + \Psi_{k+1}.
\end{aligned} \tag{3.31}$$

This completes the proof. □

Remark 3.1. *We explain for any $(\tau, s) \in \Delta$ and $\gamma \in (0.5, 1]$, all of parameters $\overline{\omega}_0$ and $\omega_i (i = 0, \cdots, 3)$ are nonnegative. First of all, for any $\tau > -1, \tau + s < 2$ and $\sigma_3 < 1$, by choosing $\varsigma \geq \frac{1-\tau}{(1+\tau)(2-\tau-s)(1-\sigma_3)}$, we know $\varsigma > 0$ and it satisfies (3.21).*

- *It follows from $\varsigma > 0$ that $(2-\tau-s)(1-\sigma_3) - \frac{1-\tau}{\varsigma(1+\tau)} \geq 0$, which together with $\beta > 0$ indicates $\overline{\omega}_0 \geq 0$. By the following relationship*

$$\omega_0 = \overline{\omega}_0 + \sigma_3(2-\tau-s)\beta,$$

*we have by $\overline{\omega}_0 \geq 0$ that $\omega_0 \geq 0$. Also, both $\omega_1$ and $\omega_2$ are nonnegative for any $\tau > -1$ and $\varsigma > 0$. For any*

$$\varsigma \leq \frac{(5-3\tau)\gamma + \tau^2 + \tau - 4}{(1-\tau)(1-s)^2},$$

*it holds that $\gamma - \tau - \frac{1-\tau}{1+\tau}\Big[2(2-2\gamma+\tau) + \varsigma(1-s)^2\Big] \geq 0$ and hence $\omega_3 \geq 0$.*

- *Combining the above lower and upper bound constraints for $\varsigma$, we further have*

$$\frac{1-\tau}{(1+\tau)(2-\tau-s)(1-\sigma_3)} \leq \frac{(5-3\tau)\gamma+\tau^2+\tau-4}{(1-\tau)(1-s)^2},$$

*which shows*

$$\gamma \geq \frac{(1-\tau)^2(1-s)^2}{(1+\tau)(2-\tau-s)(1-\sigma_3)(5-3\tau)} - \frac{\tau^2+\tau-4}{5-3\tau}. \tag{3.32}$$

*If $\gamma < 1$, then the proximal matrix $L_0$ is positive indefinite. If $\gamma = 1$, then, by the last inequality we have*

$$(1-s)^2 < (1-\sigma_3)(2-\tau-s)(1+\tau).$$

*Moreover, when $\sigma_3 = 0$ the above inequality reduces to the third inequality of $\Delta_0$ proposed by [6], and $\tau = 0$ shows $s \in (0, \frac{1+\sqrt{5}}{2})$ which is the region in the classical ADMM. By (3.32), we can see the lower bound of the parameter $\gamma$ is more flexible than the lower bound 0.75 provided in e.g. [16, 22]. For example, if we take $s = 1$ and $\tau \to -1$, then we have $\gamma > 0.5$, which meets the same lower bound in [25].*

**3.3. Iteration complexity.** In this subsection, we present the following key convergence result based on Lemma 3.7.

THEOREM 3.8. *Let $L \succeq \max\left\{(\tau-\gamma)\beta B^\top B, \frac{1-\tau}{(1-\sigma_1-\sigma_2)(1+\tau)}\mathbf{I}\right\}$ and $(\tau, s) \in \Delta$. If for some integers $\kappa, T > 0$, the following conditions hold for all $k \in [\kappa, \kappa+T]$:*

*(i) $\vartheta_k \in (0, \frac{1}{2v})$ and the sequence $\left\{\vartheta_k m_k(m_k+1)\right\}$ is nondecreasing;*

*(ii) $\widetilde{\mathcal{D}}_k \succeq \widetilde{\mathcal{D}}_{k+1} \succeq \mathbf{0}$, $\mathbb{E}(\|\delta_t\|^2_{\mathcal{H}^{-1}}) \leq \sigma^2$ for some $\sigma > 0$, where $\delta_t$ and $\widetilde{\mathcal{D}}_k$ are defined in Lemma 3.1.*

*Then, we have*

$$\mathbb{E}\big[F(\overline{w}_T)-F(w)+\big\langle \overline{w}_T-w, \mathcal{J}(w)\big\rangle\big] \leq \frac{1}{2(1+T)}\left\{\sigma^2\sum_{k=\kappa}^{\kappa+T}\vartheta_k m_k + \frac{4\big\|x-\breve{x}^\kappa\big\|^2_{\mathcal{H}}}{m_\kappa(m_\kappa+1)\vartheta_\kappa} + \Phi_\kappa(w)\right\} \tag{3.33}$$

*for any $w \in \Omega$, where $\overline{w}_T = \frac{1}{1+T}\sum_{k=\kappa}^{\kappa+T}\widetilde{w}^k$ and $\Phi_k(w)$ is defined by (3.27).*

