AN INEXACT PROXIMAL-INDEFINITE STOCHASTIC ADMM WITH APPLICATIONS IN 3D CT RECONSTRUCTION *

JIANCHAO BAI†, YANG CHEN‡, AND XUE YU§

Abstract. In this paper, we develop an Inexact Proximal-indefinite Stochastic ADMM (abbreviated as IPS-ADMM) for solving a class of separable convex optimization problems whose objective functions consist of two parts: one is an average of many smooth convex functions and another is a convex but possibly nonsmooth function. The involved smooth subproblem is tackled by an inexact accelerated stochastic gradient method based on an adaptive expansion step to avoid the case that the sample size can be huge so that computing the objective function value or its gradient is much expensive. The restulting nonsmooth subproblem is solved inexactly under a relative error criterion to avoid the case that the proximal operator is potentially unavailable. Since the dual variable updates twice, it allows a more flexible and larger stepsize region compared with standard deterministic and stochastic ADMMs. 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. The efficacy of our IPS-ADMM is demonstrated by comparing with several state-of-the-art methods for solving the three-dimensional (3D) CT reconstruction problem in medical imaging.

Key words. Convex optimization; Stochastic ADMM; Proximal-indefinite term; Larger stepsize; Convergence complexity; 3D CT reconstruction

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

1. Introduction.

One of important tasts in large-scale 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 K real-valued convex functions, that is, $f(x) = \frac{1}{K}\sum_{i=1}^{K} f_i(x)$; $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ 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 data. 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) arise in 3D CT image reconstruction, graph-guided fused lasso and 6G multi-access edge computing networks, cf. [1, 2, 3, 5, 20, 25] to list a few.

Let the Lagrangian function of (1.1) be

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

* This research was partially supported by Guangdong Basic and Applied Basic Research Foundation (2023A1515012405), Shaanxi Fundamental Science Research Project for Mathematics and Physics (No. (22JSQ001) and National Natural Science Foundation of China (12171479).

† Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 518057, P.R. China; 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).

§ School of Statistics, Renmin University of China, Beijing 100872, P.R. China (xueyu_2019@ruc.edu.cn).

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

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

A popular Lagrangian-based method for solving the problem (1.1) is the Alternating Direction Method of Multipliers (ADMM, [13, 15]). Starting with given (y^k, λ^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 λ. 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 also has certain relationships with some first-order algorithms. As explained in [18], 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-based first-order algorithms. Especially, if the Peaceman-Rachford splitting method [22] is applied to the dual form of (1.1) and the dual variable is updated twice with suitable stepsizes, 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 complexity had been carefully established by He, et al. [17]. 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 y-subproblem of the standard ADMM, it is easy to obtain the following linearized ADMM iterations

$$\begin{cases} x^{k+1} = \arg\min\limits_x \mathcal{L}_\beta(x, y^k, \lambda^k), \\ y^{k+1} = \arg\min\limits_y \Big\{\theta_2(y) + \frac{r}{2}\big\|y - y^k - \frac{1}{r}B^\top\big[\lambda^k - \beta(Ax^{k+1} + By^k - b)\big]\big\|^2\Big\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{cases}$$

Generally speaking, the proximal parameter r needs to satisfy $r > \beta\big\|B^\top B\big\|$ to further guarantee convergence of the algorithm. However, in this case r might be very large while a smaller value of r is preferred from the viewpoint of numerical performance. In 2020, He, et al.[19] 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

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

The above parametric conditions imply that the matrix D might be positive indefinite since a smaller proximal parameter is allowed. For more details on using the indefinite proximal matrix, we refer to [3, 8, 24] to list a few. In practice, it is difficult to solve the core subproblems exactly when applying ADMM-type methods to solve real application problems in e.g. image restoration [7] 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. [16] extends the inexact method [10] to solve a class of variational inequality problems. Recently, some simple yet useful relative error criteria were proposed in [7, 11, 12]. Inspired by these inexact techniques, in this paper we would 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 positive definite(semidefinite) matrix. For any symmetric matrix G, we simply denote $x^\top Gx := \|x\|_G^2$ and specially $\sqrt{x^\top Gx} = \|x\|_G$ 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

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

2. Development of IPS-ADMM. In this section, we describe the development of our IPS-ADMM based on a preliminary assumption which reduces to the general Lipschitz continuity when $\mathcal{H}$ is an identity matrix.

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

$$\big\|\nabla f_i(x_1) - \nabla f_i(x_2)\big\|_{\mathcal{H}^{-1}} \le v\big\|x_1 - x_2\big\|_{\mathcal{H}}$$

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

By the well-known Taylor expansion the above assumption suggests that the function f is v-bounded in the sense of

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

Now, we present our inexact stochastic ADMM as shown in ALG. 2.1 which enjoys an adaptive expansion step with the Armijo-type linesearch for the x-iterate and shares

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

Parameters: $\beta > 0,\ \mathcal{H} \succ \mathbf{0},\ \varpi > 1,\ (\tau, s) \in \Delta$ given by (2.9).
Initialization: $(x^0, y^0, \lambda^0, v^0) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n,\ \breve{x}^0 = x^0$.
For $k = 0, 1, \cdots$
1. Choose $m_k > 0,\ \vartheta_k > 0$, and $\mathcal{M}_k$ such that $\frac{1}{\alpha_k}\mathcal{M}_k - \beta A^\top A \succeq \mathbf{0}$.
2. $h^k = -A^\top\left[\lambda^k - \beta(Ax^k + By^k - b)\right]$.
3. $(\overline{x}^k, \breve{x}^{k+1}) = \mathbf{xsub}(x^k, \breve{x}^k, h^k)$.
4. Update the expansion step: $x^{k+1} = x^k + \alpha_k \overline{d}^k$, where $\overline{d}^k = \overline{x}^k - x^k$, $\alpha_k = \varpi^j$ with $j \ge 0$ being the largest integer such that
$$\mathcal{L}_\beta(\mathbf{x}^{k+1}, \mathbf{y}^k, \lambda^k) \le \mathcal{L}_\beta(\overline{\mathbf{x}}^k, \mathbf{y}^k, \lambda^k) - \tfrac{1}{\alpha_k - 1}\left\|x^{k+1} - \overline{x}^k\right\|^2_{\mathcal{M}_k - \frac{1+\alpha_k}{2}\beta A^\top A},$$
if $j > 0$; otherwise, $\alpha_k = 1$ and $x^{k+1} = \overline{x}^k$(without expansion step).
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

