ON THE LINEAR CONVERGENCE OF THE GENERAL FIRST ORDER PRIMAL-DUAL ALGORITHM

KAI WANG
School of Science, Nanjing University of Science and Technology
Xiaolingwei 200, Nanjing, 210094, China

DEREN HAN∗
School of Mathematical Sciences
Beijing Advanced Innovation Center for Big Data and Brain Computing (BDBC)
Beihang University, Beijing, 100191, China

(Communicated by Jin-Yan Fan)

Abstract. In this paper, we consider the general first order primal-dual algorithm, which covers several recent popular algorithms such as the one proposed in [Chambolle, A. and Pock T., A first-order primal-dual algorithm for convex problems with applications to imaging, J. Math. Imaging Vis., 40 (2011) 120-145] as a special case. Under suitable conditions, we prove its global convergence and analyze its linear rate of convergence. As compared to the results in the literature, we derive the corresponding results for the general case and under weaker conditions. Furthermore, the global linear rate of the linearized primal-dual algorithm is established in the same analytical framework.

1. Introduction. Consider the following saddle point problem
\[
\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \Phi(x, y) := \theta_1(x) - \langle y, Ax \rangle - \theta_2(y),
\]
where \( A \in \mathbb{R}^{m \times n} \) is a given matrix; \( \mathcal{X} \) and \( \mathcal{Y} \) are two nonempty, closed and convex subsets of \( \mathbb{R}^n \) and \( \mathbb{R}^m \), respectively; \( \theta_1 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) and \( \theta_2 : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\} \) are proper lower semi-continuous convex functions; \( \langle \cdot, \cdot \rangle \) denotes the standard inner product of vectors; and \( \| \cdot \| \) is the Euclidean norm. The solution set of (1) is assumed to be nonempty throughout our discussion. In the fields of constrained minimization problems, zero-sum games, statistical learning and image processing (see [3, 5, 6, 8, 10, 14, 19]), a variety of application problems can be formulated into (1).

In order to solve the saddle point problem (1), a series of primal-dual algorithms have been extensively studied with various versions and different names. The so-called First-Order Primal-Dual Algorithm (FOPDA) was developed by Chambolle...
and Pock in [3], whose iterative scheme is given as

\[
\begin{align*}
    x^{k+1} &= \arg \min_{x \in X} \left\{ \Phi(x, y^k) + \frac{r}{2} \|x - x^k\|^2 \right\}, \\
    \bar{x}^k &= x^{k+1} + \tau(x^{k+1} - x^k), \\
    y^{k+1} &= \arg \max_{y \in Y} \left\{ \Phi(\bar{x}^k, y) - \frac{s}{2} \|y - y^k\|^2 \right\},
\end{align*}
\] (2a)

where \(\tau\) is a relaxation (extrapolation) parameter, and \(r > 0\) and \(s > 0\) are proximal parameters (step sizes). FOPDA has been successfully applied to solve image processing problems. As delineated in [3], the Primal-Dual Hybrid Gradient (PDHG) method [13, 27], as well as the Arrow-Hurwicz-Uzawa method [1], can be regarded as special cases of the FOPDA method with different parameters setting. The study of FOPDA can be summarized in three aspects.

1. If \(\tau = 0\), FOPDA method reduces to PDHG, whose convergence was established in [6] under some restrictive conditions on its step sizes. Then, [13] provided an example to demonstrate that PDHG may fail to converge even when the step sizes are selected as tiny constants.

2. If \(\tau = 1\), FOPDA method reduces to the modified Primal-Dual Hybrid Gradient method (PDHG-M) and [3] analyzed the global convergence and the worst case \(O(1/k)\) complexity under the assumption that \(rs > \|A^\top A\|\). Additionally, the \(O(1/k^2)\) sub-linear and linear rate of convergence were also established under the conditions that one of the functions is strongly convex and both are strongly convex, respectively.

3. Relaxed requirements on the parameters but with additional conditions on the functions, or with the cost of an additional step. The numerical results in [2, 14] indicated that FOPDA [3] could achieve better numerical performance if its parameters are tuned beyond their theoretically guaranteed limits. Motivated by these results, a set of primal-dual algorithms in prediction-correction fashion were presented in [14] where the range of the involved parameters in [3] was enlarged to further enhance the efficiency and flexibility of FOPDA. In the same vein, an improved version of the algorithm in [14] was put forward under the umbrella of prediction-correction in [2], which further relaxes the condition of \(\tau\) and also improves its computational efficiency. To simplify the sub-problem, [16] presented a linearized primal-dual algorithm by incorporating generalized proximal regularization technique. Moreover, [22] put forward a linearized primal-dual method for solving PDE problems and analyzed the sub-linear convergence rate in the ergodic sense. [4] presented some generalizations of the primal-dual methods, which include non-linear proximal regularization and inertial variants. In [12], an algorithmic framework of generalized PDHG scheme was proposed, which updates the output by correction steps with constant step sizes; they also presented the algorithmic framework of convergence analysis and analyzed the sub-linear convergence rate measured by the iteration complexity in both ergodic and nonergodic senses. Based on the assumption of the metric subregularity, [23] proved a posteriori linear convergence of the modified PDHG for solving some image processing problems. By combining a line-search technique, [18] developed a new primal-dual method. Then, [17] and [21] proposed some inexact primal-dual approaches by adopting some inexact rules, respectively. Furthermore,
[24] proposed a double extrapolation primal-dual algorithm for saddle point problems.

The aim of this paper is to analyze the global convergence and the rate of linear convergence of FOPDA for the general case that $\tau \in \mathbb{R}$, and without the correction step. We focus on the case that one of the functions, e.g., $\theta_1$, is strongly convex, and there is no restrict on the parameter $\tau$. We first prove the global convergence for this general case. Moreover, we prove the linear rate of convergence of the first order primal-dual method for the more general case of $\tau \in \mathbb{R}$, under the further assumption that $\nabla \theta_1$ is Lipschitz continuous. Note that this condition is also weaker than the required condition in [3], where the same result was established under the condition that both $\theta_1(\cdot)$ and $\theta_2(\cdot)$ are strongly convex. Under an alternative condition, i.e., the error bound condition, we also derive the global linear rate of convergence. In addition, we extend our theoretical results to the linearized version of FOPDA and prove the global linear convergence rate under two different scenarios.

The remainder of this paper is organized as follows. In Section 2, we summarize some basic notations and definitions, and describe an equivalent variational inequality characterization of the saddle point problem (1). In Section 3 we present the algorithmic framework of FOPDA, and in Section 4 we prove the global convergence of this primal-dual method without any requirement on the relaxation parameter $\tau$. In Section 5, we establish the global linear rate of convergence of FOPDA. In Section 6, we propose a linearized primal-dual algorithm and derive the corresponding global linear rate. Section 7 demonstrates some numerical simulation results, and finally, Section 8 draws the conclusions.

2. Preliminaries. We begin this section by summarizing some notations and definitions.

2.1. Notations and definitions. Let $\mathbb{R}^n$ be an $n$-dimensional Euclidean space. The symbol $\top$ represents the transpose. For a given symmetric and positive definite matrix $H$, we let $\|x\|_H = \sqrt{\langle x, Hx \rangle}$ be the $H$-norm of $x$. Furthermore, we denote $\|N\|$ to be the matrix norm of an arbitrary matrix $N$,

$$\|N\| := \sup_{x \neq 0} \left\{ \frac{\|Nx\|}{\|x\|} \right\}.$$  

**Definition 2.1.** A function $\theta(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is said to be convex, if

$$\theta(tu + (1 - t)v) \leq t\theta(u) + (1 - t)\theta(v), \quad \forall u, v \in \mathbb{R}^n, \; t \in [0, 1].$$

Furthermore, $\theta(\cdot)$ is said to be $\mu$-strongly convex if there exists a constant $\mu > 0$ such that

$$\theta(tu + (1 - t)v) \leq t\theta(u) + (1 - t)\theta(v) - t(1 - t)\frac{\mu}{2}\|u - v\|^2, \quad \forall u, v \in \mathbb{R}^n, \; t \in [0, 1].$$

**Definition 2.2.** Let $\theta(\cdot) : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function and the domain of function $\theta(\cdot)$ is denoted by $\text{dom } \theta$. Then, the subdifferential of $\theta(\cdot)$ at a point $v \in \text{dom } \theta$ is defined by

$$\partial \theta(v) = \left\{ \xi \mid \theta(u) \geq \theta(v) + \langle \xi, u - v \rangle, \; \forall u \in \text{dom } \theta \right\},$$

and the vector $\xi$ is said to be a subgradient of $\theta(\cdot)$ at $v$. 
Accordingly, if $\xi$ represents the subgradient of a $\mu$-strongly convex function $\theta(\cdot)$ at a point $v \in \text{dom } \theta$, from [20] it follows that
\[
\theta(u) \geq \theta(v) + \langle \xi, u - v \rangle + \frac{\mu}{2} \|u - v\|^2, \quad \forall u \in \text{dom } \theta.
\]

**Definition 2.3.** An operator $f : \Omega \to \mathbb{R}^n$ is said to be Lipschitz continuous on $\Omega$ if there exists a constant $L_f > 0$ such that
\[
\|f(x_1) - f(x_2)\| \leq L_f \|x_1 - x_2\|, \quad \forall x_1, x_2 \in \Omega. \tag{3}
\]