*Proof.* By the assumptions together with $\Psi_k \geq 0$, it follows from Lemma 3.7 that

$$F(\widetilde{w}^k) - F(w) + \big\langle \widetilde{w}^k - w, \mathcal{J}(w)\big\rangle \leq -\zeta^k + \frac{1}{2}\big(\Phi_k(w) - \Phi_{k+1}(w)\big).$$

Summing the inequality over $k$ between $\kappa$ and $\kappa+T$, we have by Lemma 3.3 that

$$\sum_{k=\kappa}^{\kappa+T} F(\widetilde{w}^k) - (1+T)\big[F(w) - \big\langle \overline{w}_T - w, \mathcal{J}(w)\big\rangle\big] \leq -\sum_{k=\kappa}^{\kappa+T}\zeta^k + \frac{1}{2}\Phi_\kappa(w). \tag{3.34}$$

The convexity of the composite function $F$ shows $F(\overline{w}_T) \leq \frac{1}{1+T}\sum_{k=\kappa}^{\kappa+T} F(\widetilde{w}^k)$. Then, divide (3.34) by $1+T$ to obtain

$$F(\overline{w}_T) - F(w) + \big\langle \overline{w}_T - w, \mathcal{J}(w)\big\rangle \leq \frac{1}{2(1+T)}\Big(-2\sum_{k=\kappa}^{\kappa+T}\zeta^k + \Phi_\kappa\Big). \tag{3.35}$$

In what follows, we estimate the expectation about terms involving $\zeta^k$. Since $\{m_k(m_k+1)\vartheta_k\}$ is nondecreasing for any $k \in [\kappa, \kappa+T]$ and $\mathcal{H} \succ \mathbf{0}$, it holds that

$$\sum_{k=\kappa}^{\kappa+T} \frac{2}{m_k(m_k+1)\vartheta_k}\left(\|x-\breve{x}^k\|_{\mathcal{H}}^2 - \|x-\breve{x}^{k+1}\|_{\mathcal{H}}^2\right) \le \frac{2\,\|x-\breve{x}^\kappa\|_{\mathcal{H}}^2}{m_\kappa(m_\kappa+1)\vartheta_\kappa}.$$

Because $\delta_t = \nabla f(\widehat{x}_t) - d_t = \nabla f(\widehat{x}_t) - \nabla f_{\xi_t}(\widehat{x}_t) - e_t$ relies on the index $\xi_t$, we have $\mathbb{E}[\delta_t] = \mathbf{0}$ since the random variable $\xi_t \in \{1, 2, \ldots, N\}$ is selected with uniform probability and $\mathbb{E}[e_t] = \mathbf{0}$. Besides, we have

$$\mathbb{E}\big[\langle \delta_t, \breve{x}_t - x\rangle\big] = \mathbf{0} \tag{3.36}$$

because $\breve{x}_t$ relies on $\xi_{t-1}$, $\xi_{t-2}$, .... The assumption $\mathbb{E}(\|\delta_t\|_{\mathcal{H}^{-1}}^2) \le \sigma^2$ together with $m_k \ge 1$ implies

$$\mathbb{E}\Big[\sum_{t=1}^{m_k} t^2\,\|\delta_t\|_{\mathcal{H}^{-1}}^2\Big] \le \frac{\sigma^2 m_k(m_k+1)(2m_k+1)}{6} \le \frac{\sigma^2}{2} m_k^2(m_k+1). \tag{3.37}$$

Based on the above discussions and the condition $\vartheta_k \le 1/(2v)$, we have

$$-\mathbb{E}\Big[\sum_{k=\kappa}^{\kappa+T} \zeta^k\Big] \le \frac{2\|x-\breve{x}^\kappa\|_{\mathcal{H}}^2}{m_\kappa(m_\kappa+1)\vartheta_\kappa} + \frac{\sigma^2}{2}\sum_{k=\kappa}^{\kappa+T} \vartheta_k m_k.$$

Substituting the last inequality into (3.35) confirms the result (3.33). □

Remark 3.2. *Suppose the conditions in Theorem 3.8. Let*

$$\vartheta_k = \min\Big\{\frac{c_1}{m_k(m_k+1)}, c_2\Big\} \quad \text{and} \quad m_k = \max\Big\{\lceil c_3 k^\varrho\rceil, m\Big\}, \tag{3.38}$$

