Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 19, No. 4, pp. 551–567, 2023

A NOTE ON THE LINEAR CONVERGENCE OF GENERALIZED PRIMAL-DUAL HYBRID GRADIENT METHODS*

Kai Wang and Qun Wang†

Abstract: In this paper, we revisit a class of primal-dual algorithms proposed in [He et al., An Algorithmic Framework of Generalized Primal-Dual Hybrid Gradient Methods for Saddle Point Problems, J. Math. Imaging Vis., 58 (2017) 279-293], and focus on investigating the global linear convergence rate of these approaches under two scenarios. One scenario is assuming that one of the objectives is strongly convex and its gradient is Lipschitz continuous, and the other one is the hypothesis of some error bound conditions. Furthermore, some theoretical results are verified by numerical simulation.

Key words: *saddle point problem, primal-dual hybrid gradient algorithm, predictor-corrector, linear convergence rate, error bound*

Mathematics Subject Classification: *90C47, 65K15, 41A25, 47J20*

1 Introduction

The saddle point problem has attracted a lot of attentions due to its wide applications in practice. In the fields of constrained minimization problems, zero-sum games, statistical learning and image processing (see [4, 9, 10, 8, 11, 16, 22, 30]), a variety of application problems can be formulated into saddle point problems. In this paper, we consider the following saddle point problem

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}} \Phi(x,y) := \theta_1(x) - \langle y, Ax\rangle - \theta_2(y), \tag{1.1}$$

where $\theta_1(\cdot):\mathbb{R}^n\to\mathbb{R}$ and $\theta_2(\cdot):\mathbb{R}^m\to\mathbb{R}$ are convex but not necessarily smooth functions; $A\in\mathbb{R}^{m\times n}$ 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; $\langle\cdot,\cdot\rangle$ denotes the standard inner product of vectors; and $\|\cdot\|$ is the Euclidean norm. The solution set of (1.1) is assumed to be nonempty throughout our paper.

The primal-dual algorithms can be considered as benchmark solvers for solving the saddle point problem (1.1). One of the most representative primal-dual algorithms was proposed

*The research of the first author was supported by National Natural Science Foundation of China (NSFC) at Grant No. 11901294 and Natural Science Foundation of Jiangsu Province at Grant No. BK20190429. The research of the second author was supported by National Natural Science Foundation of China (NFSC) at Grant No. 11801502 and First class Discipline B(Zhejiang University of Finance and Economics-Mathematics).

†Corresponding author.

by Chambolle and Pock in [4], which is named FOPDA, whose iterative scheme is described as

$$\begin{cases} x^{k+1} = \arg\min_{x\in\mathcal{X}} \left\{ \Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \right\}, & \text{(1.2a)} \\ \bar{x}^k = x^{k+1} + \tau(x^{k+1} - x^k), & \text{(1.2b)} \\ y^{k+1} = \arg\max_{y\in\mathcal{Y}} \left\{ \Phi(\bar{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \right\}, & \text{(1.2c)} \end{cases}$$

where τ is a relaxation (extrapolation) parameter, and $r > 0$ and $s > 0$ are proximal parameters. In recent works [4, 16], we have witnessed many successful applications of this method in the areas of the image processing problems. More importantly, a series of Primal-Dual Hybrid Gradient (PDHG) based methods [10, 15, 32] can be viewed as special cases of the FOPDA method with different choices of τ, which are originated from the Arrow-Hurwicz-Uzawa method [1].

Considering the case of $\tau = 0$, the FOPDA method turns into PDHG. By implementing some restriction on the step sizes, the convergence of this method was proved in [10]. However, a counter example in [15] shows that PDHG may be divergent when the step sizes are fixed as tiny constants. [15] also analyzed the global convergence of this method by assuming that one of the objective functions is strongly convex. In the case of $\tau = 1$, the global convergence and the worst case $O(1/k)$ convergence rate of FOPDA was proved in [4] under the condition that $rs > \|A^\top A\|$. Based on this algorithm framework, [4] proposed two accelerated primal-dual algorithms and analyzed the $O(1/k^2)$ sub-linear and linear convergence, respectively. Then, [28] proposed a linearized primal-dual algorithm for minimizing the sum of three convex functions. By adopting some approximation rules or line-search technique, some inexact primal-dual algorithms were proposed for problem (1.1) in [18, 19, 20, 21, 24].

Considering the involved parameter $\tau \neq 1$, it is noteworthy that FOPDA could perform better from the viewpoint of computing [2, 16]. However, the convergence is not easily guaranteed in this circumstance. In order to ensure the global convergence, a series of predictor-corrector primal-dual algorithms were presented in [2, 16], and demonstrated the computational efficiency and flexibility in numerical simulations. By adopting some proximal regularization and linearized techniques, [17, 25] proposed some linearized primal-dual approaches. Furthermore, [5] presented some generalizations of the primal-dual methods, which covers non-linear proximal regularization and inertial variants as special cases. Recently, [26] established the global linear convergence rate of FOPDA for more general case of $\tau \in \mathbb{R}$ under the strongly convex assumption or some error bound conditions. Furthermore, [27] proposed a double extrapolation primal-dual algorithm for saddle point problem.