Now, let $(x^*, y^*)$ be a solution of the saddle point problem (1). Then, from saddle point optimality conditions, we get that
\[
\Phi(x^*, y) \leq \Phi(x^*, y^*) \leq \Phi(x, y^*), \quad \forall x \in \mathcal{X}, \quad \forall y \in \mathcal{Y},
\]
which reduces to the following mixed variational inequalities (MVI):
\[
\begin{aligned}
\theta_1(x) - \theta_1(x^*) + \langle x - x^*, -A^\top y^* \rangle &\geq 0, \quad \forall x \in \mathcal{X}, \\
\theta_2(y) - \theta_2(y^*) + \langle y - y^*, Ax^* \rangle &\geq 0, \quad \forall y \in \mathcal{Y}.
\end{aligned}
\]
The above variational characterization can be compactly rewritten as a problem $\text{MVI}(\mathcal{U}, \theta, \mathcal{G})$ of finding $u^* \in \mathcal{U}$, such that
\[
\theta(u) - \theta(u^*) + \langle u - u^*, \mathcal{G}(u^*) \rangle \geq 0, \quad \forall u \in \mathcal{U}, \tag{4a}
\]
where
\[
u := \left(\begin{array}{c} x \\ y \end{array}\right), \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad \mathcal{G}(u) := \left(\begin{array}{c} -A^\top y \\ Ax \end{array}\right), \quad \text{and } \mathcal{U} := \mathcal{X} \times \mathcal{Y}. \tag{4b}
\]
The underlying mapping $\mathcal{G}$ defined in (4b) is skew-symmetric, i.e.,
\[
\langle u_1 - u_2, \mathcal{G}(u_1) - \mathcal{G}(u_2) \rangle = 0, \quad \forall u_1, u_2 \in \mathcal{U}.
\]
Hence, $\mathcal{G}$ is monotone. As the solution set of (1) is assumed to be nonempty, the solution set of problem (4), denoted by $\mathcal{U}^*$, is also nonempty.

### 2.2. Projection operator and its properties.

Let $\mathcal{C}$ be a nonempty closed convex set of $\mathbb{R}^n$ and let $P_{\mathcal{C}}$ be the projection operator from $\mathbb{R}^n$ onto $\mathcal{C}$,
\[
P_{\mathcal{C}}(x) = \arg \min_{z \in \mathcal{C}} \|x - z\|_2.
\]
The projection operator $P_{\mathcal{C}}$ plays an important role in the field of convex analysis, which has many interesting properties and can be utilized in our paper. A property is that $P_{\mathcal{C}}$ is a nonexpansive map,
\[
\|P_{\mathcal{C}}(x) - P_{\mathcal{C}}(z)\| \leq \|x - z\|, \quad \forall x, z \in \mathbb{R}^n. \tag{5}
\]
We use $\text{dist}(x, \mathcal{C}) = \min_{y \in \mathcal{C}} \|x - y\|$ to denote the distance from a vector $x \in \mathbb{R}^n$ to a set $\mathcal{C} \subset \mathbb{R}^n$, and $\text{dist}_M(x, \mathcal{C}) = \min_{y \in \mathcal{C}} \|x - y\|_M$ denotes the distance in the sense of matrix norm, where $M$ is a given symmetric and positive definite matrix.

The mixed variational inequality (4) can be also equivalently transformed into some generalized projection equations.

**Lemma 2.4.** The mixed variational inequality (4) ($\text{MVI}(\mathcal{U}, \theta, \mathcal{G})$) amounts to finding $u^*$, such that $0 \in \mathcal{E}(u^*, \gamma)$, i.e.
\[
\text{dist}^2(0, \mathcal{E}(u^*, \gamma)) = 0,
\]
where the set-valued mapping $\mathcal{E}(u^*, \gamma)$ is defined as
\[
\mathcal{E}(u, \gamma) := \left(\begin{array}{c}
\mathcal{E}_\mathcal{X}(u, \gamma) := x - P_{\mathcal{X}}[x - \gamma(x - A^\top y)] \\
\mathcal{E}_\mathcal{Y}(u, \gamma) := y - P_{\mathcal{Y}}[y - \gamma(y - Ax)]
\end{array}\right), \tag{6}
\]
where $\xi_x \in \partial \theta_1(x)$, $\zeta_y \in \partial \theta_2(y)$, and $\gamma > 0$ is an arbitrary scalar.

In this paper, our convergence rate analysis under the error bound condition is based on the variational characterization (4) and the associated theory of variational inequalities. Since $\mathcal{U}^*$ denotes the solution set of $\text{MVI}(\mathcal{U}, \theta, \mathcal{G})$, it follows that

$$\mathcal{U}^* = \{u | \text{dist}(0, E(u^*, \gamma)) = 0\}.$$ 

The following theorem is established in [26, Theorem 3.3] and plays a fundamental role in our linear rate of convergence analysis under error bound condition.

**Theorem 2.5.** Let $F$ be a piecewise linear multifunction. For any $\omega > 0$, there exists $\eta > 0$ such that

$$\text{dist}(u, F^{-1}(0)) \leq \eta \text{dist}(0, F(u)), \quad \forall \|u\| < \omega.$$ 

3. The general primal-dual algorithm. Define

$$M := \begin{pmatrix} \tau I_n & \tau A^T_m \\ \tau A & sI_m \end{pmatrix} \quad \text{and} \quad H := \begin{pmatrix} \tau I_n & \tau A^T_m \\ \tau A & (s - 2\mu^{-1}(1 - \tau)^2\|A^T_m A\|)I_m \end{pmatrix},$$

where $\mu > 0$ is the strongly convex modulus of $\theta_1$, $\tau \in \mathbb{R}$, $r > 0$, $s > 2\mu^{-1}(1 - \tau)^2\|A^T_m A\| \geq 0$ and $r(s - 2\mu^{-1}(1 - \tau)^2\|A^T_m A\|) > \tau^2\|A^T_m A\|$. Under these conditions, the matrices $M$ and $H$ are positive definite.

**Remark 1.** The positive definiteness of the matrices $M$ and $H$ are crucial for the convergence analysis of primal-dual algorithms. Note that if $\tau = 1$, then $\mu$ does not play any role for the positive definiteness of the matrices $M$ and $H$, i.e., $M$ and $H$ are positive definite under the condition that $rs > \|A^T_m A\|$. Then, we can recover the convergence result in [3] without the strong convexity assumption on $\theta_1$.

Now, we are ready to formally present the algorithmic framework of the primal-dual algorithm.

**Algorithm 1** A General First-Order Primal-Dual Algorithm.

1. Select $r > 0$, $s > 0$, $\tau \in \mathbb{R}$ and $r(s - 2\mu^{-1}(1 - \tau)^2\|A^T_m A\|) > \tau^2\|A^T_m A\|$.

2. for $k = 0, 1, 2, \ldots$ do

3. Generate the points $u^{k+1} := (x^{k+1}, y^{k+1})$ via

$$\begin{cases} x^{k+1} = \arg \min_{x \in \mathcal{X}} \left\{ \Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \right\}, & (8a) \\
\bar{x}^k = x^{k+1} + \tau(x^{k+1} - x^k), & (8b) \\
y^{k+1} = \arg \max_{y \in \mathcal{Y}} \left\{ \Phi(x^k, y) - \frac{s}{2}\|y - y^k\|^2 \right\}. & (8c) \end{cases}$$

4. end for

**Remark 2.** If $\tau = 0$, Algorithm 1 reduces to the primal-dual hybrid gradient method. In [13], the global convergence result can be proven under the condition that one of the objective functions is strongly convex.

**Remark 3.** If the matrix $H$ is positive definite, it is obvious that the matrix $M$ is also positive definite. If we want to obtain the converge result of Algorithm 1, we need to select the proximal parameters $r$ and $s$ such that

$$r(s - 2\mu^{-1}(1 - \tau)^2\|A^T_m A\|) > \tau^2\|A^T_m A\|.$$
Since \( \tau, \mu \) and \( A \) are known, we can choose the proximal parameters by the following inequalities

\[
\frac{1}{rs} < \frac{1}{\|A^\top A\|} \leq \frac{1}{(2r\mu^{-1}(1-\tau)^2 + r^{-1}\tau^2)\|A^\top A\|}.
\]

If \( \tau \neq 1 \) and \( 2r\mu^{-1}(1-\tau)^2 + \tau^2 \leq 1 \), we have

\[
\frac{1}{rs} < \frac{1}{\|A^\top A\|} \leq \frac{1}{(2r\mu^{-1}(1-\tau)^2 + \tau^2)\|A^\top A\|}.
\]

4. Global convergence analysis. In this section, we prove that Algorithm 1 is globally convergent to a solution of (1).

Lemma 4.1. Let the sequence \( \{u_k := (x^k, y^k)\} \) be generated by Algorithm 1. Then, we have

\[
\theta(u) - \theta(u^{k+1}) + \langle u - u^{k+1}, G(u^{k+1}) \rangle \\
\geq \langle u^{k+1} - u, Q(u^{k+1} - u^k) \rangle + \frac{\mu}{2}\|x^{k+1} - x^k\|^2, \quad \forall u \in U,
\]

where \( G(u^{k+1}) \) is given in (4b) and

\[
Q := \begin{pmatrix} rI_n & A^\top \\ \tau A & sI_m \end{pmatrix}.
\]

Proof. By invoking the first-order optimality conditions of (8a) and (8c), we have

\[
\langle x - x^{k+1}, \xi_x^{k+1} - A^\top y^k + r(x^{k+1} - x^k) \rangle \geq 0, \quad \forall x \in \mathcal{X},
\]

and

\[
\langle y - y^{k+1}, \zeta_y^{k+1} + A[(1 + \tau)x^{k+1} - \tau x^k] + s(y^{k+1} - y^k) \rangle \geq 0, \quad \forall y \in \mathcal{Y},
\]

where \( \xi_x^{k+1} \in \partial \theta_1(x^{k+1}) \) and \( \zeta_y^{k+1} \in \partial \theta_2(y^{k+1}) \).