*where $c_1, c_2, c_3 > 0, \varrho \ge 1$ are constants and $m > 0$ is a given integer. Then, set $w = w^*$ in (3.26) together with the property in (3.1) and $\Psi_k \ge 0$ to obtain*

$$\begin{aligned}
0 \le& 2\big[F(\widetilde{w}^k) - F(w^*) + \langle \widetilde{w}^k - w^*, \mathcal{J}(w^*)\rangle\big] \\
\le& \Phi_k(w^*) - \Phi_{k+1}(w^*) - \frac{4}{m_k(m_k+1)\vartheta_k}\Big(\|x^*-\breve{x}^{k+1}\|_{\mathcal{H}}^2 - \|x^*-\breve{x}^k\|_{\mathcal{H}}^2\Big) \\
&+ \frac{1}{m_k(m_k+1)}\left\{4\sum_{t=1}^{m_k} t\langle \delta_t, \breve{x}_t - x^*\rangle + \frac{\vartheta_k}{(1-\vartheta_k v)}\sum_{t=1}^{m_k} t^2\|\delta_t\|_{\mathcal{H}^{-1}}^2\right\},
\end{aligned}$$

*which, by the choice in (3.38), gives*

$$\begin{aligned}
&\Phi_{k+1}(w^*) + \frac{4}{c_1}\|x^*-\breve{x}^{k+1}\|_{\mathcal{H}}^2 \\
\le& \Phi_k(w^*) + \frac{4}{c_1}\|x^*-\breve{x}^k\|_{\mathcal{H}}^2 \\
&+ \frac{1}{m_k(m_k+1)}\left\{4\sum_{t=1}^{m_k} t\langle \delta_t, \breve{x}_t - x^*\rangle + \frac{\vartheta_k}{(1-\vartheta_k v)}\sum_{t=1}^{m_k} t^2\|\delta_t\|_{\mathcal{H}^{-1}}^2\right\} \\
\le& \cdots\cdots \\
\le& \Phi_1(w^*) + \frac{4}{c_1}\|x^*-\breve{x}^1\|_{\mathcal{H}}^2 \\
&+ \sum_{i=1}^{k}\frac{1}{m_i(m_i+1)}\left\{4\sum_{t=1}^{m_i} t\langle \delta_t, \breve{x}_t - x^*\rangle + \frac{\vartheta_i}{(1-\vartheta_i v)}\sum_{t=1}^{m_i} t^2\|\delta_t\|_{\mathcal{H}^{-1}}^2\right\}.
\end{aligned}$$

*Taking expectation on both sides of the last inequality together with the previous results in (3.36)-(3.37), we have*

$$\mathbb{E}\Big[\Phi_{k+1}(w^*) + \frac{4}{c_1}\big\|x^* - \breve{x}^{k+1}\big\|_{\mathcal{H}}^2\Big] \le \Phi_1(w^*) + \frac{4}{c_1}\big\|x^* - \breve{x}^1\big\|_{\mathcal{H}}^2 + c_1\sigma^2\sum_{i=1}^{k}\frac{1}{1+c_3 i^{\varrho}}$$

*due to* $\vartheta_i m_i \le \frac{c_1}{1+c_3 i^{\varrho}}$ *by the choice in (3.38). And the last term in the above inequality is* $\mathcal{O}(1)$ *if* $\varrho > 1$*, while it is* $\mathcal{O}(\log k)$ *if* $\varrho = 1$*. Hence, the sequence* $\{w^k\}$ *is bounded in expectation.*

The following theorem shows that by a proper choice for the algorithm parameters, we can establish the convergence rate of IPS-ADMM in the expectation of both the objective function value gap and the constraint violation.

THEOREM 3.9. *Suppose the conditions in Theorem 3.8 and (3.38) hold. Then, we have*

$$\mathbb{E}\big[F(\overline{w}_T) - F(w^*)\big] = E_{\varrho}(T) = \mathbb{E}\big[\big\|Ax_T + By_T - b\big\|\big], \tag{3.39}$$

*where* $E_{\varrho}(T) = \mathcal{O}(1/T)$ *for* $\varrho > 1$ *and* $E_{\varrho}(T) = \mathcal{O}(T^{-1}\log T)$ *for* $\varrho = 1$*.*

*Proof.* The proof is omitted since it is same as that of [5, Theorem 4.2]. □

REMARK 3.3. *The objective error and the constraint violation converge to zero in expectation as the result in (3.39), but this does not imply the boundedness of the ergodic iterates. If there exists* $c > 0$ *such that the matrix* $\widetilde{\mathcal{D}}_k$ *in Lemma 3.1 satisfying* $\widetilde{\mathcal{D}}_k \ge c\mathbf{I}$*, then the iterates* $(x^k, y^k, \lambda^k)$ *are bounded in expectation. Furthermore, under a strong convexity assumption, the ergodic iterates converge expectation, see our previous analysis in [5, Appendix]. Note that the total iteration of ALG 2.1 is* $T(m_k + t)$*, where* $t$ *is the iteration of solving* $y$*-subproblem which can be tackled by some first-order methods such as the proximal gradient method in [29]. If one wish ALG 2.1 has* $\mathcal{O}(1/T)$ *convergence rate, both* $m_k$ *and* $t$ *should be large. This can be approximated since the proximal gradient method in [29] includes FISTA as a special case and the* **xsub** *is also an accelerated method.*

**4. Numerical experiments.** In this section, we evaluate the performance of the proposed IPS-ADMM for solving the 3D CT reconstruction problem in medical imaging and the graph-guided fused lasso problem in machine learning.