In [14], a generalized algorithmic framework of Primal-Dual Hybrid Gradient (GPDHG) was put forward for problem (1.1), which updates the iterative points in predictor-corrector fashion:

$$\begin{cases} \tilde{x}^k = \arg\min_{x\in\mathcal{X}} \left\{ \Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \right\}, & \text{(1.3a)} \\ \bar{x}^k = \tilde{x}^k + \tau(\tilde{x}^k - x^k), & \text{(1.3b)} \\ \tilde{y}^k = \arg\max_{y\in\mathcal{Y}} \left\{ \Phi(\bar{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \right\}; & \text{(1.3c)} \\ x^{k+1} = x^k - \alpha(x^k - \tilde{x}^k), & \text{(1.3d)} \\ y^{k+1} = y^k - \beta(y^k - \tilde{y}^k), & \text{(1.3e)} \end{cases}$$

and utilizes the constant step lengths α and β in the corrector step. This framework can be transformed into diversified specific PDHG-like algorithms by setting different parameters.

As we can see, this algorithm turns into the FOPDA in [4] in the case of $\alpha = \beta = 1$. If $\tau = 1$ and $\alpha = \beta \in (0, 2)$, the approach (1.3) be equivalent to the fourth algorithm in [16]. Additionally, a completely symmetric PDHG scheme with the golden-ratio step size can be yielded by setting $\alpha = 1$ and $\beta = 1/\tau$. Under ergodic and non-ergodic conditions, the proof frameworks of convergence and sub-linear convergence rate were also presented in [14]. Based on this work, Chang and Yang presented a golden ratio primal-dual algorithm and its accelerated version for structured convex programs in [6].

Although the convergence and sub-linear convergence of GPDHG are established in [14], the analysis of linear convergence results is still missing. Hence, the main contribution of this paper is analyzing the linear convergence rate of the GPDHG under two scenarios. Firstly, we focus on the case that one of the objectives is strongly convex and its gradient is Lipschitz continuous. Secondly, we also establish the linear convergence rate under the error bound condition.

The remainder of this paper is organized as follows. In Section 2, we summarize some basic notations and definitions, and present an equivalent variational inequality characterization of the saddle point problem (1.1). Section 3 presents the algorithmic framework of GPDHG, and Section 4 proves the global linear convergence rate. In section 5, we show some computational results. Finally, Section 6 makes the conclusions.

2 Preliminaries

In this section, we present some notations and definitions for facilitating the following convergence analysis.

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, and $\|x\|$ be the Euclidean norm. Furthermore, the matrix norm of an arbitrary matrix B be denoted by $\|B\|$,

$$\|B\| := \sup_{x\neq 0}\left\{\frac{\|Bx\|}{\|x\|}\right\}.$$

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 μ-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 $\mathbf{dom}\ \theta$. Then, the subdifferential of $\theta(\cdot)$ at a point $v \in \mathbf{dom}\ \theta$ is defined by

$$\partial\theta(v) = \left\{\ \xi \ \middle|\ \theta(u) \geq \theta(v) + \langle \xi,\ u - v\rangle\ ,\ \ \forall u \in \mathbf{dom}\ \theta\right\},$$

and the vector ξ is said to be a subgradient of $\theta(\cdot)$ at v.

Accordingly, if ξ represents the subgradient of a μ-strongly convex function $\theta(\cdot)$ at a point $v \in \mathbf{dom}\ \theta$, from [23] it follows that

$$\theta(u) \geq \theta(v) + \langle \xi,\ u - v\rangle + \frac{\mu}{2}\|u - v\|^2,\ \ \forall u \in \mathbf{dom}\ \theta.$$

Definition 2.3. An operator f: $\Omega \to \mathbb{R}^n$ is said to be Lipschitz continuous on Ω if there exists a constant $L_f > 0$ such that

$$\|f(x_1) - f(x_2)\| \leq L_f \|x_1 - x_2\|, \forall x_1, x_2 \in \Omega. \tag{2.1}$$

Definition 2.4. A function $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is multi-function if $F(x)$ is a set in $\mathbb{R}^n$. We say that F is linear if $Gr(F) := \{(x, y)|y \in F(x)\}$ is a polyhedron. If $Gr(F)$ is the union of finitely many polyhedra, F is said to be a piecewise linear multi-function.