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

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

Note: In step 4, $\mathcal{M}_k - \frac{1+\alpha_k}{2}\beta A^\top A \succ \mathbf{0}$ since $\alpha_k > 1$ for any $j > 0$.

the same routine `xsub` as in [5] to obtain an inexact solution. The routine `xsub` is actually deduced by solving the proximal problem inexactly:

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

that is,

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

More precisely, we linearize the bracketed terms $\frac{\beta}{2}(\cdot)$ in (2.3) as $\left\langle -A^\top[\lambda^k - \beta(Ax^k + By^k - b)], x\right\rangle$, 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 inexact solution of the y-subproblem, we update it by employing an indefinite proximal term:

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

where

$$L_0 = L - (1-\gamma)\beta B^\top B \quad \text{with} \quad \gamma \in (\eta, 1] \;\; \text{and} \;\; \eta \in (0.75, 1) \tag{2.5}$$

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

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

holds with $v^{k+1} = v^k - d^{k+1}$ 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 the last inequality, $\sigma_i (i = 1, 2, 3)$ are some negative constants to control the accuracy of an inexact solution of the y-subproblem and they satisfy

$$\sigma_1 + \sigma_2 \in [0, 1), \;\; \sigma_i \in [0, 1), \;\; i = 1, 2, 3, \tag{2.8}$$

(τ, s) are the stepsize parameters of the dual variable belonging to

$$\Delta = \left\{ (\tau, s) \middle| \begin{array}{l} \tau + s > 0, \quad \tau \le 3\eta - 2 - \sqrt{(1-\eta)(13-9\eta)}, \\ (2-\gamma-\eta)(1-s)^2 - (1-\eta)(1-\sigma_3)(1+\tau)(2-\tau-s) \le 0 \end{array} \right\}. \tag{2.9}$$

The relative error criteria in (2.6) implies that the 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. Several main features and contributions of the proposed IPS-ADMM are summarized as 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 in (2.9). Especially, IPS-ADMM reduces to the existing SAS-ADMM [6] with an extra expansion linesearch step if we take $\gamma = 1$, $\sigma_3 = 0$ and $\eta \to 1$. By special choices for the parameters (σ_3, γ, η), four instances of the region Δ are depicted in yellow area of Figure 2.1 and they seem to be enlarged or shrunken of the existing region Δ_0:
   - If $(\sigma_3, \gamma) = (0, 1)$ and $\eta \to 1$, the domain of (τ, s) is depicted as Figure 2.1 (a), where the top boundary curve is actually obtained by the inequality in the sequel (3.30), that is, $-\tau^2 - s^2 - \tau s + \tau + s + 1 = 0$.
   - If $(\sigma_3, \gamma) = (0, 1)$ and $\eta = 0.99$, the domain of (τ, s) is depicted in the shadow area of Figure 2.1 (b).
   - If $(\sigma_3, \gamma) = (0, 1)$ and $\eta = 0.9$, the domain of (τ, s) is depicted in the shadow area of Figure 2.1 (c).
   - If $(\sigma_3, \gamma) = (0, 0.95)$ and $\eta = 0.9$, the domain of (τ, s) is depicted in the shadow area of Figure 2.1 (d).

The first three cases shows that Δ contains the classical region $(0, \frac{1+\sqrt{5}}{2})$. Moreover, in the first two cases, the stepsize s could be $5/3$ which is larger than $\frac{1+\sqrt{5}}{2}$. The subsequent Remark 3.2 presents another stepsize region Δ_1 under a different relative error criteria and two instances of Δ_1 are depicted in Figure 3.1. These instances suggest that our dual variable allows more flexible stepsizes than some in the literature.

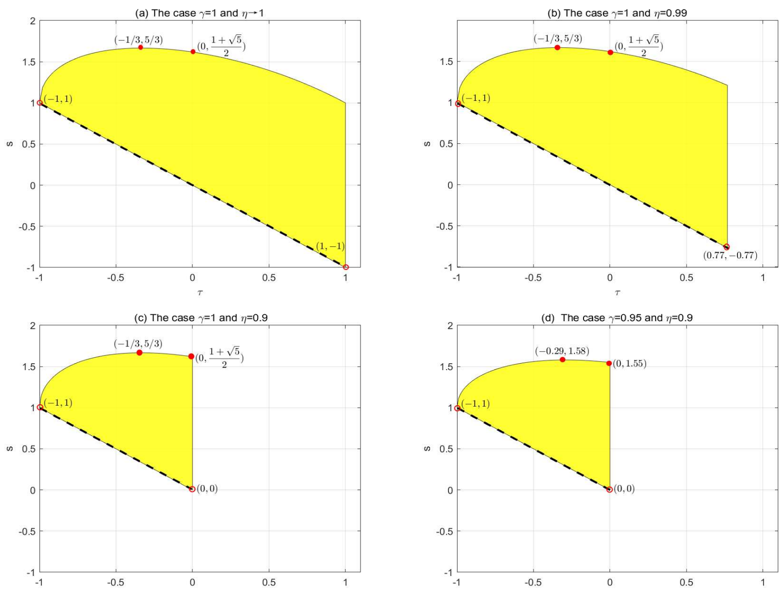

Fig. 2.1: Four instances of the stepsize region Δ

(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 criteria (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, different from most of ADMM-type methods [4, 9, 21, 23] that employ a positive definite/semidefinite matrix, our proximal matrix L_0 is possibly positive-indefinite according to (2.5) and the lower bound of the parameter involved in L_0 is still 0.75 (the same value to [19]). Except the relative error criteria (2.6), Remark 3.2 also provides another relative error criteria while still maintaining a similar stepsize region to Δ. These cases implies the generality and flexibility of IPS-ADMM.

(iii) **Convergence and performance guaranteed**. By the way of a unified variational characterization for the saddle-point of the problem and the generated iterates, we eventually show that IPS-ADMM under the generalized Lipschitz condition 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 e.g. [5, 21, 25]. A key step for proving the convergence of IPS-ADMM is to estimate the lower bound of $\left\| w^k - \widetilde{w}^k \right\|_{G_k}^2$.