**4.1. 3D CT reconstruction problem.** Recall the so-called 3D CT reconstruction problem

$$\begin{aligned} &\min_{x,y} F(x,y) := \frac{1}{N}\sum_{j=1}^{N}(\mathcal{R}_j x - b_j)^2 + \mu\|y\|_1 \\ &\text{s.t.} \quad \nabla x = y, \end{aligned} \tag{4.1}$$

where $\mathcal{R}_j$ is the 3D $x$-ray transform of the beam indexed by [14] and $\nabla$ is the discrete gradient operator. The size of 3D image to be reconstructed is $256 \times 256 \times 64$, the detector plane is $512 \times 384$ and the number of views is 668. Clearly, the size of the observed data $b$, i.e., $N = 131334144$ and is a big data. So, the strategy S1 is used in the following experiments. Problem (4.1) fits (1.1) with $(A, B, b) = (\nabla, -\mathbf{I}, \mathbf{0})$ and applying IPS-ADMM to (4.1) results in the following key iterations

$$\begin{cases} \breve{x}_{t+1} = [\gamma_t\mathcal{H} + \mathcal{M}_k]^{-1}[\gamma_t\mathcal{H}\breve{x}_t + \mathcal{M}_k x^k - d_t - h^k], \\ y^{k+1} = \text{Shrink}\big(\frac{\mu}{\iota+\gamma\beta}, y^k - \frac{\lambda^{k+\frac{1}{2}} - \beta(\nabla x^{k+1} - y^k)}{\iota+\gamma\beta}\big), \end{cases}$$

where Shrink$(\cdot,\cdot)$ is the soft shrinkage operator and can be computed by the built-in MATLAB function "wthresh". Since the $y$-subproblem has a closed-form solution due to $B = \mathbf{I}$, we can compute $y^{k+1}$ exactly, i.e., $d^{k+1} = \mathbf{0}$. That is, the parameters $\sigma_i (i = 1, 2, 3)$ used to control the accuracy of the inexact solution are set to 0. The regularization parameter of the problem is set as $\mu = 10^{-1}$ and the penalty parameter $\beta = 10^{-8}$. For computational efficiency, we let the number of inner iteration $m_k = 10$. We use the tuned parameters $(\mathcal{H}, \mathcal{M}_k) = (10^{-5}\mathbf{I}, 10^{-2}\mathbf{I})$, $\iota = 0.5$ (i.e., $L = \iota\mathbf{I}$) and $(\tau, s, \gamma) = (0.9, 1.09, 0.51)$. By these choices, the parameters satisfy the previous region in (2.8).

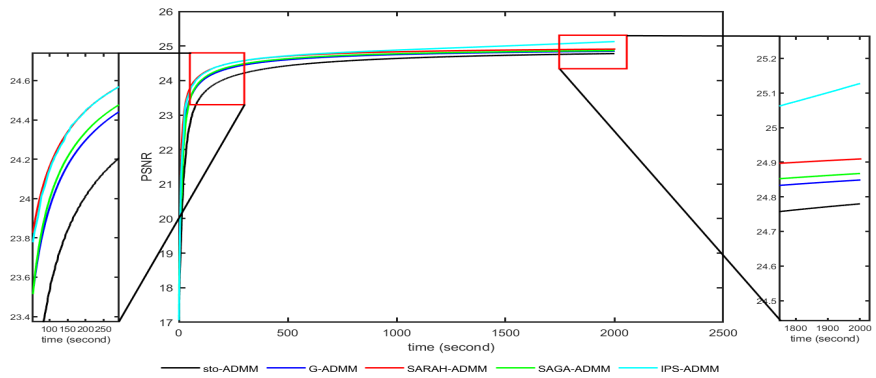


Fig. 4.1: PSNR results of five comparative algorithms

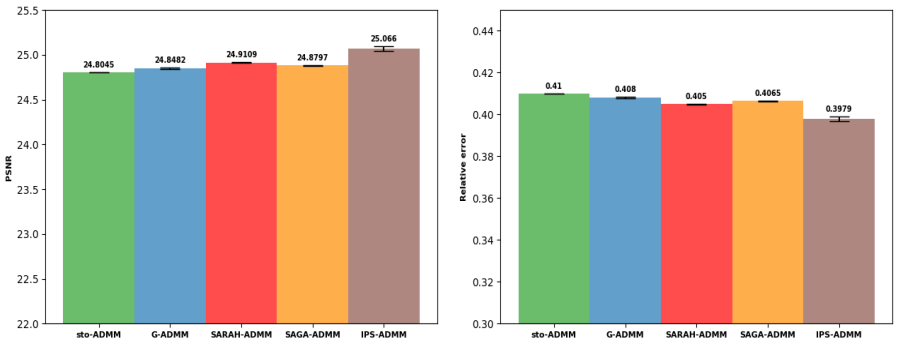


Fig. 4.2: The average PSNR and relative error of each algorithm

In order to quantitatively evaluate the performance of IPS-ADMM, we consider four state-of-the-art baselines: stochastic ADMM (sto-ADMM, [23]) with mini-batch stochastic gradients; Generalized ADMM (G-ADMM, [10]); SARAH-ADMM with the SARAH gradient estimator and SAGA-ADMM with the SAGA gradient estimator, see [2]. We terminate these algorithms when the maximal iterations $10^4$ or the allowable running time 2000 seconds is satisfied. The quality of the reconstructed image is