Now, let (x^*, y^*) be a solution of the saddle point problem (1.1). Then, from saddle point optimality conditions, we get that

$$\Phi(x^*, y) \leq \Phi(x^*, y^*) \leq \Phi(x, y^*), \ \forall x \in \mathcal{X}, \ \forall y \in \mathcal{Y},$$

which reduces to the following mixed variational inequalities (**MVI**):

$$\begin{cases} \theta_1(x) - \theta_1(x^*) + \langle x - x^*, -A^\top y^* \rangle \geq 0, & \forall x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(y^*) + \langle y - y^*, Ax^* \rangle \geq 0, & \forall y \in \mathcal{Y}. \end{cases} \tag{2.2}$$

The above variational inequality characterization can be compactly rewritten as a problem **MVI**$(\mathcal{U}, \theta, \mathcal{G})$: Finding $\boldsymbol{u}^* \in \mathcal{U}$, such that

$$\theta(\boldsymbol{u}) - \theta(\boldsymbol{u}^*) + \langle \boldsymbol{u} - \boldsymbol{u}^*, \mathcal{G}(\boldsymbol{u}^*) \rangle \geq 0, \quad \forall \boldsymbol{u} \in \mathcal{U}, \tag{2.3a}$$

where

$$\boldsymbol{u} := \begin{pmatrix} x \\ y \end{pmatrix}, \ \theta(\boldsymbol{u}) = \theta_1(x) + \theta_2(y), \ \mathcal{G}(\boldsymbol{u}) := \begin{pmatrix} -A^\top y \\ Ax \end{pmatrix}, \quad \text{and } \mathcal{U} := \mathcal{X} \times \mathcal{Y}. \tag{2.3b}$$

The underlying mapping $\mathcal{G}$ defined in (2.3b) is monotone because that

$$\langle \boldsymbol{u}_1 - \boldsymbol{u}_2, \ \mathcal{G}(\boldsymbol{u}_1) - \mathcal{G}(\boldsymbol{u}_2) \rangle = 0, \quad \forall \boldsymbol{u}_1, \ \boldsymbol{u}_2 \in \mathcal{U}.$$

As the solution set of (1.1) is assumed to be nonempty, the solution set of problem (2.3), 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 $P_\mathcal{C}$ be the projection operator from $\mathbb{R}^n$ onto $\mathcal{C}$,

$$P_\mathcal{C}(x) = \arg\min_{z \in \mathcal{C}} \|x - z\|.$$

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,

$$\|P_\mathcal{C}(x) - P_\mathcal{C}(z)\| \leq \|x - z\|, \ \forall x, \ z \in \mathbb{R}^n. \tag{2.4}$$

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 $\text{dist}_H(x, \mathcal{C}) = \min_{y \in \mathcal{C}}\{\|x - y\|_H\}$ denotes the distance in the sense of matrix norm, where H is a given symmetric and positive definite matrix.

The above variational inequalities (2.3) can be also equivalently transformed into some generalized projection equations.

Lemma 2.5. *The variational inequality problem* **MVI**$(\mathcal{U},\theta,\mathcal{G})$ *amounts to finding* $\boldsymbol{u}^*$*, such that* $0\in\mathcal{E}(\boldsymbol{u}^*,\gamma)$*, i.e.*

$$dist^2(0,\mathcal{E}(\boldsymbol{u}^*,\gamma))=0,$$

where the set-valued mapping $\mathcal{E}(\boldsymbol{u},\gamma)$ *is defined as*

$$\mathcal{E}(\boldsymbol{u},\gamma):=\begin{pmatrix}\mathcal{E}_{\mathcal{X}}(\boldsymbol{u},\gamma):=x-P_{\mathcal{X}}[x-\gamma(\xi_x-A^\top y)]\\ \mathcal{E}_{\mathcal{Y}}(\boldsymbol{u},\gamma):=y-P_{\mathcal{Y}}[y-\gamma(\zeta_y+Ax)]\end{pmatrix}, \tag{2.5}$$

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 inequality characterization (2.3) and the associated theory of variational inequalities. Since $\mathcal{U}^*$ denotes the solution set of **MVI**$(\mathcal{U},\theta,\mathcal{G})$, it follows that

$$\mathcal{U}^*=\{\boldsymbol{u}^*\mid \mathrm{dist}(0,\mathcal{E}(\boldsymbol{u}^*,\gamma))=0\}.$$

The following theorem is established in [31, Theorem 3.3] and plays a fundamental role in our linear rate of convergence analysis under error bound condition.

Theorem 2.6. *Let* F *be a piecewise linear multi-function. For any* $\omega>0$*, there exists* $\eta>0$ *such that*

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

3 Algorithm Framework

Define

$$M:=\begin{pmatrix}\alpha I_n & 0\\ 0 & \beta I_m\end{pmatrix}\quad\text{and}\quad Q:=\begin{pmatrix} rI_n & A^\top\\ \tau A & sI_m\end{pmatrix}, \tag{3.1}$$

where $\alpha>0$, $\beta>0$, $\tau\in(0,1]$, $rs>\tau\|A^\top A\|\ge 0$. Under these conditions, the matrices M and Q are positive definite. The positive definiteness of the matrices M and Q are crucial for the convergence analysis of primal-dual algorithms.