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 the real 3D CT reconstruction problem in medical imaging 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 variational reformulation for the iterates generated by IPS-ADMM. Then, we analyze the convergence of IPS-ADMM according to the first-order optimality conditions 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

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

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

$$\begin{cases} f(x) - f(x^*) + \big\langle x - x^*, -A^\top \lambda^* \big\rangle \ge 0, \\ g(y) - g(y^*) + \big\langle y - y^*, -B^\top \lambda^* \big\rangle \ge 0, \\ Ax^* + By^* - b = \mathbf{0}. \end{cases}$$

Write these inequalities in a more compact form to obtain

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

namely,

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

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

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

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 the sequence can be characterized as a similar variational inequality to (3.2). To verify this conjecture, we first provide the following lemma about the iterates generated by the routine `xsub`.

LEMMA 3.1. *Let $\delta_t = \nabla f(\widehat{x}_t) - d_t$ and $\widetilde{\mathcal{D}}_k = \frac{1}{\alpha_k}\mathcal{M}_k - \beta A^\top A$. Suppose $\vartheta_k \in (0, 1/v)$ and Assumption 2.1 holds. Then, the iterates generated by IPS-ADMM satisfy*

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

where $\widetilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b)$ and

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

Proof. On the one hand, for any $j > 0$ we have by [5, Lemma 3.2] that

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

where $\mathcal{D}_k$ and ζ^k are given by (2.2) and (3.5) repsectively. If $j = 0$, then (3.6) becomes (3.4) with $\alpha_k = 1$ and $\overline{x}^k = x^{k+1}$. On the other hand, it follows from the inequality in the fourth step of IPS-ADMM that

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

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

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

Notice that by the expansion step $x^{k+1} = \alpha_k \overline{x}^k + (1 - \alpha_k)x^k$, we have

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

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

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

Inserting this relation into (3.8) with simple algebra confirms the result in (3.4) immediately. □

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

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

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

for all $w \in \Omega$, where

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

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

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

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

$$\begin{aligned} &g(y) - g(\widetilde{y}^k) + \Big\langle y - \widetilde{y}^k, -B^\top \big(\widetilde{\lambda}^k + (\tau - 1)(\widetilde{\lambda}^k - \lambda^k)\big) \\ &\quad + \beta B^\top (A\widetilde{x}^k + B\widetilde{y}^k - b) + L_0(\widetilde{y}^k - y^k) - d^{k+1} \Big\rangle \geq 0, \ \forall y \in \mathbb{R}^n. \end{aligned} \quad (3.12)$$

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

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

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

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

Finally, the result (3.9) is confirmed by the previous (3.4), (3.13) and (3.14). ∎

COROLLARY 3.3. *Suppose $\vartheta_k \in (0, 1/v)$. Then, we have*

$$\begin{aligned} &F(w) - F(\widetilde{w}^k) + \langle w - \widetilde{w}^k, \mathcal{J}(w) \rangle - \langle y - \widetilde{y}^k, d^{k+1} \rangle \\ \geq& \frac{1}{2}\Big(\|w - w^{k+1}\|^2_{H_k} - \|w - w^k\|^2_{H_k} \Big) + \frac{1}{2}\|w^k - \widetilde{w}^k\|^2_{G_k} + \zeta^k \end{aligned} \quad (3.15)$$

for all $w \in \Omega$, where

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

and

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

Proof. It holds by the updates of $\lambda^{k+1}, \widetilde{\lambda}^k$, and (3.11) that

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

that is,

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

Combine this equality together with $\widetilde{x}^k = x^{k+1}$ and $\widetilde{y}^k = y^{k+1}$ to have

$$w^k - w^{k+1} = P(w^k - \widetilde{w}^k), \quad \text{where } P = \begin{bmatrix} \mathbf{I} & & \\ & \mathbf{I} & \\ & -s\beta B & (\tau+s)\mathbf{I} \end{bmatrix} \tag{3.16}$$

is clearly invertible. Let

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

Then, we have from (3.3), (3.9) and (3.16) that

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

where the equality uses the following popularly used identity

$$\langle u - v, H_k(p - q) \rangle = \frac{1}{2}\Big\{ \|u - q\|_{H_k}^2 - \|u - p\|_{H_k}^2 + \|v - p\|_{H_k}^2 - \|v - q\|_{H_k}^2 \Big\} \tag{3.19}$$

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

Now, we focus on the last $\{\cdot\}$ term in (3.18). It holds by (3.16) 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

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

Then, (3.15) follows from (3.18) and the above discussions. □

3.2. More technical results. If both H_k and G_k are positive semidefinite matrices, then convergence of IPS-ADMM can be proved by Corollary 3.3. However, these two matrices are not necessarily positive semidefinite for any stepsize parameters (τ, s), which brings new challenges in the convergence analysis of IPS-ADMM. To proceed, a feasible scheme is to provide a condition to ensure $H_k \succeq \mathbf{0}$ and to estimate a lower bound of $\|w^k - \widetilde{w}^k\|_{G_k}^2$.

LEMMA 3.4. *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 first step of IPS-ADMM ensures $\widetilde{\mathcal{D}}_k \succeq \mathbf{0}$. 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. □

Notice that, the condition in Lemma 3.4 with $\gamma = 1$ is the same to that in [6, Lemma 4.3], which together with the condition below (2.5) shows that L is a positive semidefinite matrix. This suggests that our proximal matrix is more relaxed than that in [19]. Before evaluating the term $\|w^k - \widetilde{w}^k\|_{G_k}^2$, we estimate the involved crossing term $(Ax^k + By^k - b)^\top B(y^k - y^{k+1})$ in quadratic forms based on a variant of the Cauchy-Schwarz inequality.