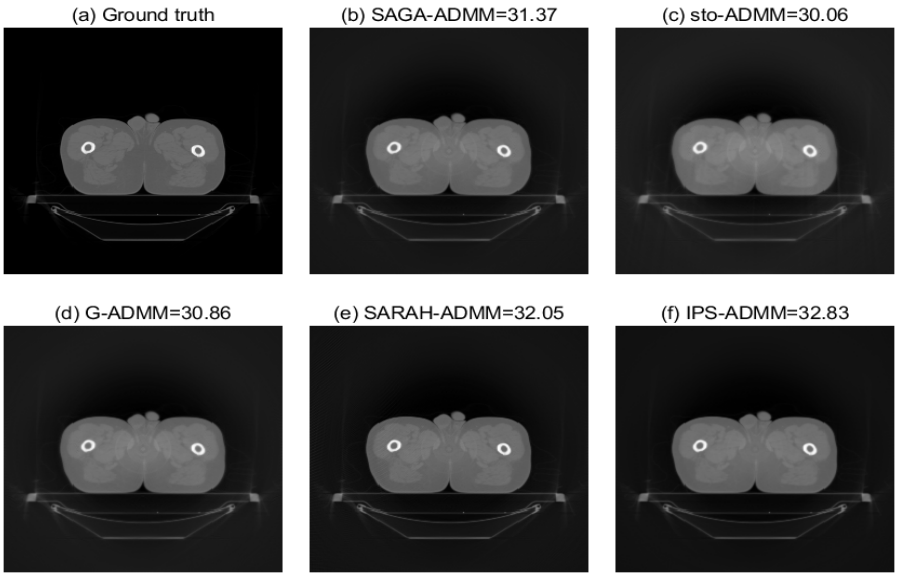


Fig. 4.3: Final reconstruction images of different methods for the **7th** slice

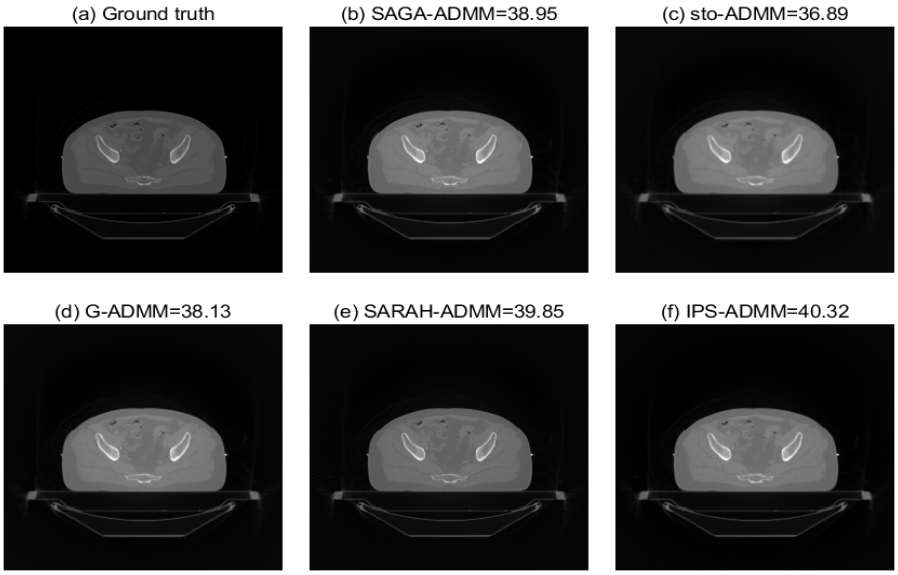


Fig. 4.4: Final reconstruction images of different methods for the **57th** slice

evaluated by the widely-used Peak Signal-to-Noise Ratio (PSNR):

$$\text{PSNR} = 10\log_{10}\left(\frac{d_x \times d_y \times d_z}{\text{MSE}}\right) \quad \text{with} \quad \text{MSE} = \|x - \widetilde{x}\|^2,$$

where $x$ and $\widetilde{x}$ are the original and reconstructed images, respectively. We also calcu-

late the Relative Error $= \|x - \widetilde{x}\|/\|x\|$. For each algorithm, we calculate the average and standard deviation of these two metrics over 5 independent runs.