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

4 Global Convergence Analysis

4.1 Global convergence

In this section, we review the result in [14] that Algorithm 1 is globally convergent to a solution of (1.1) under the following assumption, which can be derived by [14, Theorem 4.1] and [14, Theorem 4.2].

Assumption 4.1. Assume that $\theta_1(x)$ in (1.1) is strongly convex function with modulus $\mu>0$, $\theta_2(x)$ is convex and $rs\ge\|A^\top A\|$. Define

$$H:=QM^{-1}=\begin{pmatrix}\frac{r}{\alpha}I_n & \frac{1}{\beta}A^\top\\ \frac{\tau}{\alpha}A & \frac{s}{\beta}I_m\end{pmatrix}, \tag{4.1}$$

and set $\beta=\frac{\alpha}{\tau}$. If we choose

$$\tau=1\quad\text{and}\quad\alpha\in(0,2)$$

Algorithm 1 Generalized Primal-Dual Hybrid Gradient Algorithm.

1: Select $r > 0$, $s > 0$, $\tau \in (0, 1]$ and $rs > \|A^\top A\|$.
2: **for** $k = 0, 1, 2, \cdots$ **do**
3: Generate the intermediate points $\tilde{\boldsymbol{u}}^k := (\tilde{x}^k, \tilde{y}^k)$ via

$$\begin{cases} \tilde{x}^k = \arg\min_{x\in\mathcal{X}} \left\{ \Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \right\}, & (3.2a) \\ \bar{x}^k = \tilde{x}^k + \tau(\tilde{x}^k - x^k), & (3.2b) \\ \tilde{y}^k = \arg\max_{y\in\mathcal{Y}} \left\{ \Phi(\bar{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \right\}. & (3.2c) \end{cases}$$

4: Generate the next points $\boldsymbol{u}^{k+1} := (x^{k+1}, y^{k+1})$ via

$$\begin{cases} x^{k+1} = x^k - \alpha(x^k - \tilde{x}^k), & (3.3a) \\ y^{k+1} = y^k - \beta(y^k - \tilde{y}^k), & (3.3b) \end{cases}$$

where

$$\begin{cases} \alpha = \beta \in (0, 2), \quad \text{if } \tau = 1; & (3.4a) \\ \alpha \in \left(0, (1+\tau) - \sqrt{1-\tau}\right], \beta = \frac{\alpha}{\tau}, \quad \text{if } \tau \in (0, 1]. & (3.4b) \end{cases}$$

5: **end for**

or

$$\tau \in (0, 1) \quad \text{and} \quad 0 < \alpha \leq (1+\tau) - \sqrt{1-\tau},$$

the matrices $H \succ 0$, $H = H^\top$,

$$G := Q^\top + Q - M^\top H M = \begin{pmatrix} (2-\alpha)rI_n & (1+\tau-\alpha)A^\top \\ (1+\tau-\alpha)A & (2-\frac{\alpha}{\tau})sI_m \end{pmatrix} \succ 0.$$

Lemma 4.1. *Suppose Assumption 4.1 holds. Let the sequence $\{\boldsymbol{u}^k := (x^k, y^k)\}$ be generated by Algorithm 1. Then,*

$$\begin{aligned} &\theta(\boldsymbol{u}) - \theta(\tilde{\boldsymbol{u}}^k) + \langle \boldsymbol{u} - \tilde{\boldsymbol{u}}^k, \mathcal{G}(\tilde{\boldsymbol{u}}^k) \rangle \\ &\geq \langle \boldsymbol{u} - \tilde{\boldsymbol{u}}^k, Q(\boldsymbol{u}^k - \tilde{\boldsymbol{u}}^k) \rangle + \frac{\mu}{2}\|\tilde{x}^k - x\|^2, \quad \forall \boldsymbol{u} \in \mathcal{U}, \end{aligned} \tag{4.2}$$

where $\mathcal{G}(\cdot)$ is given in (2.3b) *and Q is defined in* (3.1).

Proof. It follows from the first-order optimality conditions of (3.2a) and (3.2c) that

$$\langle x - \tilde{x}^k, \xi_{\tilde{x}^k} - A^\top y^k + r(\tilde{x}^k - x^k) \rangle \geq 0, \ \ \forall x \in \mathcal{X}, \tag{4.3}$$

and

$$\langle y - \tilde{y}^k, \zeta_{\tilde{y}^k} + A[(1+\tau)\tilde{x}^k - \tau x^k] + s(\tilde{y}^k - y^k) \rangle \geq 0, \ \forall y \in \mathcal{Y}, \tag{4.4}$$

where $\xi_{\tilde{x}^k} \in \partial\theta_1(\tilde{x}^k)$ and $\zeta_{\tilde{y}^k} \in \partial\theta_2(\tilde{y}^k)$.