LEMMA 3.5. *For any constants n_1, n_2 and $\delta \in (0, 1/2]$, it holds*

$$n_1 n_2 \langle Ax^k + By^k - b, B(y^k - y^{k+1})\rangle$$
$$\geq -(\frac{1}{4} + \frac{1}{2}\delta)n_1^2\|B(y^k - y^{k+1})\|^2 - (1-\delta)n_2^2\|Ax^k + By^k - b\|^2.$$

Moreover, take $\delta = \frac{\gamma-\eta}{2(1-\eta)}$ with $\eta \in (0,1)$ to obtain

$$n_1 n_2 \langle Ax^k + By^k - b, B(y^k - y^{k+1})\rangle$$
$$\geq -\frac{1+\gamma-2\eta}{4(1-\eta)}n_1^2\|B(y^k - y^{k+1})\|^2 - \frac{2-\gamma-\eta}{2(1-\eta)}n_2^2\|Ax^k + By^k - b\|^2. \quad (3.20)$$

Proof. For any $\delta \in (0,1)$, we have from the Cauchy-Schwarz inequality that

$$\begin{aligned} & n_1 n_2 \langle Ax^k + By^k - b, B(y^k - y^{k+1}) \rangle \\ \geq & -\frac{1}{4(1-\delta)} n_1^2 \|B(y^k - y^{k+1})\|^2 - (1-\delta) n_2^2 \|Ax^k + By^k - b\|^2. \end{aligned}$$

Let $\delta \in (0, 0.5]$, then $\frac{1}{4(1-\delta)} \leq \frac{1}{4} + \frac{1}{2}\delta$ and hence the first result is confirmed. In particular, by taking $\delta = \frac{\gamma-\eta}{2(1-\eta)}$ with $\eta \in (0,1)$, the inequality (3.20) is obtained. Finally, $\delta \in (0, 0.5]$ implies

$$0 < \frac{\gamma - \eta}{2(1-\eta)} \leq \frac{1}{2}.$$

The left inequality is ensured by $\eta \in (0,1)$ and $\gamma \in (\eta, 1]$ as in (2.5), while the right inequality is ensured by $\gamma \leq 1$. ∎

LEMMA 3.6. *Let*

$$\begin{cases} \omega_0 = \left(2 - \tau - s - \dfrac{(2-\eta-\gamma)(1-s)^2}{(1-\eta)(1+\tau)}\right)\beta, \\ \omega_1 = \dfrac{(2-\eta-\gamma)(1-s)^2}{(1-\eta)(1+\tau)}\beta, \quad \omega_2 = \dfrac{1-\tau}{1+\tau}, \\ \omega_3 = \left(\dfrac{\tau^2 + (1-3\gamma)\tau + 5\gamma - 4}{1+\tau} - \dfrac{(1+\gamma-2\eta)(1-\tau)^2}{2(1-\eta)(1+\tau)}\right)\beta. \end{cases} \tag{3.21}$$

Then, the iterates generated by IPS-ADMM satisfy

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

Proof. Combine the definition of $\widetilde{\lambda}^k$ in Lemma 3.1 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

$$\begin{aligned} \|w^k - \widetilde{w}^k\|_{G_k}^2 = & \|x^k - x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2 + \|y^k - y^{k+1}\|_{L + (\gamma - s)\beta B^\top B}^2 \\ & + 2(s-1)\langle \lambda^k - \widetilde{\lambda}^k, B(y^k - y^{k+1}) \rangle + \frac{2 - \tau - s}{\beta}\|\lambda^k - \widetilde{\lambda}^k\|^2 \\ = & \|x^k - x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2 + \|y^k - y^{k+1}\|_L^2 \\ & + (2-\tau-s)\beta\|Ax^{k+1} + By^{k+1} - b\|^2 + (\gamma - \tau)\beta\|B(y^k - y^{k+1})\|^2 \\ & + 2(1-\tau)\beta\langle Ax^{k+1} + By^{k+1} - b, B(y^k - y^{k+1}) \rangle. \end{aligned} \tag{3.23}$$

Next, we analyze the crossing term in the second equality of (3.23). Setting $y = y^k$ in (3.12) and $y = y^{k+1}$ in (3.12) at the $(k-1)$-th iteration respectively, we have

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

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 $1 + \tau > 0$)

$$\begin{aligned}
&\langle Ax^{k+1} + By^{k+1} - b, B(y^k - y^{k+1})\rangle \\
\geq & \frac{1-s}{1+\tau}\langle Ax^k + By^k - b, B(y^k - y^{k+1})\rangle - \frac{\tau}{1+\tau}\|B(y^k - y^{k+1})\|^2 \\
&+ \frac{1}{\beta(1+\tau)}\Big\langle y^k - y^{k+1}, (L - (1-\gamma)\beta B^\top B)\big[(y^k - y^{k+1}) - (y^{k-1} - y^k)\big]\Big\rangle \\
&+ \frac{1}{\beta(1+\tau)}\langle y^k - y^{k+1}, d^{k+1} - d^k\rangle \\
= & \frac{1-s}{1+\tau}\langle Ax^k + By^k - b, B(y^k - y^{k+1})\rangle - \frac{\tau}{1+\tau}\|B(y^k - y^{k+1})\|^2 \\
&+ \frac{1}{\beta(1+\tau)}\Big[\|y^k - y^{k+1}\|_L^2 - \langle y^k - y^{k+1}, L(y^{k-1} - y^k)\rangle - (1-\gamma)\beta\|B(y^k - y^{k+1})\|^2 \\
&+ (1-\gamma)\beta\langle y^k - y^{k+1}, B^\top B(y^{k-1} - y^k)\rangle\Big] + \frac{1}{\beta(1+\tau)}\langle y^k - y^{k+1}, d^{k+1} - d^k\rangle \\
\geq & \frac{1-s}{1+\tau}\langle Ax^k + By^k - b, B(y^k - y^{k+1})\rangle - \frac{2(1-\gamma)+\tau}{1+\tau}\|B(y^k - y^{k+1})\|^2 \\
&+ \frac{1}{2\beta(1+\tau)}\Big(\|y^k - y^{k+1}\|_L^2 - \|y^{k-1} - y^k\|_L^2 + (1-\gamma)\beta\big(\|B(y^k - y^{k+1})\|^2 \\
&- \|B(y^{k-1} - y^k)\|^2\big)\Big) + \frac{1}{\beta(1+\tau)}\langle y^k - y^{k+1}, d^{k+1} - d^k\rangle,
\end{aligned} \tag{3.24}$$

where the last inequality uses the Cauchy-Schwarz inequality. Combining (3.23) and (3.24), we immediately get