We report some experimental results in Figures 4.1-4.4, and all experiments are run in MATLAB R2019a on a heterogeneous high-performance computational cluster equipped with Tesla V100 GPU with 24 cores and 192GB memory. Figure 4.1 shows how the PSNR of reconstructed images changes by different algorithms, and we can see the performance of all methods improves as the running time increases. Although the left zoom-in view in Figure 4.1 indicates that SARAH-ADMM can achieve a specific PSNR value in a relatively short time, the running time of our IPS-ADMM is similar to SARAH-ADMM and significantly faster than the other three algorithms. Moreover, the right zoom-in view in Figure 4.1 shows that IPS-ADMM achieves obviously the best restoration result and the PSNR is still improving, most likely due to the two updates of the dual variable and the use of a more flexible stepsize region. Figure 4.2 describes the obtained average PSNR and relative error of each algorithm, which indicates that IPS-ADMM consistently has the best quality of reconstructed images in terms of the obtained PSNR and relative error. Figures 4.3 and 4.4 visualize the 7th and 57th slices of the reconstructed 3D CT image, respectively. It can be seen that IPS-ADMM produces visually superior results than SAGA-ADMM, G-ADMM and sto-ADMM, and also slightly better results than SARAH-ADMM employing the biased SARAH gradient estimator. Compared to SARAH-ADMM, IPS-ADMM has a relative improvement of 2.43% and 1.17% on the 7th slice and 57th slice, respectively. As shown in Figure 4.3, many fuzzy circular contours can be observed in the 7th slice images reconstructed by the algorithms sto-ADMM, G-ADMM SAGA-ADMM and SARAH-ADMM. However, these circular contours are not obvious in images reconstructed by IPS-ADMM.

**4.2. Graph-guided fused lasso problem.** Consider the following graph-guided fused lasso problem in machine learning:

$$\begin{aligned} &\min_{x,y} F(x,y) := f(x) + \frac{\mu}{2}\|y\|^2 + \delta_{\mathcal{C}}(y) \\ &\text{s.t.} \quad Ax = y, \end{aligned} \tag{4.2}$$

where $f(x) := \frac{1}{N}\sum_{j=1}^{N} f_j(x)$ with $f_j(x) = \log\left(1 + \exp(-b_j a_j^\top x)\right)$ denoting the logistic loss function on the feature-label pair $(a_j, b_j) \in \mathbb{R}^m \times \{-1, 1\}$, $N(> m)$ is the data size, $\mu > 0$ is a given regularization parameter, and $\delta_{\mathcal{C}}$ is the indicator function of the set $\mathcal{C} = \{y \in \mathbb{R}^l : \mathbf{0} \le y \le u,\ \mathbf{1}^\top y = c\}$, i.e., $\delta_{\mathcal{C}}(y) = 0$ if $y \in \mathcal{C}$; $\delta_{\mathcal{C}}(y) = \infty$, otherwise. Here, $\mathbf{1}$ is the vector of ones, $c$ is a positive number, $A = [G; \mathbf{I}]$ is a matrix encoding the feature sparsity pattern, and $G$ is the sparsity pattern of the graph that is obtained by sparse inverse covariance estimation [12]. Applying our IPS-ADMM to the above problem, we have the following key iterations

$$\begin{cases} \breve{x}_{t+1} = [\gamma_t \mathcal{H} + \mathcal{M}_k]^{-1}[\gamma_t \mathcal{H}\breve{x}_t + \mathcal{M}_k x^k - d_t - h^k], \\ y^{k+1} \approx \arg\min_{y \in \mathbb{R}^l} \delta_{\mathcal{C}}(y) + \frac{\mu+\tau\beta}{2}\left\|y - y_c^k\right\|^2, \end{cases}$$

where $y_c^k = \frac{\tau\beta y^k - \lambda^{k+\frac{1}{2}} + \beta(Ax^{k+1} - y^k)}{\mu+\tau\beta}$. Note that the involved $y$-subproblem, which needs a projection on a simplex set, has no closed-form-solution. Hence, we solve it inexactly by the method developed in [29] using the previous criterion (2.6). Although the above $\breve{x}_{t+1}$ has a closed formula, the original $x$-subproblem in (2.3) is solved inexactly by our procedure **xsub**.