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

$$\theta_1(x) - \theta_1(\tilde{x}^k) \geq \langle x - \tilde{x}^k, \xi_{\tilde{x}^k} \rangle + \frac{\mu}{2}\|\tilde{x}^k - x\|^2, \ \ \forall x \in \mathcal{X}, \tag{4.5}$$

and

$$\theta_2(y) - \theta_2(\tilde{y}^k) \geq \langle y - \tilde{y}^k, \zeta_{\tilde{y}^k} \rangle, \ \forall y \in \mathcal{Y}. \tag{4.6}$$

Combining (4.3) and (4.5), (4.4) and (4.6) yields

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

and

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

Adding the above two inequalities

$$\begin{aligned} &\theta(\boldsymbol{u}) - \theta(\tilde{\boldsymbol{u}}^k) - \frac{\mu}{2}\|\tilde{x}^k - x\|^2 \\ &+ \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \end{pmatrix}^\top \left\{ \begin{pmatrix} -A^\top \tilde{y}^k \\ A\tilde{x}^k \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) + A^\top(\tilde{y}^k - y^k) \\ \tau A(\tilde{x}^k - x^k) + s(\tilde{y}^k - y^k) \end{pmatrix} \right\} \geq 0. \end{aligned} \tag{4.7}$$

Then, it follows from the definitions of $\mathcal{G}(\cdot)$ and Q, that the assertion of this lemma is obtained.

□

Following the line of reasoning presented in [14] , we can also prove that Algorithm 1 is globally convergent. Here we only state the following theorem and omit the proof.

Theorem 4.2. *Let the sequence $\{\boldsymbol{u}^k := (x^k, y^k)\}$ be generated by Algorithm 1. Then, we have*

$$\begin{aligned} &\theta(\boldsymbol{u}) - \theta(\tilde{\boldsymbol{u}}^k) + \langle \boldsymbol{u} - \tilde{\boldsymbol{u}}^k, \mathcal{G}(\tilde{\boldsymbol{u}}^k) \rangle \\ &\geq \frac{1}{2}\left(\|\boldsymbol{u}^{k+1} - \boldsymbol{u}\|_H^2 - \|\boldsymbol{u}^k - \boldsymbol{u}\|_H^2\right) + \frac{1}{2}\|\boldsymbol{u}^k - \tilde{\boldsymbol{u}}^k\|_G^2 + \frac{\mu}{2}\|\tilde{x}^k - x\|^2, \ \forall \boldsymbol{u} \in \mathcal{U}, \end{aligned} \tag{4.8}$$

where $\mathcal{G}(\cdot)$ is given in (2.3b).

Proof. The proof details can be found in [14, Theorem 3.1] □

Remark 4.3. If $\theta_1(\cdot)$ is convex, the modulus $\mu = 0$. Then, [14, Theorem 3.1] can be recovered and

$$\begin{aligned} &\theta(\boldsymbol{u}) - \theta(\tilde{\boldsymbol{u}}^k) + \langle \boldsymbol{u} - \tilde{\boldsymbol{u}}^k, \mathcal{G}(\tilde{\boldsymbol{u}}^k) \rangle \\ &\geq \frac{1}{2}\left(\|\boldsymbol{u}^{k+1} - \boldsymbol{u}\|_H^2 - \|\boldsymbol{u}^k - \boldsymbol{u}\|_H^2\right) + \frac{1}{2}\|\boldsymbol{u}^k - \tilde{\boldsymbol{u}}^k\|_G^2, \ \forall \boldsymbol{u} \in \mathcal{U}. \end{aligned} \tag{4.9}$$

The following theorem indicates that the sequence $\{\boldsymbol{u}^k\}$ generated by Algorithm 1 is Fejér monotone with respect to the solution set of (2.3).

Theorem 4.4. *Let $\boldsymbol{u}^*$ be an arbitrary solution of* (1.1). *Then, the sequence $\{\boldsymbol{u}^k\}$ generated by Algorithm 1 satisfies*

$$\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^*\|_H^2 \leq \|\boldsymbol{u}^k - \boldsymbol{u}^*\|_H^2 - \|\boldsymbol{u}^k - \tilde{\boldsymbol{u}}^k\|_G^2 - \mu\|\tilde{x}^k - x^*\|^2. \tag{4.10}$$

Proof. The proof details can be found in [14, Theorem 3.2]. □

Remark 4.5. If $\theta_1(\cdot)$ is convex, the modulus $\mu = 0$. Then, [14, Theorem 3.2] can be recovered and

$$\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^*\|_H^2 \leq \|\boldsymbol{u}^k - \boldsymbol{u}^*\|_H^2 - \|\boldsymbol{u}^k - \tilde{\boldsymbol{u}}^k\|_G^2. \tag{4.11}$$

Theorem 4.6. *The sequence $\{\boldsymbol{u}^k\}$ generated by Algorithm 1 is globally convergent to a solution point of saddle point problem* (1.1).

4.2 Linear convergence under strategy 1

In this section, we prove the linear rate of convergence of Algorithm 1 under the following assumption.