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

$$
\begin{aligned}
&\geq \|y^k - y^{k+1}\|_L^2 + \left(2-\tau-s-\frac{(2-\eta-\gamma)(1-s)^2}{(1-\eta)(1+\tau)}\right)\beta\|Ax^{k+1}+By^{k+1}-b\|^2 \\
&\quad + \|x^k - x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2 + \frac{(2-\gamma-\eta)(1-s)^2}{(1-\eta)(1+\tau)}\beta\left(\begin{array}{c}\|Ax^{k+1}+By^{k+1}-b\|^2 \\ -\|Ax^k+By^k-b\|^2\end{array}\right) \\
&\quad + \frac{1-\tau}{1+\tau}\left[\|y^k-y^{k+1}\|_L^2 - \|y^{k-1}-y^k\|_L^2 + (1-\gamma)\beta\left(\begin{array}{c}\|B(y^k-y^{k+1})\|^2 \\ -\|B(y^{k-1}-y^k)\|^2\end{array}\right)\right] \\
&\quad + \left(\gamma-\tau-\frac{2(2-2\gamma+\tau)(1-\tau)}{1+\tau}-\frac{(1+\gamma-2\eta)(1-\tau)^2}{2(1-\eta)(1+\tau)}\right)\beta\|B(y^k-y^{k+1})\|^2 \\
&\quad + \frac{2(1-\tau)}{1+\tau}\langle y^k-y^{k+1}, d^{k+1}-d^k\rangle \\
&= \|x^k - x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2 + \|y^k-y^{k+1}\|_L^2 \\
&\quad + \left(2-\tau-s-\frac{(2-\eta-\gamma)(1-s)^2}{(1-\eta)(1+\tau)}\right)\beta\|Ax^{k+1}+By^{k+1}-b\|^2 \\
&\quad + \frac{(2-\eta-\gamma)(1-s)^2}{(1-\eta)(1+\tau)}\beta\left(\|Ax^{k+1}+By^{k+1}-b\|^2 - \|Ax^k+By^k-b\|^2\right) \\
&\quad + \frac{1-\tau}{1+\tau}\left[\|y^k-y^{k+1}\|_L^2 - \|y^{k-1}-y^k\|_L^2 + (1-\gamma)\beta\left(\begin{array}{c}\|B(y^k-y^{k+1})\|^2 \\ -\|B(y^{k-1}-y^k)\|^2\end{array}\right)\right] \\
&\quad + \left(\frac{\tau^2+(1-3\gamma)\tau+5\gamma-4}{1+\tau}-\frac{(\gamma+1-2\eta)(1-\tau)^2}{2(1-\eta)(1+\tau)}\right)\beta\|B(y^k-y^{k+1})\|^2 \\
&\quad + \frac{2(1-\tau)}{1+\tau}\langle y^k-y^{k+1}, d^{k+1}-d^k\rangle,
\end{aligned}
$$

where the second inequality follows from (3.20) with specifications $n_1 := 1-\tau,\ n_2 := 1-s$. As a result, (3.22) holds with $\omega_i (i=0,1,2,3)$ given in (3.21). ◻

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

$$
2\left[F(w) - F(\widetilde{w}^k) + \langle w - \widetilde{w}^k, \mathcal{J}(w)\rangle\right] \geq \Phi_{k+1}(w) - \Phi_k(w) + \Psi_{k+1} + 2\zeta^k, \tag{3.25}
$$

where

$$
\begin{aligned}
\Phi_k(w) = &\|w - w^k\|_{H_k}^2 + \|y - v^k\|^2 + \omega_1\|Ax^k + By^k - b\|^2 \\
&+ (\omega_2 + \sigma_2)\|y^{k-1} - y^k\|_L^2 + \omega_2(1-\gamma)\beta\|B(y^k - y^{k+1})\|^2
\end{aligned} \tag{3.26}
$$

and

$$
\begin{aligned}
\Psi_k = &\|x^k - x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2 + (1-\sigma_1-\sigma_2)\|y^k - y^{k+1}\|_L^2 \\
&+ \overline{\omega}_0\|Ax^{k+1} + By^{k+1} - b\|^2 + \omega_3\|B(y^{k-1} - y^k)\|^2
\end{aligned} \tag{3.27}
$$

with $\overline{\omega}_0 = \left[(2-\tau-s)(1-\sigma_3) - \frac{(2-\eta-\gamma)(1-s)^2}{(1-\eta)(1+\tau)}\right]\beta.$

Proof. Rearrange (3.15) with $\widetilde{y}^k = y^{k+1}$ to get

$$\begin{aligned}
&2\big[F(w) - F(\widetilde{w}^k) + \langle w - \widetilde{w}^k, \mathcal{J}(w)\rangle\big] \\
&- \|w^k - \widetilde{w}^k\|_{G_k}^2 - 2\langle v^k - y^{k+1}, d^{k+1}\rangle + \|d^{k+1}\|^2 - 2\zeta^k \\
\geq&\Big(\|w - w^{k+1}\|_{H_k}^2 - \|w - w^k\|_{H_k}^2\Big) - 2\langle v^k - y, d^{k+1}\rangle + \|d^{k+1}\|^2 \\
=&\Big(\|w - w^{k+1}\|_{H_k}^2 - \|w - w^k\|_{H_k}^2\Big) + 2\langle y - v^k, v^k - v^{k+1}\rangle + \|v^k - v^{k+1}\|^2 \\
=&\Big(\|w - w^{k+1}\|_{H_k}^2 - \|w - w^k\|_{H_k}^2\Big) + \|y - v^{k+1}\|^2 - \|y - v^k\|^2,
\end{aligned} \tag{3.28}$$

where the first equality follows from the seventh step of IPS-ADMM and the last equality uses the aforementioned identity in (3.19).

Based on the previous inequalities (2.6) and (3.22) as well as the assumption that $\widetilde{\mathcal{D}}_k \succeq \widetilde{\mathcal{D}}_{k+1} \succeq \mathbf{0}$, it holds by (3.28) that