Then, using the strong convexity of \( \theta_1(x) \) and the convexity of \( \theta_2(y) \) respectively, we have

\[
\theta_1(x) - \theta_1(x^{k+1}) \geq \langle x - x^{k+1}, \xi_x^{k+1} \rangle + \frac{\mu}{2}\|x^{k+1} - x\|^2, \quad \forall x \in \mathcal{X},
\]

and

\[
\theta_2(y) - \theta_2(y^{k+1}) \geq \langle y - y^{k+1}, \zeta_y^{k+1} \rangle, \quad \forall y \in \mathcal{Y}.
\]

Combining (11) and (13), (12) and (14) yields

\[
\theta_1(x) - \theta_1(x^{k+1}) + \langle x - x^{k+1}, -A^\top y^k + r(x^{k+1} - x^k) \rangle \geq \frac{\mu}{2}\|x^{k+1} - x\|^2, \quad \forall x \in \mathcal{X},
\]

and

\[
\theta_2(y) - \theta_2(y^{k+1}) + \langle y - y^{k+1}, A[(1 + \tau)x^{k+1} - \tau x^k] + s(y^{k+1} - y^k) \rangle \geq 0, \quad \forall y \in \mathcal{Y}.
\]
Adding the above inequalities
\[ \theta(u) - \theta(u^{k+1}) - \frac{\mu}{2} \|x^{k+1} - x\|^2 \]
\[ + \left( x - x^{k+1} \right)^T \left( -A^T y^{k+1} \right) + \left( -\mu \|x^{k+1} - x\|^2 \right) \geq 0. \]

Then, making use of the definitions of \( G(\cdot) \) and \( Q \), the assertion of this lemma is obtained.

**Lemma 4.2.** Let the sequence \( \{u^k := (x^k, y^k)\} \) be generated by Algorithm 1. Then,
\[ \theta(u) - \theta(u^{k+1}) + \langle u - u^{k+1}, G(u) \rangle + \langle u - u^{k+1}, M(u^{k+1} - u^k) \rangle \]
\[ \geq \frac{\mu}{4} \|x^{k+1} - x\|^2 - \frac{(1 - \tau)^2}{\mu} \|A^T (y^k - y^{k+1})\|^2, \quad \forall u \in \mathcal{U}, \]
where \( M \) is given in (7).

**Proof.** Notice that
\[ \langle u - u^{k+1}, G(u^{k+1}) \rangle = \langle u - u^{k+1}, G(u) \rangle. \] (16)

Substituting (16) into (9) leads to
\[ \theta(u) - \theta(u^{k+1}) + \langle u - u^{k+1}, G(u) \rangle \geq \langle u^{k+1} - u, Q(u^{k+1} - u^k) \rangle + \frac{\mu}{2} \|x^{k+1} - x\|^2. \] (17)

It follows from the definitions of \( Q \) and \( M \) that
\[ \langle u^{k+1} - u, Q(u^{k+1} - u^k) \rangle \]
\[ = \langle u^{k+1} - u, M(u^{k+1} - u^k) \rangle + (1 - \tau)(x^{k+1} - x, A^T (y^{k+1} - y^k)). \] (18)

Then, using Young’s inequality leads to
\[ (1 - \tau)(x^{k+1} - x, A^T (y^{k+1} - y^k)) \]
\[ \geq -\frac{\mu}{4} \|x^{k+1} - x\|^2 - \frac{(1 - \tau)^2}{\mu} \|A^T (y^k - y^{k+1})\|^2. \] (19)

Combining (18) and (19), we obtain
\[ \langle u^{k+1} - u, Q(u^{k+1} - u^k) \rangle \]
\[ \geq \langle u^{k+1} - u, M(u^{k+1} - u^k) \rangle \]
\[ - \frac{\mu}{4} \|x^{k+1} - x\|^2 - \frac{(1 - \tau)^2}{\mu} \|A^T (y^k - y^{k+1})\|^2. \] (20)

Thus, combining (17) and (20) yields the assertion of this lemma. \qed

**Remark 5.** If we use the Young’s inequality
\[ \langle x^{k+1} - x, A^T (y^{k+1} - y^k) \rangle \geq -\frac{\mu}{2} \|x^{k+1} - x\|^2 - \frac{1}{2\mu} \|A^T (y^k - y^{k+1})\|^2, \]
we obtain
\[ \theta(u) - \theta(u^{k+1}) + \langle u - u^{k+1}, G(u) \rangle + \langle u - u^{k+1}, M(u^{k+1} - u^k) \rangle \]
\[ \geq \frac{\mu \tau}{2} \|x^{k+1} - x\|^2 - \frac{(1 - \tau)^2}{2\mu} \|A^T (y^k - y^{k+1})\|^2, \quad \forall u \in \mathcal{U}. \]

If we set \( \tau = 0 \), we can recover the convergence result in [13] under the condition \( r > 0 \) and \( s > \mu^{-1} \|A^T A\| \). Here, we utilize the Young’s inequality (19) for facilitating the proof of linear convergence.
Lemma 4.3. Let the sequence \( \{ \mathbf{u}^k := (x^k, y^k) \} \) be generated by Algorithm 1. Then,
\[
\theta(\mathbf{u}) - \theta(\mathbf{u}^{k+1}) + \langle \mathbf{u} - \mathbf{u}^{k+1}, \mathcal{G}(\mathbf{u}) \rangle + \frac{1}{2} \| \mathbf{u} - \mathbf{u}^k \|^2_M - \frac{1}{2} \| \mathbf{u} - \mathbf{u}^{k+1} \|^2_M
\geq \frac{1}{2} \| \mathbf{u}^k - \mathbf{u}^{k+1} \|^2_H + \frac{\mu}{4} \| x^{k+1} - x \|^2, \quad \forall \mathbf{u} \in \mathcal{U},
\]
where \( M \) and \( H \) are given in (7), and \( \mathcal{G} \) is defined in (4b).

Proof. According to the identity
\[
\langle b, M(a-b) \rangle = \frac{1}{2} (\| a \|^2_M - \| b \|^2_M - \| a - b \|^2_M),
\]
and setting \( a = \mathbf{u} - \mathbf{u}^k \) and \( b = \mathbf{u} - \mathbf{u}^{k+1} \) we get
\[
\langle \mathbf{u} - \mathbf{u}^{k+1}, M(\mathbf{u}^{k+1} - \mathbf{u}^k) \rangle = \frac{1}{2} \| \mathbf{u} - \mathbf{u}^k \|^2_M - \frac{1}{2} \| \mathbf{u} - \mathbf{u}^{k+1} \|^2_M - \frac{1}{2} \| \mathbf{u}^{k+1} - \mathbf{u}^k \|^2_M.
\]
Then, substituting the above equality into (15) leads to
\[
\theta(\mathbf{u}) - \theta(\mathbf{u}^{k+1}) + \langle \mathbf{u} - \mathbf{u}^{k+1}, \mathcal{G}(\mathbf{u}) \rangle + \frac{1}{2} \| \mathbf{u} - \mathbf{u}^k \|^2_M - \frac{1}{2} \| \mathbf{u} - \mathbf{u}^{k+1} \|^2_M
\geq \frac{1}{2} \| \mathbf{u}^k - \mathbf{u}^{k+1} \|^2_M - \mu^{-1} (1 - \tau)^2 \frac{\| \mathbf{u}^{k+1} - \mathbf{u}^k \|^2}{\| A \|^2} (y^k - y^{k+1})^\top (y^k - y^{k+1}) + \frac{\mu}{4} \| x^{k+1} - x \|^2,
\geq \frac{1}{2} \| \mathbf{u}^k - \mathbf{u}^{k+1} \|^2_M - \frac{\mu}{4} \| x^{k+1} - x \|^2, \quad \forall \mathbf{u} \in \mathcal{U},
\]
where the last inequality is derived from the definition of \( H \).

The following theorem proves that the sequence \( \{ \mathbf{u}^k \} \) generated by Algorithm 1 is Fejér monotone with respect to the solution set of (4).

Theorem 4.4. Let \( \mathbf{u}^* \) be an arbitrary solution of (1). Then, the sequence \( \{ \mathbf{u}^k \} \) generated by Algorithm 1 satisfies
\[
\| \mathbf{u}^{k+1} - \mathbf{u}^* \|^2_M \leq \| \mathbf{u}^k - \mathbf{u}^* \|^2_M - \| \mathbf{u}^k - \mathbf{u}^{k+1} \|^2_H - \frac{\mu}{2} \| x^{k+1} - x^* \|^2.
\]

Proof. Setting \( \mathbf{u} = \mathbf{u}^* \) in the inequality (21) and using the fact
\[
\theta(\mathbf{u}^{k+1}) - \theta(\mathbf{u}^*) + \langle \mathbf{u}^{k+1} - \mathbf{u}^*, \mathcal{G}(\mathbf{u}^*) \rangle \geq 0,
\]
we yield
\[
\frac{1}{2} \| \mathbf{u}^* - \mathbf{u}^k \|^2_M - \frac{1}{2} \| \mathbf{u}^* - \mathbf{u}^{k+1} \|^2_M \geq \frac{\mu}{4} \| x^{k+1} - x^* \|^2 + \frac{1}{2} \| \mathbf{u}^k - \mathbf{u}^{k+1} \|^2_H.
\]
Hence, the assertion of the theorem is proved.