Assumption 4.2. In problem (1.1), assume that $\mathcal{X} = \mathbb{R}^n$, A is full row rank, θ_2 is convex, θ_1 is strongly convex with modulus $\mu > 0$ and $\nabla\theta_1$ is Lipschitz continuous with constant L. For Algorithm 1, $rs > \|A^\top A\|$, α, β and τ satisfy the conditions in Assumption 4.1.

First, we begin our analysis with a fundamental inequality, which gives the bound of $\|y^{k+1} - y^*\|^2$ by using the terms $\|\tilde{x}^k - x^*\|^2$, $\|\tilde{x}^k - x^k\|^2$ and $\|\tilde{y}^k - y^k\|^2$.

Lemma 4.7. *Suppose Assumption 4.2 holds. Let $\boldsymbol{u}^*$ be a solution of* (1.1). *Then, the sequence $\{\boldsymbol{u}^k\}$ generated by Algorithm 1 satisfies*

$$\|y^{k+1} - y^*\|^2 \leq 3\kappa \left[L^2\|\tilde{x}^k - x^*\|^2 + \beta^2\|A^\top A\|\|\tilde{y}^k - y^k\|^2 + r^2\|\tilde{x}^k - x^k\|^2\right], \tag{4.12}$$

where

$$\kappa = [\lambda_{\min}(AA^\top)]^{-1} > 0, \tag{4.13}$$

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

Proof. Since A is full row rank, we have

$$\|y^{k+1} - y^*\|^2 \leq \ \kappa\|A^\top(y^{k+1} - y^*)\|^2. \tag{4.14}$$

Then, it follows from the first order optimality condition of the sub-problem (3.2a) and $\mathcal{X} = \mathbb{R}^n$ that

$$\nabla\theta_1(\tilde{x}^k) - A^\top y^k + r(\tilde{x}^k - x^k) = 0.$$

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

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

Then, it follows from the above two equations that

$$\begin{aligned}
&\left\|A^\top y^{k+1} - A^\top y^*\right\|^2 \\
&= \left\|\nabla\theta_1(\tilde{x}^k) + A^\top(y^{k+1} - y^k) + r(\tilde{x}^k - x^k) - \nabla\theta_1(x^*)\right\|^2 \\
&\leq 3\left\|\nabla\theta_1(\tilde{x}^k) - \nabla\theta_1(x^*)\right\|^2 + 3\|A^\top(y^{k+1} - y^k)\|^2 + 3r^2\left\|\tilde{x}^k - x^k\right\|^2 \\
&\leq 3\left[L^2\|\tilde{x}^k - x^*\|^2 + \|A^\top A\|\|y^{k+1} - y^k\|^2 + r^2\left\|\tilde{x}^k - x^k\right\|^2\right] \\
&\leq 3\left[L^2\|\tilde{x}^k - x^*\|^2 + \beta^2\|A^\top A\|\|\tilde{y}^k - y^k\|^2 + r^2\left\|\tilde{x}^k - x^k\right\|^2\right].
\end{aligned} \tag{4.15}$$

Thus, combining the inequalities (4.14) and (4.15) leads to (4.12). □

Theorem 4.8. *Suppose Assumption 4.2 holds. Let $\boldsymbol{u}^*$ be a solution of* (1.1). *Then, the sequence $\{\boldsymbol{u}^k\}$ generated by Algorithm 1 converges Q-linearly, i.e.,*

$$\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^*\|_H^2 \leq \frac{1}{1+\delta}\|\boldsymbol{u}^k - \boldsymbol{u}^*\|_H^2, \tag{4.16}$$

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

$$\left\{\begin{array}{l}
\delta_1 := \frac{\mu}{\|H\|(2+3\kappa L^2)}, \\
\delta_2 := \frac{\lambda_{\min}(G)}{\|H\|(3r^2\kappa+2(1-\alpha)^2)}, \\
\delta_3 := \frac{\lambda_{\min}(G)}{3\kappa\|H\|\|A^\top A\|\beta^2}.
\end{array}\right.$$

Proof. It follows from definition of the matrix H

$$\begin{aligned}
&\delta\|\boldsymbol{u}^{k+1}-\boldsymbol{u}^*\|_H^2 \\
&\leq \delta\|H\|\left(\|x^{k+1}-x^*\|^2+\|y^{k+1}-y^*\|^2\right) \\
&\leq \delta\|H\|\left(2\|x^{k+1}-\tilde{x}^k\|^2+2\|\tilde{x}^k-x^*\|^2+\|y^{k+1}-y^*\|^2\right) \\
&\leq \delta\|H\|\left(2(1-\alpha)^2\|x^k-\tilde{x}^k\|^2+2\|\tilde{x}^k-x^*\|^2+\|y^{k+1}-y^*\|^2\right).
\end{aligned}$$