$$\begin{aligned}
&2\big[F(w) - F(\widetilde{w}^k) + \langle w - \widetilde{w}^k, \mathcal{J}(w)\rangle\big] \\
\geq&\Big(\|w - w^{k+1}\|_{H_{k+1}}^2 + \|y - v^{k+1}\|^2\Big) - \Big(\|w - w^k\|_{H_k}^2 + \|y - v^k\|^2\Big) \\
&+ \|w^k - \widetilde{w}^k\|_{G_k}^2 + 2\langle v^k - y^{k+1}, d^{k+1}\rangle - \|d^{k+1}\|^2 + 2\zeta^k \\
\geq&2\zeta^k + \frac{2(1-\tau)}{1+\tau}\langle y^k - y^{k+1}, d^{k+1} - d^k\rangle + 2\langle v^k - y^{k+1}, d^{k+1}\rangle - \|d^{k+1}\|^2 \\
&+ \|x^k - x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2 + \|y^k - y^{k+1}\|_L^2 + \omega_0\|Ax^{k+1} + By^{k+1} - b\|^2 + \omega_3\|B(y^{k-1} - y^k)\|^2 \\
&+ \Phi_{k+1}(w) - \Phi_k(w) + \sigma_2\Big(\|y^k - y^{k-1}\|_L^2 - \|y^{k+1} - y^k\|_L^2\Big) \\
\geq&2\zeta^k + \Phi_{k+1}(w) - \Phi_k(w) + \sigma_2\Big(\|y^k - y^{k-1}\|_L^2 - \|y^{k+1} - y^k\|_L^2\Big) \\
&+ \|x^k - x^{k+1}\|_{\widetilde{\mathcal{D}}_k}^2 + \|y^k - y^{k+1}\|_L^2 + \omega_0\|Ax^{k+1} + By^{k+1} - b\|^2 + \omega_3\|B(y^{k-1} - y^k)\|^2 \\
&- \sigma_1\|y^{k+1} - y^k\|_L^2 - \sigma_2\|y^k - y^{k-1}\|_L^2 - (2 - \tau - s)\sigma_3\beta\|Ax^{k+1} + By^{k+1} - b\|^2 \\
=&2\zeta^k + \Phi_{k+1}(w) - \Phi_k(w) + \Psi_{k+1}.
\end{aligned} \tag{3.29}$$

This completes the proof. □

REMARK 3.1. *We explain why the region of the dual stepsize is the previous* Δ *in (2.9).*

- *The first inequality in* Δ *comes from the proof of Lemma 3.4.*
- *The third inequality in (2.9) comes from the requirement that* $\overline{\omega}_0 \geq 0$, *i.e.,*

$$\gamma \geq 2 - \eta - \frac{(1-\eta)(2-\tau-s)(1-\sigma_3)(1+\tau)}{(1-s)^2}.$$

Combine the last inequality and the constraints $\gamma \leq 1$ *and* $\eta < 1$ *to obtain*

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

When $\sigma_3 = 0$, *it gives*

$$-\tau^2 - s^2 - \tau s + \tau + s + 1 \geq 0 \tag{3.30}$$

which reduces one of the constraints in the region of the dual stepsize [6]. Moreover, when $\tau = 0$ *we will have* $s \in (0, \frac{1+\sqrt{5}}{2}]$ *which is relatively larger than that in the classical ADMM. Besides,* $\overline{\omega}_0 \geq 0$ *implies* $\omega_0 \geq 0$.

- *The requirement on* $\omega_1 \geq 0$ *has been ensured by* $\gamma \in (\eta, 1]$ *and* $\eta \in (0.75, 1)$. *And we will explain the lower bound of* η *in the next item.*
- *The second inequality in (2.9) is obtained by requiring* $\omega_3 \geq 0$, *i.e.,*

$$\begin{aligned} & 0 \geq 2(1-\eta)\big[4-5\gamma+(3\gamma-1)\tau-\tau^2\big]+(1+\gamma-2\eta)(1-\tau)^2 \\ \Longleftrightarrow\ & 0 \geq (\gamma-1)\big[\tau^2+(4-6\eta)\tau+10\eta-9\big], \end{aligned}$$

which together with $\gamma < 1$ *shows*

$$\tau \leq 3\eta - 2 - \sqrt{(1-\eta)(13-9\eta)}.$$

Here, the identity $(1-\eta)(13-9\eta) > 0$ *has been ensured by* $\eta < 1$. *Finally, combine this inequality and the constraint* $1 \geq \tau > -1$ *to obtain* $\eta > 0.75$.

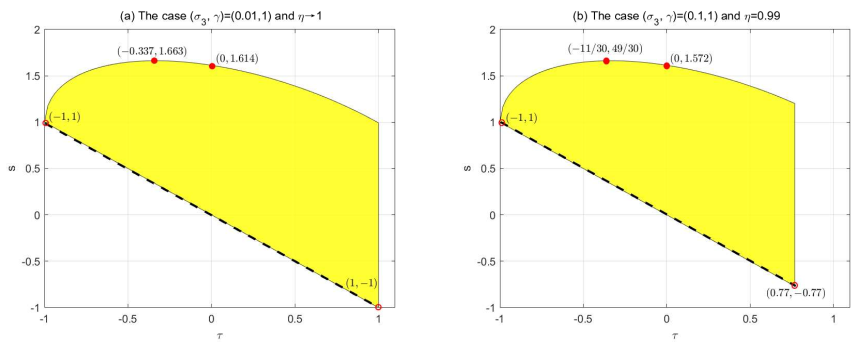

Fig. 3.1: Two instances of the stepsize region Δ_1

Remark 3.2. *According to the analysis of Lemma 3.7, especially the inequality (3.29), the following relative error criteria*

$$\begin{aligned} & \frac{2(1-\tau)}{1+\tau}\big|\langle y^k - y^{k+1}, d^{k+1}-d^k\rangle\big| + 2\big|\langle v^k - y^{k+1}, d^{k+1}\rangle\big| + \big\|d^{k+1}\big\|^2 \\ \leq & \sigma_1\big\|y^{k+1}-y^k\big\|_L^2 + \sigma_3\beta\big\|Ax^{k+1}+By^{k+1}-b\big\|^2,\ \ \sigma_1,\sigma_3\in[0,1), \end{aligned} \tag{3.31}$$

can be also exploited for the y-iterate. In this case, the parameter σ_2 *in (3.26)-(3.27) will be removed and* $\overline{\omega}_0 = \Big[(2-\tau-s) - \frac{(2-\eta-\gamma)(1-s)^2}{(1-\eta)(1+\tau)} - \sigma_3\Big]\beta$. *As a result, the stepsize parameters* (τ, s) *shall satisfy*

$$(\tau,s)\in\Delta_1 = \left\{(\tau,s)\,\middle|\, \begin{array}{l} \tau+s>0,\ \tau\leq 3\eta-2-\sqrt{(1-\eta)(13-9\eta)}, \\ (2-\gamma-\eta)(1-s)^2-(1-\eta)(1+\tau)(2-\tau-s-\sigma_3)\leq 0 \end{array}\right\}. \tag{3.32}$$