Following the line of reasoning presented in [13, Theorem 4.4], we can also prove that Algorithm 1 is globally convergent. Here we only state the following theorem and omit the proof.

Theorem 4.5. The sequence \( \{ \mathbf{u}^k \} \) generated by Algorithm 1 is globally convergent to a solution point of saddle point problem (1).

5. Convergence rate. In this section, we turn our attention to the linear rate of convergence of Algorithm 1.
5.1. Linear convergence under strongly convexity assumption. In this section, we prove the linear rate of convergence of Algorithm 1 under the following assumption.

**Assumption 1.** In problem (1), assume that $X = \mathbb{R}^n$, $A$ is full row rank, $\theta_2$ is convex, $\theta_1$ is strongly convex with modulus $\mu > 0$ and $\nabla \theta_1$ is Lipschitz continuous with constant $L$. For Algorithm 1, $r > |\tau||A^TA|$ and $s > |\tau|+2\mu^{-1}(1-\tau)^2||A^TA||$.

We now begin our analysis with a fundamental inequality, which relates the bound $\|y^{k+1} - y^*\|_2$ to the terms $\|x^{k+1} - x^*\|_2$, $\|x^{k+1} - x^k\|_2$ and $\|y^{k+1} - y^k\|_2$.

**Lemma 5.1.** Suppose Assumption 1 holds. Let $u^*$ be a solution of (1). Then, the sequence $\{u^k\}$ generated by Algorithm 1 satisfies

$$
\|y^{k+1} - y^*\|_2^2 \leq 3\kappa \left[ L^2 \|x^{k+1} - x^*\|_2^2 + \|A^TA\||y^{k+1} - y^k\|_2^2 + r^2 \|x^{k+1} - x^k\|_2^2 \right],
$$

where

$$
\kappa = \left( \lambda_{\min}(AA^T) \right)^{-1} > 0,
$$

and $\lambda_{\min}(\cdot)$ is the smallest eigenvalue of a positive definite matrix.

**Proof.** It follows from the full row rank of $A$ that

$$
\|y^{k+1} - y^*\|_2^2 \leq \kappa \|A^T(y^{k+1} - y^*)\|_2^2.
$$

Since $X = \mathbb{R}^n$, the first order optimality condition of the sub-problem (8a) can be written as

$$
\nabla \theta_1(x^{k+1}) - A^Ty^k + r(x^{k+1} - x^k) = 0.
$$

Since $(x^*, y^*)$ is a solution, we have

$$
\nabla \theta_1(x^*) = A^Ty^*.
$$

Then, combining the above two equations leads to

$$
\|A^Ty^{k+1} - A^Ty^*\|_2^2 = \|\nabla \theta_1(x^{k+1}) + A^T(y^{k+1} - y^k) + r(x^{k+1} - x^k) - \nabla \theta_1(x^*)\|_2^2
$$

$$
\leq 3 \|\nabla \theta_1(x^{k+1}) - \nabla \theta_1(x^*)\|_2^2 + 3\|A^T(y^{k+1} - y^k)\|_2^2 + 3r^2 \|x^{k+1} - x^k\|_2^2
$$

$$
\leq 3 \left[ L^2 \|x^{k+1} - x^*\|_2^2 + \|A^TA\||y^{k+1} - y^k\|_2^2 + r^2 \|x^{k+1} - x^k\|_2^2 \right],
$$

where the first inequality is derived by

$$
\|a + b + c\|_2^2 \leq 3\|a\|_2^2 + 3\|b\|_2^2 + 3\|c\|_2^2,
$$

and $L$ is the Lipschitz constant of $\nabla \theta_1(\cdot)$. Then, combining the inequalities (25) and (26) yields (23).

**Theorem 5.2.** Suppose Assumption 1 holds. Let $u^*$ be a solution of (1). Then, the sequence $\{u^k\}$ generated by Algorithm 1 converges Q-linearly, i.e.,

$$
\|u^{k+1} - u^*\|_M \leq \frac{1}{1+\delta} \|u^k - u^*\|_M,
$$

where $\delta := \min\{\delta_1, \delta_2, \delta_3\} > 0$ and

$$
\delta_1 := \frac{\mu}{2(1+3\kappa L^2)||M||},
$$

$$
\delta_2 := \frac{r|\tau||A^TA|}{3\kappa ||M||},
$$

$$
\delta_3 := \frac{s\tau}{3\kappa ||M||||A^TA||}.
$$
Linear convergence under error bound conditions.

5.2. Proof. Using the definition of the matrix $M$, we have

$$
\begin{align*}
\delta \| u^{k+1} - u^* \|_M^2 \\
\leq \delta \| M \| \left( \| x^{k+1} - x^* \|^2 + \| y^{k+1} - y^* \|^2 \right) \\
\leq \delta \| M \| \left( (1 + 3\kappa L^2) \| x^{k+1} - x^* \|^2 + 3\kappa \| A^T A \| \| y^{k+1} - y^* \|^2 \\
+ 3\kappa r^2 \| x^{k+1} - x^* \|^2 \right).
\end{align*}
$$

On the other hand,

$$
\begin{align*}
\| u^{k+1} - u_k \|_H^2 \\
= r \| x^{k+1} - x^* \|^2 + [s - 2\mu^{-1}(1 - \tau)^2 \| A^T A \|] \| y^{k+1} - y^* \|^2 \\
+ 2\tau (A(x^{k+1} - x^k), y^{k+1} - y^k) \\
\geq (r - |\tau| \| A^T A \|) \| x^{k+1} - x^* \|^2 + (s - 2\mu^{-1}(1 - \tau)^2 \| A^T A \| - |\tau|) \| y^{k+1} - y^* \|^2,
\end{align*}
$$

where the inequality is derived from the Young’s inequality. Then, combining the above two inequalities with (22) leads to

$$
\begin{align*}
(1 + \delta) \| u^{k+1} - u^* \|_M^2 \\
\leq \| u^k - u^* \|_M^2 - \| u^k - u^{k+1} \|_H^2 - \frac{\mu}{2} \| x^{k+1} - x^* \|^2 + \delta \| u^{k+1} - u^* \|_M^2 \\
\leq \| u^k - u^* \|_M^2 - \left( \frac{\mu}{2} - (1 + 3\kappa L^2) \| M \| \delta \right) \| x^{k+1} - x^* \|^2 \\
- (r - |\tau| \| A^T A \| - 3\kappa r^2 \| M \| \delta) \| y^{k+1} - y^* \|^2 \\
- (s - |\tau| - 2\mu^{-1}(1 - \tau)^2 \| A^T A \| - 3\kappa \| M \| \| A^T A \| \delta) \| y^{k+1} - y^* \|^2.
\end{align*}
$$

It follows from the definition of $\delta$ and the above inequality that

$$(1 + \delta) \| u^{k+1} - u^* \|_M^2 \leq \| u^k - u^* \|_M^2,$$

and the assertion of the theorem is proved. \hfill \Box

5.2. Linear convergence under error bound conditions. In this subsection, we prove the global linear convergence of Algorithm 1 under some error bound conditions.

Assumption 2. In problem (1), $\theta_2$ is convex, $\theta_1$ is strongly convex with modulus $\mu > 0$. For Algorithm 1, $r(s - 2\mu^{-1}(1 - \tau)^2 \| A^T A \|) > \tau^2 \| A^T A \|$. Furthermore, assume that for any $\omega > 0$, there exists $\eta > 0$, such that

$$
dist(u, U^*) \leq \eta \dist(0, E(u, 1)), \quad \forall \| u \| \leq \omega, \quad u \in U, \tag{28}
$$

where $E(\cdot)$ is defined by (6).

If the subdifferentials $\partial \theta_1$ and $\partial \theta_2$ are piecewise linear multi-functions and $X$ and $Y$ are polyhedral sets, then $E(u, \gamma)$ are piecewise linear multi-functions. According to Theorem 2.5, (28) holds, and this fact was adopted in [9, 11, 25] to prove the linear rate of convergence of alternating direction methods of multipliers.

Lemma 5.3. Let the sequence $\{ u^k := (x^k, y^k) \}$ be generated by Algorithm 1. Then, there exists a constant $\sigma_1 > 0$ such that

$$
\text{dist}^2(0, E(u^{k+1}, 1)) \leq \sigma_1 \{ \| y^k - y^{k+1} \|^2 + \| x^k - x^{k+1} \|^2 \}, \tag{29}
$$

where $\text{dist}(\cdot, \cdot)$ is the Hausdorff distance.
Theorem 5.4. Suppose Assumption 2 holds. Let \( \sigma \)
where

\[
\sigma_1 = \max\{2(\|A^T A\| + s^2), 2(r^2 + \tau^2 \|A^T A\|)\} > 0
\]

Proof. The optimality condition of \( x^-\)subproblem (8a) indicates that

\[
x^{k+1} = P_x[x^{k+1} - (\xi_x^{k+1} - A^T y^k + r(x^{k+1} - x^k))].
\]

Then,

\[
dist^2(0, E_x(u^{k+1}, 1))
= dist^2 (x^{k+1}, P_x(x^{k+1} - (\xi_x^{k+1} - A^T y^k)))
= \left\| P_x[x^{k+1} - (\xi_x^{k+1} - A^T y^k) - P_x[x^{k+1} - (\xi_x^{k+1} - A^T y^k)]\right\|^2
\leq \|A^T(y^k - y^{k+1}) + r(x^k - x^{k+1})\|^2
\leq 2\|A^T A\||y^k - y^{k+1}|^2 + 2\|\|y^k - y^{k+1}\|^2,
\]