Then, utilizing the above inequality and (4.12) leads to

$$\begin{aligned}
&\delta\|\boldsymbol{u}^{k+1}-\boldsymbol{u}^*\|_H^2 \\
&\leq \delta\|H\|\left[(2+3\kappa L^2)\|\tilde{x}^k-x^*\|^2+3\kappa\beta^2\|A^\top A\|\|\tilde{y}^k-y^k\|^2\right. \\
&\qquad\left.+(3\kappa r^2+2(1-\alpha)^2)\|\tilde{x}^k-x^k\|^2\right].
\end{aligned}$$

Furthermore,

$$\left\|\tilde{\boldsymbol{u}}^k-\boldsymbol{u}^k\right\|_G^2 \geq \lambda_{\min}(G)\left(\left\|\tilde{x}^k-x^k\right\|^2+\left\|\tilde{y}^k-y^k\right\|^2\right).$$

where $\lambda_{\min}(G)>0$ because of the positive definiteness of G. Then, it follows from the above two inequalities and that

$$\begin{aligned}
&(1+\delta)\left\|\boldsymbol{u}^{k+1}-\boldsymbol{u}^*\right\|_H^2 \\
&\leq \|\boldsymbol{u}^k-\boldsymbol{u}^*\|_H^2-\|\boldsymbol{u}^k-\tilde{\boldsymbol{u}}^k\|_G^2-\mu\|\tilde{x}^k-x^*\|^2+\delta\|\boldsymbol{u}^{k+1}-\boldsymbol{u}^*\|_H^2 \\
&\leq \|\boldsymbol{u}^k-\boldsymbol{u}^*\|_H^2-\left(\mu-\delta\|H\|(2+3\kappa L^2)\right)\|\tilde{x}^k-x^*\|^2 \\
&\quad -\left(\lambda_{\min}(G)-\delta\|H\|\left(2(1-\alpha)^2+3\kappa r^2\right)\right)\left\|\tilde{x}^k-x^k\right\|^2 \\
&\quad -\left(\lambda_{\min}(G)-3\kappa\delta\|H\|\|A^\top A\|\beta^2\right)\|\tilde{y}^k-y^k\|^2.
\end{aligned}$$

Thus, utilizing the definition of δ and the above inequality derives that

$$(1+\delta)\left\|\boldsymbol{u}^{k+1}-\boldsymbol{u}^*\right\|_H^2 \leq \|\boldsymbol{u}^k-\boldsymbol{u}^*\|_H^2,$$

and the assertion of the theorem is proved. □

Hence, the above theorem demonstrates that the Algorithm 1 can possess the Q-linear convergence under Assumption 4.2.

4.3 Linear convergence under strategy 2

Next, we analyze the global linear convergence of Algorithm 1 under some error bound hypotheses.

Assumption 4.3. In problem (1.1), both θ_1 and θ_2 are convex. Furthermore, assume that for any $\omega>0$, there exists $\eta>0$, such that

$$\operatorname{dist}(\boldsymbol{u},\mathcal{U}^*)\leq\eta\operatorname{dist}(0,\mathcal{E}(\boldsymbol{u},1)),\quad\forall\|\boldsymbol{u}\|\leq\omega,\ \boldsymbol{u}\in\mathcal{U}, \tag{4.17}$$

where $\mathcal{E}(\cdot)$ is defined by (2.5). For Algorithm 1, $rs>\|A^\top A\|$, α, β and τ satisfy the conditions in Assumption 4.2.

If the subdiferentials $\partial\theta_1$ and $\partial\theta_2$ are piecewise linear multi-functions and $\mathcal{X}$ and $\mathcal{Y}$ are polyhedral sets, then $\mathcal{E}(\boldsymbol{u},\gamma)$ are piecewise linear multi-functions and (4.17) holds by Theorem 2.6. This fact was utilized in [12, 13, 29] to prove the linear rate of convergence of Alternating Direction Methods of Multipliers (ADMM).

Lemma 4.9. *Let the sequence $\{\boldsymbol{u}^k := (x^k, y^k)\}$ be generated by Algorithm 1. Then, there exists a constant $\sigma_1 > 0$ such that*

$$dist^2(0, \mathcal{E}(\tilde{\boldsymbol{u}}^k, 1)) \le \sigma_1 \|\boldsymbol{u}^k - \tilde{\boldsymbol{u}}^k\|^2, \tag{4.18}$$

where

$$\sigma_1 = \max\{2(\|A^\top A\| + s^2), 2(r^2 + \tau^2\|A^\top A\|)\} > 0. \tag{4.19}$$

Proof. It follows from the optimality condition of $x-$subproblem (3.2a) that

$$\tilde{x}^k = P_{\mathcal{X}}[\tilde{x}^k - (\xi_{\tilde{x}^k} - A^\top y^k + r(\tilde{x}^k - x^k))].$$