Figure 3.1 shows two instances of the region Δ_1, *which are similar to the first two subfigures of Figure 2.1. It follows from the last inequality in (3.32) with* $\gamma = 1$ *that*

$$\begin{aligned} & 0 \geq (1-\eta)[s^2-(1-\tau)s+\tau^2-(1-\sigma_3)\tau-(1-\sigma_3)] \\ \overset{\eta<1}{\Longrightarrow}\ & s \leq \frac{1}{2}\left(1-\tau+\sqrt{-3\tau^2+(4z-2)\tau+4z+1}\right) := f(\tau)\ \textit{where}\ z = 1-\sigma_3 \\ \overset{\nabla f(\tau)=0}{\Longrightarrow}\ & 0 = \big[3\tau+(2-z)\big](\tau-z) \Longrightarrow \tau = \frac{z-2}{3} = -\frac{1+\sigma_3}{3} \Longrightarrow s \leq \frac{5-\sigma_3}{3}. \end{aligned}$$

That is, the stepsize s could be $\frac{5-\sigma_3}{3}$ if we choose $\tau = -\frac{1+\sigma_3}{3}$ and $\gamma = 1$. One may choose $\sigma_1 = 0$ or $\sigma_3 = 0$ to derive a similar result to Lemma 3.7, and the difference lies in the region of the parameters (τ, s) and the terms in the right-hand-side of the relative error criteria for the y-iterate (i.e., users can exploit the y-residual and/or the equality error to control the solution quality), which suggests that more flexible relative error criteria could be employed for the y-subproblem.

3.3. Iteration complexity. Now, we present the following key convergence result based on the previous Lemma 3.7.

THEOREM 3.8. *Let $L \succeq (\tau-\gamma)\beta B^\top B$ and $(\tau, s) \in \Delta$. If for some integers $\kappa, N > 0$, the following conditions hold for all $k \in [\kappa, \kappa+N]$:*

(i) $\vartheta_k \in (0, 1/2\nu)$ and the sequence $\big\{\vartheta_k m_k(m_k+1)\big\}$ is nondecreasing;

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

Then, we have

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

for any $w \in \Omega$, where $\overline{w}^N = \frac{1}{1+N}\sum_{k=\kappa}^{\kappa+N}\widetilde{w}^k$ and $\Phi_k(w)$ is defined by (3.26).

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

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

Summing the inequality over k between κ and $\kappa+N$, we have by Lemma 4.4 that

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

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

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

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

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

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 ξ_t, we have $\mathbb{E}[\delta_t] = \mathbf{0}$ since the random variable $\xi_t \in \{1, 2, \ldots, K\}$ 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 ξ_{t-1}, ξ_{t-2}, ... The assumption $\mathbb{E}(\|\delta_t\|^2_{\mathcal{H}^{-1}}) \le \sigma^2$ together with $m_k \ge 1$ implies

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

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

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

Substitute the last inequality into (3.35) to confirm the result (3.33). □

Remark 3.3. *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 (3.25) together with the property in (3.1) and* $\Psi_k \ge 0$ *to obtain*

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

which, by the choice in (3.38), gives

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

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

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

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*

$$\left|\mathbb{E}\Big[F(\overline{w}^N) - F(w^*)\Big]\right| = E_\varrho(N) = \mathbb{E}\Big[\big\|Ax_N + By_N - b\big\|\Big],$$

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

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

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 we also compare it with a variety of state-of-the-art methods. 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.

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 $\mathcal{R}$ is the Radon transform generated by the cone beam scanning geometry [14] and ∇ is the discrete gradient operator. The size of 3D image to be constructed is $256 \times 256 \times 64$, the detector plane is 512×384 and the number of viewers is 668. Clearly, the size of the observed data b (that is N) is 131334144 and is a big data. Problem (4.1) is a a special case of (1.1) with $(A, B, b) = (\nabla, -\mathbf{I}, \mathbf{0})$ and applying IPS-ADMM to the problem (4.1) results in

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

where $\text{Shrink}(\cdot,\cdot)$ is the soft shrinkage operator and can be computed by the built-in MATLAB function "wthresh". Here, we compute y^{k+1} exactly (that is $d^{k+1} = \mathbf{0}$) since the corresponding subproblem has a closed-form solution and we do not need to solve it inexactly. 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 iterations $m_k = 10$, and we use the tuned parameters $(\mathcal{H}, \mathcal{M}_k) = (10^{-5}\mathbf{I}, 10^{-2}\mathbf{I})$, $\iota = 10^{-9}$, and $(\tau, s, \gamma, \eta) = (0.75, 1.09, 0.8, 0.99)$. By these choices, the proximal matrix L_0 given by (2.5) is positive indefinite and parameters satisfy (2.9).

In order to quantitatively evaluate the performance of IPS-ADMM, we consider four state-of-the-art baselines: stochastic ADMM (sto-ADMM, [20]) 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 evaluated by the following 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 calculate 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.