In the next experiments, the parameters $u$ and $c$ are chosen as 10*$\mathbf{1}$ and 5 respectively, the regularization parameter $\mu$ is fixed as $10^{-5}$, the penalty parameter $\beta$ in IPS-ADMM uses the tuned value 0.005, and the matrices $\mathcal{M}_k$ are updated adaptively by the strategy in Remark 4.2 [5] with $\rho_0 = 1$, $\rho_{\min} = 10^{-5}$ and $\mathcal{H} = 2 \times 10^{-5}\mathbf{I}$. The rest parameters and the vector $e_t$ are chosen in the same way as that used in [5, Section 7.1]. We use the following relative objective value error

$$\text{Obj\_err(k)} = \frac{|F(x^k, y^k) - F^*|}{\max\{F^*, 1\}}$$

and the optimality error

$$\text{Opt\_err(k)} = \max\left\{\|Ax^k - y^k\|, \|\nabla f(x^k) - A^\top \lambda^k\|, \|y^k - \mathcal{P}_{\mathcal{C}}(y^k - \mu y^k - \lambda^k)\|\right\}$$

to measure the performance of our IPS-ADMM. Here, $\mathcal{P}_{\mathcal{C}}(\cdot)$ denotes the projection operator onto the convex set $\mathcal{C}$, and $F^*$ is the approximate optimal objective function value obtained by running IPS-ADMM for more than 1000 minutes.

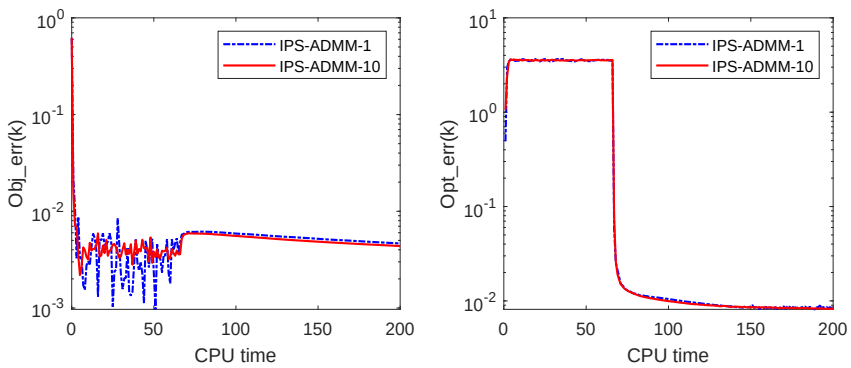


Fig. 4.5: Comparison of Obj_err(k) and Opt_err(k) against the CPU time

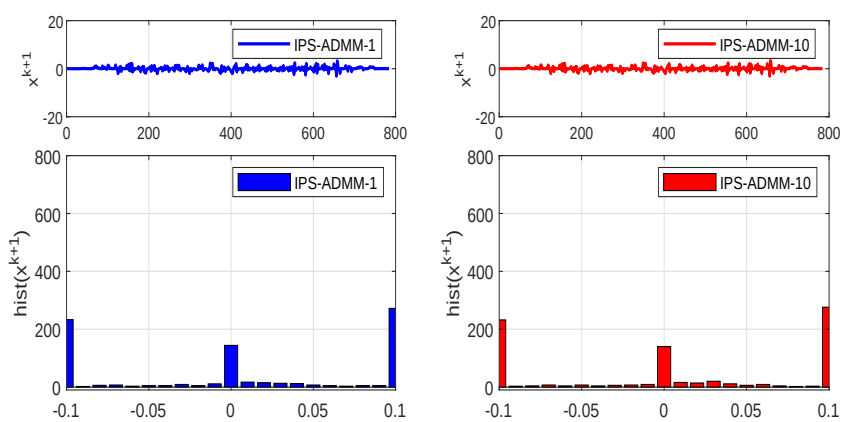


Fig. 4.6: Comparison of the final iterate $x^{k+1}$ and hist($x^{k+1}$) after 1 and 10 successive runs

By implementing IPS-ADMM in MATLAB R2018a (64-bit) and on a PC with

Windows 10 operating system, with an Intel i7-8700K CPU and 16GB RAM and starting with the initial point $(x^0, x^0, \lambda^0) = (\mathbf{0}, \mathbf{0}, \mathbf{0})$, Figure 4.5 plots the error Obj_err(k) and Opt_err(k) against the CPU time on the public dataset *mnist* (including 11,791 samples and 784 features, that is, $(N, m) = (11791, 784)$) downloaded from LIBSVM website, where we make 1 and 10 successive runs of IPS-ADMM (denoted by IPS-ADMM-1 and IPS-ADMM-10) under the CPU time budgets 200s, respectively and then report the average results. The finally obtained iterative solution $\mathbf{x}^{k+1}$ and hist($\mathbf{x}^{k+1}$) under different successive runs are shown in Figure 4.6. The reported results in both figures show that the proposed method is feasible and promising.

## REFERENCES


[1] A. Asheralieva, D. Niyato, and Y. Miyanaga, *Efficient dynamic distributed resource slicing in 6G multi-access edge computing networks with online ADMM and message passing graph neural networks*, IEEE Trans. Mobile Comput., 23 (2024) pp. 2614–2638.

[2] F. Bian, J. Liang, and X. Zhang, *A stochastic alternating direction method of multipliers for non-smooth and non-convex optimization*, Inverse Problems, 37 (2021), 075009.