(31)

where the second inequality follows from the inequality

\[
\|a + b\|^2 \leq 2a^2 + 2b^2, \quad \forall a, b \in \mathbb{R}^n,
\]

by setting \( a = A^T(y^k - y^{k+1}) \) and \( b = r(x^k - x^{k+1}) \). On the other hand, the optimality condition of the \( y^-\)subproblem (8c) implies that

\[
y^{k+1} = P_y[y^{k+1} - (\zeta_y^{k+1} + A x^{k+1} + \tau A(x^{k+1} - x^k) + s(y^{k+1} - y^k))].
\]

Hence, we obtain

\[
dist^2(0, E_y(u^{k+1}, 1))
= dist^2(y^{k+1}, P_y(y^{k+1} - (\zeta_y^{k+1} + A^T x^{k+1}))
\leq \left\| P_y[y^{k+1} - (\zeta_y^{k+1} + A^T x^{k+1} + \tau A(x^{k+1} - x^k) + s(y^{k+1} - y^k)] - P_y[y^{k+1} - (\zeta_y^{k+1} + A^T x^{k+1})]\right\|^2
\leq \|\tau A(x^{k+1} - x^k) + s(y^{k+1} - y^k)\|^2
\leq 2\tau^2\|A^T A\||x^{k+1} - x^k|^2 + 2s^2\|y^k - y^{k+1}\|^2,
\]

(32)

Then, combining (31) and (32) yields

\[
dist^2(0, E(u^{k+1}, 1))
= dist^2(0, E_x(u^{k+1}, 1)) + dist^2(0, E_y(u^{k+1}, 1))
\leq 2(\|A^T A\| + s^2)||y^k - y^{k+1}|^2 + 2(r^2 \|A^T A\| + \tau^2)\|x^k - x^{k+1}\|^2.
\]

The assertion of this lemma is derived from the definition of \( \sigma_1 \).

Now, we show the global linear convergence of Algorithm 1 under Assumption 2.

**Theorem 5.4.** Suppose Assumption 2 holds. Let \( \{u^k := (x^k, y^k)\} \) be generated by Algorithm 1. Then,

\[
(1 + \rho)dist_M^2(u^{k+1}, U^*) \leq dist_M^2(u^k, U^*),
\]

(33)

where

\[
\rho := \frac{1}{\eta^2 \sigma \|M\|},
\]

\( \sigma = \sigma_1/\lambda_{\min} (H) \), and \( \sigma_1 \) is defined in (30).
Proof. Let \((x^*, y^*)\) be a saddle point in \(\mathcal{U}^*\). Based on Theorem 4.5, the sequence \(\{(x^k, y^k)\}\) converges to a saddle point in \(\mathcal{U}^*\), which is bounded. Assumption 2, together with (29), implies that  
\[
\text{dist}^2(u^{k+1}, \mathcal{U}^*) \leq \eta^2 \sigma_1 \{ \|y^k - y^{k+1}\|^2 + \|x^k - x^{k+1}\|^2 \}.
\]
On the other hand,  
\[
\lambda_{\text{min}}(H)\{\|y^k - y^{k+1}\|^2 + \|x^k - x^{k+1}\|^2 \} \leq \|u^k - u^{k+1}\|_H^2.
\]
Then, setting \(\sigma = \sigma_1/\lambda_{\text{min}}(H)\), we conclude  
\[
\text{dist}^2(u^{k+1}, \mathcal{U}^*) \leq \eta^2 \sigma \|u^k - u^{k+1}\|_H^2.
\]
Note that  
\[
\|u^k - u^*\|_H^2 = \|u^k - u^*\|_2^2 + \|u^* - u^*\|_2^2 = \|u^k - u^*\|_2^2 + \|u^* - u^*\|_H^2 = \|u^k - u^*\|_2^2 + \|u^* - u^*\|_H^2 = \|u^k - u^*\|_2^2.
\]
and hence  
\[
\text{dist}^2_M(u^{k+1}, \mathcal{U}^*) \leq \eta^2 \sigma \|u^k - u^{k+1}\|_M^2. \tag{34}
\]
Then, it follows from Theorem 4.4 and (34) that  
\[
\|u^{k+1} - u^*\|_M^2 \leq \|u^k - u^*\|_M^2 - \|u^k - u^{k+1}\|_H^2 \leq \|u^k - u^*\|_M^2 - \frac{1}{\eta^2 \sigma \|M\|} \text{dist}^2_M(u^{k+1}, \mathcal{U}^*) \tag{35}
\]
Let \(u^* \in \mathcal{U}^*\) such that \(\text{dist}_M(u^k, \mathcal{U}^*) = \|u^k - u^*\|_M\). Then, it follows from (35) that  
\[
\text{dist}^2_M(u^{k+1}, \mathcal{U}^*) \leq \|u^k - u^*\|_M^2 - \frac{1}{\eta^2 \sigma \|M\|} \text{dist}^2_M(u^{k+1}, \mathcal{U}^*) \leq \|u^k - u^*\|_M^2 - \frac{1}{\eta^2 \sigma \|M\|} \text{dist}^2_M(u^{k+1}, \mathcal{U}^*) = \text{dist}^2_M(u^k, \mathcal{U}^*) - \frac{1}{\eta^2 \sigma \|M\|} \text{dist}^2_M(u^{k+1}, \mathcal{U}^*).
\]
Rearranging terms, we get the desired inequality immediately. \(\square\)

Remark 6. As mentioned in Remark 1, if \(\tau = 1\), we can recover the convergence result in [3] and obtain the inequality (22) without the strong convexity assumption on \(\theta_1\). In this case, we can also derive the linear convergence of Algorithm 1 under this error bound condition.

6. Linearized primal-dual algorithm. This section considers a concrete model of (1):  
\[
\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \Phi(x, y) := \frac{\nu}{2} \|x - b\|_B^2 - \langle y, Ax \rangle - \theta_2(y), \tag{36}
\]
where \(B \in \mathbb{R}^{n \times n}\) is a symmetric and positive definite matrix and \(\nu\) is a positive constant. Define  
\[
M' = \begin{pmatrix} rI_n - \nu B & \tau A^\top \\ \tau A & sI_m \end{pmatrix} \tag{37}
\]
and  
\[
H' = \begin{pmatrix} rI_n - \nu B & \tau A^\top \\ \tau A & (s - 2\mu^{-1}(1 - \tau)^2\|A^\top A\|)I_m \end{pmatrix}, \tag{38}
\]
where \(\mu = \nu \lambda_{\text{min}}(B) > 0\) is the strongly convex modulus of \(\theta_1\), and \(\tau \in \mathbb{R}\). If  
\[(r - \nu\|B\|)(s - 2\mu^{-1}(1 - \tau)^2\|A^\top A\|) > \tau^2\|A^\top A\|,
\]
the matrices \(M'\) and \(H'\) are positive definite.
Adding the above two inequalities and utilizing the definitions of \( G \) algorithm is described as follows.

**Lemma 6.1.** Let the sequence \( \{u^k \} \) be generated by Algorithm 2. Then, we have

\[
\theta(u) - \theta(u^{k+1}) + \langle u - u^{k+1}, G(u^{k+1}) \rangle \\
\geq \langle u^{k+1} - u, Q'(u^{k+1} - u^k) \rangle + \frac{\mu}{2} \|x^{k+1} - x^k\|^2, \quad \forall u \in U,
\]

where \( G(u^{k+1}) \) is given in (4b) and and

\[
Q' := \begin{pmatrix}
    rI_n - \nu B & A^T \\
    \tau A & sI_m
\end{pmatrix}.
\]

**Proof.** By invoking the first-order optimality conditions of (39a) and (39c), we have

\[
\langle x - x^{k+1}, \nu B(x^k - b) - A^T y^k + \tau (x^{k+1} - x^k) \rangle \geq 0, \quad \forall x \in \mathcal{X},
\]

and

\[
\langle y - y^{k+1}, \zeta_y^{k+1} + A[(1 + \tau)x^{k+1} - \tau x^k] + s(y^{k+1} - y^k) \rangle \geq 0, \quad \forall y \in \mathcal{Y},
\]

where \( \zeta_y^{k+1} \in \partial \theta_2(y^{k+1}) \).

Then, using the strong convexity of \( \theta_1(x) \) and the convexity of \( \theta_2(y) \) respectively, we have

\[
\theta_1(x) - \theta_1(x^{k+1}) \geq \langle x - x^{k+1}, \nu B(x^{k+1} - b) \rangle + \frac{\mu}{2} \|x^{k+1} - x\|^2, \quad \forall x \in \mathcal{X},
\]

and

\[
\theta_2(y) - \theta_2(y^{k+1}) \geq \langle y - y^{k+1}, \zeta_y^{k+1} \rangle, \quad \forall y \in \mathcal{Y},
\]

where \( \mu = \nu \lambda_{\text{min}}(B) \). Combining (42) and (44), (43) and (45) yields

\[
\theta_1(x) - \theta_1(x^{k+1}) + \langle x - x^{k+1}, -A^T y^k + (r I_n - \nu B)(x^{k+1} - x^k) \rangle \geq \frac{\mu}{2} \|x^{k+1} - x\|^2, \quad \forall x \in \mathcal{X},
\]

and

\[
\theta_2(y) - \theta_2(y^{k+1}) + \langle y - y^{k+1}, A[(1 + \tau)x^{k+1} - \tau x^k] + s(y^{k+1} - y^k) \rangle \geq 0, \quad \forall y \in \mathcal{Y}.
\]

Adding the above two inequalities and utilizing the definitions of \( G(\cdot) \) and \( Q' \), the assertion of this lemma is obtained.
Then, following the proof framework in Section 4, we can show that the sequence \( \{u^k\} \) generated by Algorithm 2 is Fejér monotone with respect to the solution set of (4).