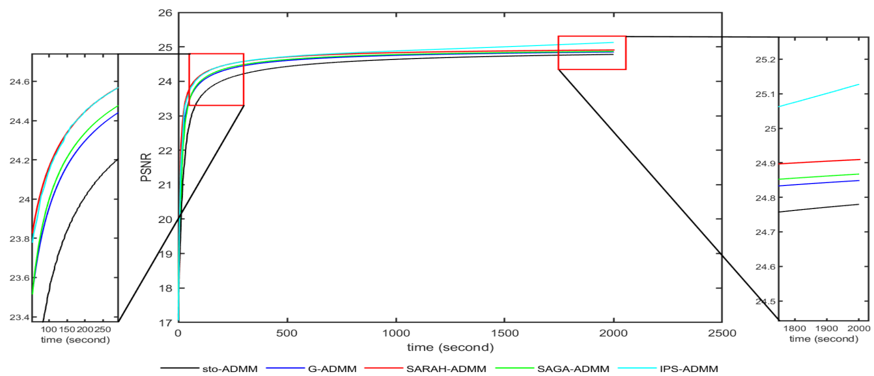

Fig. 4.1: PSNR results of five comparative algorithms

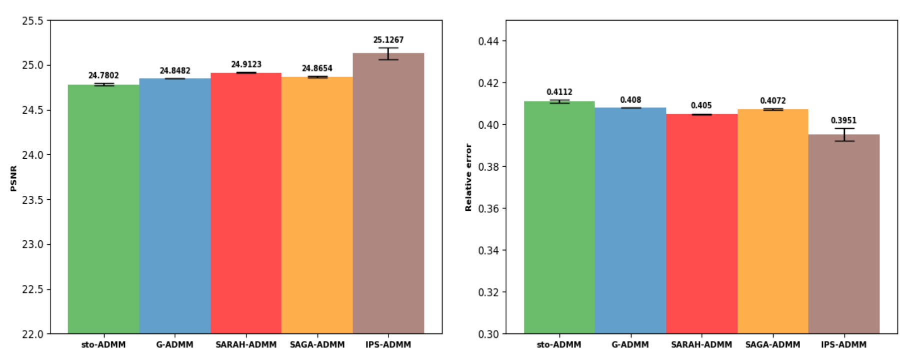

Fig. 4.2: The average PSNR and relative error of each algorithm

We report some experimental results in Figures 4.1-4.4. 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. 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 visuralize 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.

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., (2023) doi: 10.1109/TMC.2023.3262514.

[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, (2023), doi: 10.1007/s40305-023-00470-8.

[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] J. Bai, and Y. Ma, *A family of accelerated inexact augmented Lagrangian methods with applications to image restoration*, Optimization Online, (2022), https://optimization-online.org/2022/04/8872.

[8] 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.

[9] W. Deng, M. Lai, Z. Peng, and W. Yin, *Parallel multi-block ADMM with $o(1/k)$ convergence*, J. Sci. Comput., 71 (2017), pp. 712–736.

[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] 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.

[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 parélćments 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] 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.

[17] 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.

[18] 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.

[19] B. He, F. Ma, and X. Yuan, *Optimally linearizing the alternating direction method of multipliers for convex programming*, Comput. Optim. Appl., 75 (2020), pp. 361–388.

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

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

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

[23] J. Yang, and X. Yuan, *Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization*, Math. Comput., 82 (2013), pp. 301–329.

[24] N. Zhang, J. Wu, and L. Zhang, *A linearly convergent majorized ADMM with indefinite proximal terms for convex composite programming and its applications*, Math. Comput., 89 (2020), pp. 1867–1894.

[25] S. Zheng, and J. Kwok, *Fast-and-light stochastic ADMM*, The 25th Int. Joint Conf. Artificial Intelligence, (2016), pp. 2407–2413.

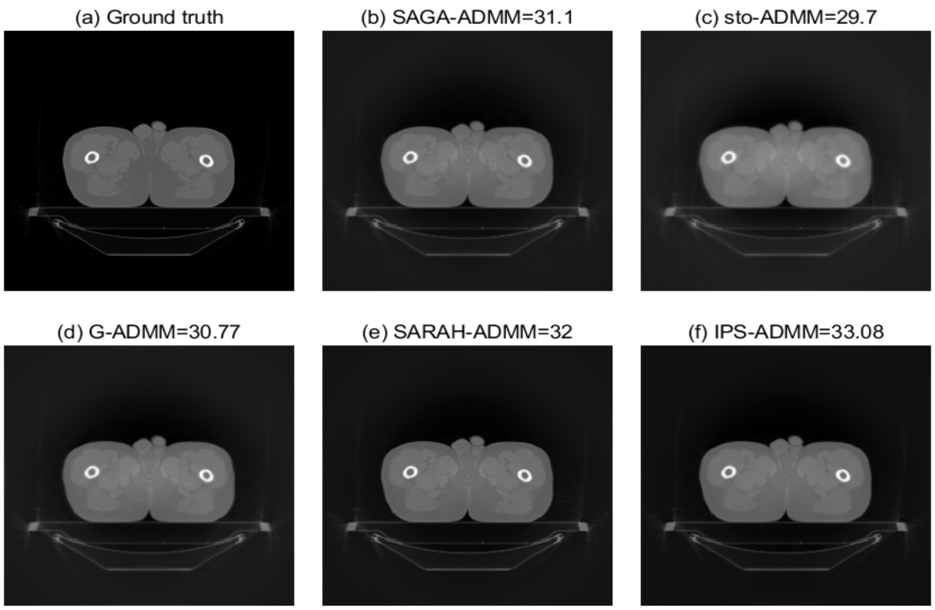

Fig. 4.3: Final reconstruction images of different methods for the **7th** slice

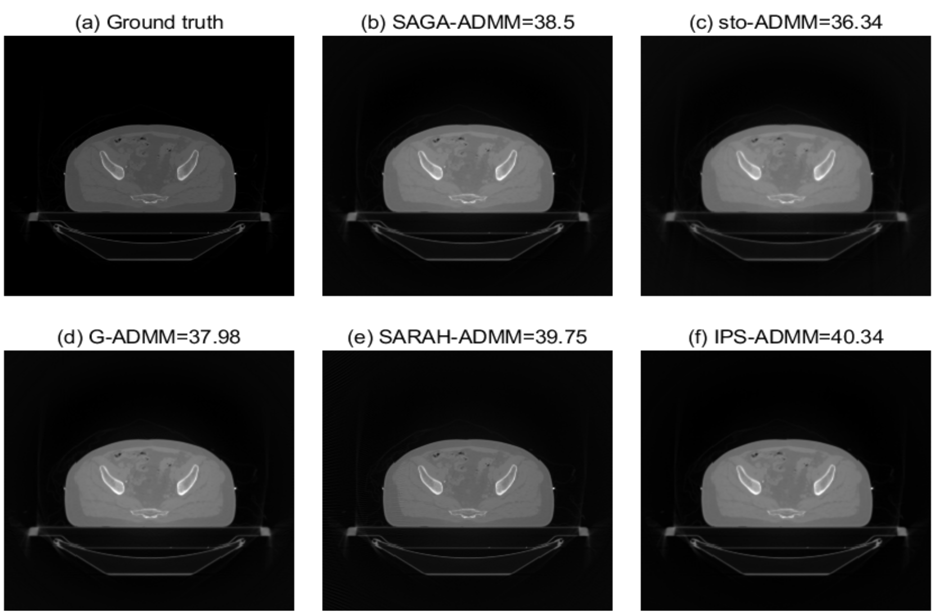

Fig. 4.4: Final reconstruction images of different methods for the **57th** slice