[3] J. Bai, F. Bian, X. Chang, and L. Du, *Accelerated stochastic Peaceman-Rachford method for empirical risk minimization*, J. Oper. Res. Soc. China, 11 (2023), pp. 783–807.

[4] J. Bai, J. Li, F. Xu, and H. Zhang, *Generalized symmetric ADMM for separable convex optimization*, Comput. Optim. Appl., 70 (2018), pp. 129–170.

[5] J. Bai, W. Hager, and H. Zhang, *An inexact accelerated stochastic ADMM for separable convex optimization*, Comput. Optim. Appl., 81 (2022), pp. 479–518.

[6] J. Bai, D. Han, H. Sun, and H. Zhang, *Convergence on a symmetric accelerated stochastic ADMM with larger stepsizes*, CSIAM Trans. Appl. Math., 3 (2022), pp. 448–479.

[7] X. Cai, D. Han, $\mathcal{O}(1/t)$ *complexity analysis of the generalized alternating direction method of multipliers*, Sci. China Math., 62 (2019), pp. 795–808.

[8] C. Chen, D. Yu, D. Han, *Exact and inexact DouglasCRachford splitting methods for solving large-scale sparse absolute value equations*, IMA J. Numer. Anal., 43 (2023), pp. 1036–1060.

[9] J. Chen, Y. Wang, H. He, and Y. Lv, *Convergence analysis of positive-indefinite proximal ADMM with a Glowinskis relaxation factor*, Numer. Algor., 83 (2020), pp. 1415–1440.

[10] J. Eckstein, and D. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318.

[11] J. Eckstein, and W. Yao, *Relative-error approximate versions of Douglas-Rachford splitting and special cases of the ADMM*, Math. Program., 170 (2018), pp. 417–444.

[12] J. Friedman, T. Hastie and R. Tibshirani, *Sparse inverse covariance estimation with the graphical lasso*, Biostatistics, 9 (2008), pp. 432–441.

[13] D. Gabay, and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Comput. Math. Appl., 2 (1976), pp. 17–40.

[14] H. Gao, *Fast parallel algorithms for the x-ray transform and its adjoint*, Med. Phys., 39 (2012), pp. 7110–7120.

[15] R. Glowinski, and A. Marrocco, *Approximation paréléments finis d'rdre un et résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires*, Rev. Fr. Autom. Inform. Rech. Opér. Anal. Numér. 2 (1975), pp. 41–76.

[16] Y. Gu, B. Jiang, D Han, *An indefinite-proximal-based strictly contractive Peaceman-Rachford splitting method*, J. Comput. Math., 41 (2002), pp. 1017–1040.

[17] Y. Gu, B. Jiang, D Han, *An augmented Lagrangian based parallel splitting method for separable convex minimization with applications to image processing*, Math. Comput. 83 (289), pp. 2263–2291.

[18] D. Han, *A survey on some recent developments of alternating direction method of multipliers*, J. Oper. Res. Soc. China, 10 (2022) pp. 1–52.

[19] B. He, L. Liao, D. Han, and H. Yang, *A new inexact alternating directions method for monotone variational inequalities*, Math. Program., 92 (2002), pp. 103–118.

[20] B. He, F. Ma, and Y. Yuan, *Convergence study on the symmetric version of ADMM with larger step sizes*, SIAM J.Imaging Sci., 9 (2016), pp. 1467–1501.

[21] B. He, *Contraction Methods for Convex Optimization and Monotone Variational Inequalities*, Lecture Notes, (2020), http://maths.nju.edu.cn/∼hebma/New-VS/HeBS_JiangYI_NEW.pdf.

[22] B. He, F. Ma, and X. Yuan, *Optimally linearizing the alternating direction method of multi-*

*pliers for convex programming*, Comput. Optim. Appl., 75 (2020), pp. 361–388.

[23] F. Huang, and S. Chen, *Mini-batch stochastic ADMMs for nonconvex nonsmooth optimization*, arXiv: 1802.03284, (2019).

[24] F. Jiang, Z. Wu, X. Cai, and H. Zhang, *A first-order inexact primal-dual algorithm for a class of convex-concave saddle point problems*, Numer. Algor., 88 (2021), pp. 1109–1136.

[25] F. Jiang, and Z. Wu, *An inexact symmetric ADMM algorithm with indefinite proximal term for sparse signal recovery and image restoration problems*, J. Comput. Appl. Math., 417 (2023), 114628.

[26] F. Ma, *Convergence study on the proximal alternating direction method with larger step size*, Numer. Algor., 85 (2020), pp. 399–425.

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

[28] J. Xie, A. Liao, and X. Yang, *An inexact alternating direction method of multipliers with relative error criteria*, Optim. Lett., 11 (2017), pp. 583–596.

[29] B. Wen, X. Cheng, and T. Pong, *Linear convergence of proximal gradient algorithm with extrapolation for a class of nonconvex nonsmooth minimization problems*, SIAM J. Optim., 27 (2017), pp. 124–145.