**Lemma 6.2.** Let the sequence \( \{u := (x^k, y^k)\} \) be generated by Algorithm 2. Then,
\[
\theta(u) - \theta(u^{k+1}) + \langle u - u^{k+1}, G(u) \rangle + \langle u - u^{k+1}, M'(u^{k+1} - u^k) \rangle \\
\geq \frac{\mu}{4} \|x^{k+1} - x\|^2 - \frac{(1 - \tau)^2}{\mu} \|A^T (y^k - y^{k+1})\|^2, \quad \forall u \in \mathcal{U},
\]
where \( M' \) is given in (37).

**Lemma 6.3.** Let the sequence \( \{u := (x^k, y^k)\} \) be generated by Algorithm 2. Then,
\[
\theta(u) - \theta(u^{k+1}) + \langle u - u^{k+1}, G(u) \rangle + \frac{1}{2} \|u - u^k\|^2_{M'} - \frac{1}{2} \|u - u^{k+1}\|^2_{M'}, \\
\geq \frac{1}{2} \|u^k - u^{k+1}\|^2_{M'} + \frac{\mu}{4} \|x^{k+1} - x\|^2, \quad \forall u \in \mathcal{U},
\]
where \( M' \) and \( H' \) are given by (37) and (38) respectively, and \( G \) is defined in (4b).

**Theorem 6.4.** Let \( u^* \) be an arbitrary solution of (36). Then, the sequence \( \{u^k\} \) generated by Algorithm 2 satisfies
\[
\|u^{k+1} - u^*\|^2_{M'} \leq \|u^k - u^*\|^2_{M'} - \|u^k - u^{k+1}\|^2_{M'} - \frac{\mu}{2} \|x^{k+1} - x^*\|^2. \tag{46}
\]

### 6.2. Linear convergence under scenario 1

In this section, we show the linear convergence rate of Algorithm 2 under the following assumption.

**Assumption 3.** In problem (36), \( B \) is symmetric and positive definite matrix. Then, \( \theta_1 \) is strongly convex with modulus \( \mu = \nu \lambda_{\text{min}}(B) > 0 \) and \( \nabla \theta_1 \) is Lipschitz continuous with constant \( L = \nu \|B\| \). Assume that \( \mathcal{X} = \mathbb{R}^n \), \( A \) is full row rank, and \( \theta_2 \) is convex. For Algorithm 2, \( r - \nu \|B\| > |r| \|A^T A\| \) and \( s > |r| + 2\mu^{-1}(1 - \tau)^2 \|A^T A\| \).

Then, utilizing the analytical tool in Section 5, we can show that the sequence \( \{u^k\} \) generated by Algorithm 2 converges linearly to the solution set.

**Lemma 6.5.** Suppose Assumption 3 holds. Let \( u^* \) be a solution of (36). Then, the sequence \( \{u^k\} \) generated by Algorithm 2 satisfies
\[
\|y^{k+1} - y^*\|^2 \leq 3\kappa \left[ L^2 \|x^{k+1} - x^*\|^2 + \|A^T A\| \|y^{k+1} - y^k\|^2 + r_2 \|x^{k+1} - x^k\|^2 \right], \tag{47}
\]
where \( \kappa \) is defined as (24) and \( r_2 = \|r I_n - \nu B\| \).

**Proof.** It follows from the full row rank of \( A \) that
\[
\|y^{k+1} - y^*\|^2 \leq \kappa \|A^T (y^{k+1} - y^*\|^2. \tag{49}
\]
Since \( \mathcal{X} = \mathbb{R}^n \), the first order optimality condition of the sub-problem (39a) can be written as
\[
\nabla \theta_1(x^k) - A^T y^k + r(x^{k+1} - x^k) = 0.
\]
Since \( (x^*, y^*) \) is a solution, we have
\[
\nabla \theta_1(x^*) = A^T y^*.
\]
Then, combining the above two equations leads to
\[
\| A^T y^{k+1} - A^T y^* \|^2 \\
= \| \nabla \theta_1(x^k) + A^T (y^{k+1} - y^k) + r(x^{k+1} - x^k) - \nabla \theta_1(x^*) \|^2 \\
= \| \nu B(x^k - x^{k+1}) + A^T (y^{k+1} - y^k) + r(x^{k+1} - x^k) + \nu B(x^{k+1} - x^*) \|^2 \\
\leq 3 \| \nu B(x^{k+1} - x^*) \|^2 + 3 \| A^T (y^{k+1} - y^k) \|^2 + 3 \| (rI_n - \nu B)(x^{k+1} - x^k) \|^2 \\
\leq 3 \left[ L^2 \| x^{k+1} - x^* \|^2 + \| A^T A \| \| y^{k+1} - y^k \|^2 + r_1^2 \| x^{k+1} - x^k \|^2 \right]. \tag{50}
\]
where \( r_1 = \| rI_n - \nu B \| \) and the first inequality is derived by (27). Then, combining the inequalities (49) and (50) yields (47).

Theorem 6.6. Suppose Assumption 3 holds. Let \( u^* \) be a solution of (36). Then, the sequence \( \{u^k\} \) generated by Algorithm 2 converges \( Q \)-linearly, i.e.,
\[
\| u^{k+1} - u^* \|^2_{M'} \leq \frac{1}{1 + \delta} \| u^k - u^* \|^2_{M'},
\]
where \( \delta := \min\{\delta_4, \delta_5, \delta_6\} > 0 \) and
\[
\begin{align*}
\delta_4 & := \frac{2\|M'\| \|1 + 3\kappa L^2\|}{\mu r - \|B\| - |r| \|A^T A\|}, \\
\delta_5 & := \frac{\kappa r_1^2 \|M'\|}{3s^2 \|A^T A\| \|M'\|}, \\
\delta_6 & := \frac{\kappa r_1^2 \|M'\|}{3s^2 \|A^T A\| \|M'\|}.
\end{align*}
\]
where \( \kappa \) and \( r_1 \) is defined in (24) and (48), respectively.

Proof. Using the definition of the matrix \( M' \), we have
\[
\delta \| u^{k+1} - u^* \|^2_{M'} \\
\leq \delta \| M' \| \left( \| x^{k+1} - x^* \|^2 + \| y^{k+1} - x^* \|^2 \right) \\
\leq \delta \| M' \| \left[ (1 + 3\kappa L^2) \| x^{k+1} - x^* \|^2 + 3\kappa \| A^T A \| \| y^{k+1} - y^k \|^2 \\
+ 3\kappa r_1^2 \| x^{k+1} - x^k \|^2 \right]
\]
where the last inequality is derived from (47). On the other hand,
\[
\| u^{k+1} - u^k \|^2_{M'} \\
= (rI_n - \nu B) \| x^{k+1} - x^k \|^2 + [s - 2\mu^{-1}(1 - \tau)^2 \| A^T A \|] \| y^{k+1} - y^k \|^2 \\
+ 2\tau \langle A(x^{k+1} - x^k), y^{k+1} - y^k \rangle \\
\geq (r - \nu \| B \| - |r| \| A^T A \|) \| x^{k+1} - x^k \|^2 \\
+ (s - 2\mu^{-1}(1 - \tau)^2 \| A^T A \| - |r|) \| y^{k+1} - y^k \|^2,
\]
where the inequality is derived from the Young’s inequality. Then, combining the above two inequalities with (46) leads to
\[
(1 + \delta) \left\| \mathbf{u}^{k+1} - \mathbf{u}^* \right\|^2_{M'} \\
\leq \left\| \mathbf{u}^k - \mathbf{u}^* \right\|^2_{M'} - \left\| \mathbf{u}^k - \mathbf{u}^{k+1} \right\|_{H^0}^2 - \frac{\mu}{2} \left\| x^{k+1} - x^* \right\|^2 + \delta \left\| \mathbf{u}^{k+1} - \mathbf{u}^* \right\|^2_{M'}
\]
\[
\leq \left\| \mathbf{u}^k - \mathbf{u}^* \right\|^2_{M'} - \left( \frac{\mu}{2} - \left\| M' \right\|(1 + 3\kappa L^2) \delta \right) \left\| x^{k+1} - x^* \right\|^2
\]
\[
- (r - \nu) |B| - |\tau| \left\| \mathbf{A}^T \mathbf{A} \right\| - 3\kappa \nu^2 \left\| M' \right\| \left\| \delta \right\| \left\| x^{k+1} - x^* \right\|^2
\]
\[
- (s - |\tau| - 2\mu^{-1}(1 - \tau^2) \left\| \mathbf{A}^T \mathbf{A} \right\| - 3\kappa \left\| \mathbf{A}^T \mathbf{A} \right\| \left\| M' \right\| \left\| \delta \right\| \left\| y^{k+1} - y^k \right\|^2.
\]

It follows from the definition of \( \delta \) and the above inequality that
\[
(1 + \delta) \left\| \mathbf{u}^{k+1} - \mathbf{u}^* \right\|^2_{M'} \leq \left\| \mathbf{u}^k - \mathbf{u}^* \right\|^2_{M'},
\]
and the assertion of the theorem is proved. \( \square \)

6.3. Linear convergence under scenario 2. In this section, we prove the global linear convergence of Algorithm 2 under some error bound conditions.

Assumption 4. In problem (36), \( \theta_2 \) is convex, \( \theta_1 \) is strongly convex with modulus \( \mu = \nu \lambda_{\text{min}}(B) > 0 \). For Algorithm 2, \( (r - \nu |B|)(s - 2\mu^{-1}(1 - \tau^2) \left\| \mathbf{A}^T \mathbf{A} \right\|) > \tau^2 \left\| \mathbf{A}^T \mathbf{A} \right\| \). Furthermore, assume that for any \( \omega > 0 \), there exists \( \eta > 0 \), such that
\[
\text{dist}(\mathbf{u}, \mathcal{U}^*) \leq \eta \text{dist}(0, \mathcal{E}(\mathbf{u}, 1)), \quad \forall \left\| \mathbf{u} \right\| \leq \omega, \quad \mathbf{u} \in \mathcal{U},
\]
where \( \mathcal{E} \) is defined by (6).

Lemma 6.7. Let the sequence \( \{ \mathbf{u}^k := (x^k, y^k) \} \) be generated by Algorithm 2. Then, there exists a constant \( \sigma_2 > 0 \) such that
\[
\text{dist}^2(0, \mathcal{E}(\mathbf{u}^{k+1}, 1)) \leq \sigma_2 \{ \left\| y^k - y^{k+1} \right\|^2 + \left\| x^k - x^{k+1} \right\|^2 \},
\]
where
\[
\sigma_2 = \max \{ 2(\left\| \mathbf{A}^T \mathbf{A} \right\| + s^2), 2(r_1^2 + \tau^2 \left\| \mathbf{A}^T \mathbf{A} \right\|) \} > 0,
\]
where \( r_1 \) is defined as (48).

Proof. The optimality condition of \( x \)-subproblem (39a) indicates that
\[
x^{k+1} = \mathbf{P}_\mathcal{X} [x^{k+1} - (\nu B(x^k - b) - \mathbf{A}^T y^k + r(x^{k+1} - x^k))].
\]

Then,
\[
\text{dist}^2(0, \mathcal{E}_\mathcal{X}(\mathbf{u}^{k+1}, 1))
\]
\[
= \text{dist}^2 (x^{k+1}, \mathbf{P}_\mathcal{X}(x^{k+1} - (\nu B(x^{k+1} - b) - \mathbf{A}^T y^{k+1})))
\]
\[
= \left\| \mathbf{P}_\mathcal{X} [x^{k+1} - (\nu B(x^k - b) - \mathbf{A}^T y^k + r(x^{k+1} - x^k))]
\]
\[
- \mathbf{P}_\mathcal{X} [x^{k+1} - (\nu B(x^{k+1} - b) - \mathbf{A}^T y^{k+1})] \right\|^2
\]
\[
\leq \left\| \mathbf{A}^T (y^k - y^{k+1}) + (r I_n - \nu B)(x^k - x^{k+1}) \right\|^2
\]
\[
\leq 2 \left\| \mathbf{A}^T \right\| \left\| y^k - y^{k+1} \right\|^2 + 2r_1^2 \left\| x^k - x^{k+1} \right\|^2,
\]
where the second inequality follows from the inequality
\[
\left\| a + b \right\|^2 \leq 2a^2 + 2b^2, \quad \forall a, b \in \mathbb{R}^n.
\]
by setting \(a = A^T(y^k - y^{k+1})\) and \(b = \|rI_n - \nu B\|(x^k - x^{k+1})\). On the other hand, the optimality condition of the \(y\)-subproblem (39c) implies that
\[
y^{k+1} = P_y[y^{k+1} - (s_y^{k+1} + A x^{k+1} + \tau A(x^{k+1} - x^k) + s(y^{k+1} - y^k))].
\]
Hence, we obtain
\[
\text{dist}^2(0, E_y(u^{k+1}, 1)) = \|P_y[y^{k+1} - (s_y^{k+1} + A^T x^{k+1})]\|
\leq \|P_y[y^{k+1} - (s_y^{k+1} + A^T x^{k+1} + \tau A(x^{k+1} - x^k) + s(y^{k+1} - y^k))]
- P_y[y^{k+1} - (s_y^{k+1} + A^T x^{k+1})]\|
\leq \|\tau A(x^{k+1} - x^k) + s(y^{k+1} - y^k)\|^2
\leq 2\rho^2\|A^T A\|\|x^{k+1} - x^k\|^2 + 2s^2\|y^k - y^{k+1}\|^2.
\]
(55)

Then, combining (54) and (55) yields
\[
\text{dist}^2(0, E_x(u^{k+1}, 1)) = \text{dist}^2(0, E_x(u^{k+1}, 1)) + \text{dist}^2(0, E_y(u^{k+1}, 1))
\leq 2(\|A^T A\| + s^2\|y^k - y^{k+1}\|^2 + 2\rho^2\|A^T A\| + r^2)\|x^k - x^{k+1}\|^2.
\]
The assertion of this lemma is derived from the definition of \(\sigma_2\). \(\square\)

Next, we show the global linear convergence of Algorithm 2 under Assumption 4.

**Theorem 6.8.** Suppose Assumption 4 holds. Let \(\{u^k := (x^k, y^k)\}\) be generated by Algorithm 2. Then,
\[
(1 + \rho)\text{dist}^2_{M'}(u^{k+1}, u^*) \leq \text{dist}^2_{M'}(u^k, u^*),
\]
where
\[
\rho := \frac{1}{\eta^2\sigma\|M'\|},
\]
and \(\sigma = \sigma_2/\lambda_{\text{min}}(H')\), \(\sigma_2\) is defined in (53).

**Proof.** Let \((x^*, y^*)\) be a saddle point in \(U^*\). Based on Theorem 6.4, the sequence \(\{x^k, y^k\}\) converges to a saddle point in \(U^*\), which is bounded. Assumption 4, together with (52), implies that
\[
\text{dist}^2(u^{k+1}, u^*) \leq \eta^2\sigma_2\{\|y^k - y^{k+1}\|^2 + \|x^k - x^{k+1}\|^2\}.
\]
On the other hand,
\[
\lambda_{\text{min}}(H')\{\|y^k - y^{k+1}\|^2 + \|x^k - x^{k+1}\|^2\} \leq \|u^k - u^{k+1}\|_{H'}^2.
\]
Then, setting \(\sigma = \sigma_2/\lambda_{\text{min}}(H')\), we conclude
\[
\text{dist}^2(u^{k+1}, u^*) \leq \eta^2\sigma\|u^k - u^{k+1}\|_{H'}^2.
\]
Note that
\[
\text{dist}^2_{M'}(u^k, u^*) \leq \|M'\|\text{dist}^2(u^k, u^*),
\]
and hence
\[
\text{dist}^2_{M'}(u^{k+1}, u^*) \leq \eta^2\sigma\|M'\|\|u^k - u^{k+1}\|_{H'}^2.
\]
(56)
Then, it follows from Theorem 6.4 and (56) that
\[ \|u^{k+1} - u^*\|_M^2 \leq \|u^k - u^*\|_M^2 - \frac{1}{\eta^2 \sigma \|M^\prime\|} \text{dist}^2_{M^\prime}(u^{k+1}, U^*) \] (57)

Let \( u^* \in U^* \) such that \( \text{dist}_{M^\prime}(u^k, U^*) = \|u^k - u^*\|_{M^\prime} \). Then it follows from (57) that
\[ \text{dist}^2_{M^\prime}(u^{k+1}, U^*) \leq \|u^k - u^*\|_M^2 - \frac{1}{\eta^2 \sigma \|M^\prime\|} \text{dist}^2_{M^\prime}(u^{k+1}, U^*) = \text{dist}^2_{M^\prime}(u^k, U^*) - \frac{1}{\eta^2 \sigma \|M^\prime\|} \text{dist}^2_{M^\prime}(u^{k+1}, U^*). \]
Rearranging terms, we get the desired inequality immediately. \( \square \)

Hence, the above theorem demonstrates the global linear convergence of Algorithm 2.

7. Numerical Test. In this section, we turn our attention to studying the numerical behavior of Algorithm 1. The tested algorithm was coded by MATLAB R2016a, and our experiments were implemented on a Lenovo laptop with Windows 10 system and Inter(R) Core(TM) i7-7500 (2.70GH) CPU processor with a 16GB memory.

7.1. Example 1. Firstly, we consider solving image processing problem by Algorithm 1. The quality of restored images is measured by the value of signal-to-noise (SNR), which is defined as
\[ \text{SNR} := 20 \log \frac{\|x^*\|}{\|x^k - x^*\|} \]
where \( x^k \) is the restored image by certain algorithm and \( x^* \) represents the original one.

Then, we apply Algorithm 1 to solve the primal-dual formulation of the ROF model in [3], which is given as
\[ \min_{x \in X} \max_{y \in Y} y^\top Ax + \frac{\lambda}{2} \|x - g\|^2 \] (58)
where \( x, g \in X \) are the unknown solution and the given noisy data, \( A \) is the matrix representation of the discrete divergence operator, \( \lambda > 0 \) is a parameter balancing the data-fidelity and TV regularization terms, and \( Y \) is the Cartesian product of some unit balls. It was noticed that \( \theta_1(x) \) is strongly convexity with modulus \( \lambda \) and this model can be viewed as a special case of (1).

As shown in Figure 1, the grayscale images (house.png (256 × 256), and lena.png (512 × 512)) were tested in our experiments. These images are degraded by the zero-mean Gaussian noise with the standard deviation 0.1, which are also displayed in Figure 1. To recover these corrupted images, we take \( \lambda = 8 \) for this model (58). In our experiment, the tested algorithm adopts the stopping criterion
\[ \text{Rer} := \frac{\|x^{k+1} - x^k\|}{\|x^{k+1}\|} < \text{Tol}, \] (59)
Table 1. Parameters Setting

| Parameters    | Values         |
|---------------|----------------|
| \((1/r, 1/s, \tau)\) | \((0.05, 3, -0.2)\) |
| \((1/r, 1/s, \tau)\) | \((0.01, 15, 0)\) |
| \((1/r, 1/s, \tau)\) | \((0.012, 12.5, 0.4)\) |
| \((1/r, 1/s, \tau)\) | \((0.012, 12.5, 0.8)\) |
| \((1/r, 1/s, \tau)\) | \((0.01, 12.5, 1)\) |
| \((1/r, 1/s, \tau)\) | \((0.012, 12.5, 1.2)\) |

where \(\{x^k\}\) is the sequence generated by Algorithm 1 and Tol is the error tolerance and set as \(10^{-5}\). The iterations is set as 150. In order to study the sensitivity of the parameter \(\tau\) of Algorithm 1, we tested six sets of the involved parameters in the following Table 1. It is noticed that the different \(\tau\) are corresponding to various sets of proximal parameters. This is because that we need to choose a variety of proximal parameters to ensure that Algorithm 1 is globally convergent. For the case of \(\tau = 1\), this method chooses the proximal parameter by the strategy in [3]. For the other cases, we tune the proximal parameters by the inequalities in Remark 3 and numerical simulation.

In Figure 2, we plotted the evolution curves of SNRs and Relative error (log(Rer)) with respect to iterations, which were obtained by solving the test instances. From the results in Figure 2, we can see that Algorithm 1 can quickly achieve higher SNR values in most of cases except that \(\tau = 0\) when recovering these two images. Furthermore, Algorithm 1 with \(\tau = 1.2\) can reach the best relative error values when comparing with the other cases, and Algorithm 1 with \(\tau = 0.8\) can also obtain very similar satisfactory results. Therefore, it demonstrates that Algorithm 1 for the
cases of $\tau \neq 1$ is feasible and efficient when dealing with some numerical problems, and is also sensitive about the values of $\tau$ and the proximal parameters. Moreover, the evolution curves in Figure 2 further support the linear convergence behaviors of Algorithm 1.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Evolution of SNRs, and Relative error (Rer) defined by (59) with respect to iterations. Left: for House. Right: for Lena.}
\end{figure}

### 7.2. Example 2
This subsection further investigates the sensitivity of $\tau$ in Algorithm 1 when dealing with the nearest correlation matrix problem (see [7, 15]). This problem is derived from the area of statistics and defined as follows:

\[
\min_X \left\{ \frac{1}{2} \| X - C \|_F^2 \mid \text{diag}(X) = 1, \ X \in S^n_+ \right\},
\]

where $\| \cdot \|_F$ denotes the standard Forbenius-norm for matrices; $C$ is a given $n \times n$ symmetric matrix; $1 = (1, \ldots, 1)^T$ is an $n$-dimensional vector; $S^n_+$ is a positive semi-definite cone. By introducing a Lagrangian multiplier $y \in \mathbb{R}^n$, the (60) can be transformed into a saddle point problem

\[
\min_X \max_{y \in \mathbb{R}^n} \left\{ \frac{1}{2} \| X - C \|_F^2 - \langle y, \text{diag}(X) - 1 \rangle \right\},
\]

which can be viewed as a special case of (1). Then, we utilize Algorithm 1 to solve the above matrix optimization problem and the tolerance of the stopping criterion is set as $10^{-5}$. Here, we consider five scenarios of the parameters: $(\tau, r, s) \in \{(1, 2, 1.01/r), (0.2, 6, 0.4/r), (0.5, 2, 0.4/r), (0.8, 2, 0.4/r), (1.2, 2, 0.4/r)\}$, and plot the bar charts with respect to iterations and computing time in Figure 3. The plots show that Algorithm 1 still performs stable for different dimensions. Comparatively,
Algorithm 1 with $\tau = 1.2$ performs better than the others, and Algorithm 1 with $\tau \neq 1$ can achieve more better results than $\tau = 1$. On the other hand, it is noteworthy that our algorithm still works effectively for the parameters beyond theoretical guarantees, which implies that the derived bound of parameters is not necessarily tight.

![Graph](image)

Figure 3. Sensitivity analysis on parameters $\tau$ of Algorithm 1 for matrix correlation problems.

8. Conclusion. In this paper, we studied the general primal-dual algorithm for the saddle point problem where one of the underlying function is strongly convex. Under suitable conditions on the step sizes, we proved the global convergence. We then considered its linear rate of convergence under further assumptions. We completed the task for two cases, the first one is that the strongly convex function is smooth with Lipschitz continuous gradient, and the second one is that the problem possesses some error bound. Additionally, we also established the global linear rate of the linearized primal-dual algorithm under these two different strategies.

REFERENCES

[1] K. Arrow, L. Hurwicz and H. Uzawa, Studies in linear and non-linear programming, in Stanford Mathematical Studies in the Social Sciences, II, Vol. 2, Stanford University Press, Stanford, Calif., 1958.

[2] X. Cai, D. Han and L. Xu, An improved first-order primal-dual algorithm with a new correction step, J. Global Optim., 57 (2013), 1419–1428.

[3] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, J. Math. Imaging Vis., 40 (2011), 120–145.

[4] A. Chambolle and T. Pock, On the ergodic convergence rates of a first-order primal-dual algorithm, Math. Program. Ser. A, 159 (2016), 253–287.

[5] Y. Chen, G. Lan and Y. Ouyang, Optimal primal-dual methods for a class of saddle point problems, SIAM J. Optim., 24 (2014), 1779–1814.

[6] E. Esser, X. Zhang and T. Chan, A general framework for a class of first-order primal-dual algorithms for convex optimization in imaging science, SIAM J. Imaging Sci., 3 (2010), 1015–1046.

[7] Y. Gao and D. Sun, Calibrating least squares semidefinite programming with equality and inequality constraints, SIAM J. Matrix Anal. Appl., 31 (2009), 1432–1457.

[8] G. Gu, B. He and X. Yuan, Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: A unified approach, Comput. Optim. Appl., 59 (2014), 135–161.

[9] D. Han, D. Sun and L. Zhang, Linear rate convergence of the alternating direction method of multipliers for convex composite programming, Math. Oper. Res., 43 (2018), 622–637.
[10] D. Han, W. Xu and H. Yang, An operator splitting method for variational inequalities with partially unknown mappings, *Numer. Math.*, **111** (2008), 207–237.

[11] D. Han and X. Yuan, Local linear convergence of the alternating direction method of multipliers for quadratic programs, *SIAM J. Numer. Anal.*, **51** (2013), 3446–3457.

[12] B. He, F. Ma and X. Yuan, An algorithmic framework of generalized primal-dual hybrid gradient methods for saddle point problems, *J. Math. Imaging Vision*, **58** (2017), 279–293.

[13] B. He, Y. You and X. Yuan, On the convergence of primal-dual hybrid gradient algorithm, *SIAM J. Imaging Sci.*, **7** (2014), 2526–2537.

[14] B. He and X. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, *SIAM J. Imaging Sci.*, **5** (2012), 119–149.

[15] B. He, M. Xu and X. Yuan, Solving large-scale least squares semidefinite programming by alternating direction methods, *SIAM J. Matrix Anal. Appl.*, **32** (2011), 136–152.

[16] H. He, J. Desai and K. Wang, A primal-dual prediction-correction algorithm for saddle point optimization, *J. Global Optim.*, **66** (2016), 573–583.

[17] F. Jiang, X. Cai, Z. Wu and D. Han, Approximate first-order primal-dual algorithms for saddle point problems, *Math. Comput.*, **90** (2021), 1227–1262.

[18] Y. Malitsky and T. Pock, A first-order primal-dual algorithm with linesearch, *SIAM J. Optim.*, **28** (2018), 411–432.

[19] A. Nemirovski, Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operator and smooth convex-concave saddle point problems, *SIAM J. Optim.*, **15** (2004), 229–251.

[20] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Applied Optimization, Kluwer Academic Publishers, Boston, MA, 2004.

[21] J. Rasch and A. Chambolle, Inexact first-order primal-dual algorithms, *Comput. Optim. Appl.*, **76** (2020), 381–430.

[22] W. Tian and X. Yuan, Linearized primal-dual methods for linear inverse problems with total variation regularization and finite element discretization, *Inverse Problems*, **32** (2016), 115011.

[23] T. Valkonen, Preconditioned proximal point methods and notions of partial subregularity, *J. Convex Anal.*, **28** (2021), 251–278.

[24] K. Wang and H. He, A double extrapolation primal-dual algorithm for saddle point problems, *J. Sci. Comput.*, **85** (2020), 1–30.

[25] W. Yang and D. Han, Linear convergence of the alternating direction method of multipliers for a class of convex optimization problems, *SIAM J. Numer. Anal.*, **54** (2016), 625–640.

[26] X. Zheng and K. Ng, Metric subregularity of piecewise linear multifunctions and applications to piecewise linear multiobjective optimization, *SIAM J. Optim.*, **24** (2014), 154–174.

[27] M. Zhu and T. Chan, An efficient primal-dual hybrid gradient algorithm for total variation image restoration, *CAM Reports*, UCLA, Los Angeles, CA, 2008, 08–34.

Received March 2021; revised May 2021; early access August 2021.

*E-mail address: wangkaihawk@njust.edu.cn*

*E-mail address: handr@buaa.edu.cn*