Then,

$$\begin{aligned}
&\text{dist}^2\left(0, \mathcal{E}_{\mathcal{X}}(\tilde{\boldsymbol{u}}^k, 1)\right)\\
&= \text{dist}^2\big(\tilde{x}^k, P_{\mathcal{X}}(\tilde{x}^k - (\xi_{\tilde{x}^k} - A^\top \tilde{y}^k))\big)\\
&= \|P_{\mathcal{X}}[\tilde{x}^k - (\xi_{\tilde{x}^k} - A^\top y^k + r(\tilde{x}^k - x^k))] - P_{\mathcal{X}}[\tilde{x}^k - (\xi_{\tilde{x}^k} - A^\top \tilde{y}^k)]\|^2\\
&\le \|A^\top(y^k - \tilde{y}^k) + r(x^k - \tilde{x}^k)\|^2\\
&\le 2\|A^\top A\|\|y^k - \tilde{y}^k\|^2 + 2r^2\|x^k - \tilde{x}^k\|^2,
\end{aligned} \tag{4.20}$$

where the last inequality is derived by the inequality

$$\|a + b\|^2 \le 2a^2 + 2b^2,\ \forall a, b \in \mathbb{R}^n.$$

On the other hand, it follows from the optimality condition of the $y-$subproblem (3.2c) that

$$\tilde{y}^k = P_{\mathcal{Y}}[\tilde{y}^k - (\zeta_{\tilde{y}^k} + A\tilde{x}^k + \tau A(\tilde{x}^k - x^k) + s(\tilde{y}^k - y^k))].$$

Thus, we yield

$$\begin{aligned}
&\text{dist}^2(0, \mathcal{E}_{\mathcal{Y}}(\tilde{\boldsymbol{u}}^k, 1))\\
&= \text{dist}^2\big(\tilde{y}^k, P_{\mathcal{Y}}(\tilde{y}^k - (\zeta_{\tilde{y}^k} + A^\top \tilde{x}^k))\big)\\
&\le \|P_{\mathcal{Y}}[\tilde{y}^k - (\zeta_{\tilde{y}^k} + A^\top \tilde{x}^k + \tau A(\tilde{x}^k - x^k) + s(\tilde{y}^k - y^k))]\\
&\quad - P_{\mathcal{Y}}[\tilde{y}^k - (\zeta_{\tilde{y}^k} + A^\top \tilde{x}^k)]\|^2\\
&\le \left\|\tau A\big(\tilde{x}^k - x^k\big) + s(\tilde{y}^k - y^k))\right\|^2\\
&\le 2\tau^2\|A^\top A\|\|\tilde{x}^k - x^k\|^2 + 2s^2\|\tilde{y}^k - y^k\|^2,
\end{aligned} \tag{4.21}$$

Hence, combining (4.20) and (4.21) leads to

$$\begin{aligned}
&\text{dist}^2(0, \mathcal{E}(\tilde{\boldsymbol{u}}^k, 1))\\
&= \text{dist}^2(0, \mathcal{E}_{\mathcal{X}}(\boldsymbol{u}^{k+1}, 1)) + \text{dist}^2(0, \mathcal{E}_{\mathcal{Y}}(\boldsymbol{u}^{k+1}, 1))\\
&\le 2(\|A^\top A\| + s^2)\|y^k - \tilde{y}^k\|^2 + 2(\tau^2\|A^\top A\| + r^2)\|x^k - \tilde{x}^k\|^2.
\end{aligned} \tag{4.22}$$

The assertion of this lemma is in turn obtained by the definition of σ_1. $\square$

Now, we show the global linear convergence of Algorithm 1 under Assumption 4.3.

Theorem 4.10. *Suppose Assumption 4.3 holds. Let* $\{\boldsymbol{u}^k := (x^k, y^k)\}$ *be generated by Algorithm 1 and* $\mathcal{U}^*$ *be the solution set of problem* (1.1)*. Then,*

$$(1+\rho)\,dist_H^2(\boldsymbol{u}^{k+1},\mathcal{U}^*) \leq dist_H^2(\boldsymbol{u}^k,\mathcal{U}^*), \tag{4.23}$$

where

$$\rho := \frac{1}{\sigma\|H\|},\ \sigma = \frac{\sigma_2}{\sigma_3},\ \sigma_3 = \lambda_{\min}(G),$$

$$\textit{and}\quad \sigma_2 = \max\left\{2[(1-\alpha)^2+\eta^2\sigma_1], 2[(1-\beta)^2+\eta^2\sigma_1]\right\}$$

Proof. It follows from Theorem 4.6 that the sequence $\{(x^k, y^k)\}$ converges to a saddle point in $\mathcal{U}^*$, which is bounded. Then, using the Assumption 4.3, and (4.18) leads to that

$$\mathrm{dist}^2(\tilde{\boldsymbol{u}}^k,\mathcal{U}^*) \leq \eta^2\mathrm{dist}^2(0,\mathcal{E}(\tilde{\boldsymbol{u}}^k,1)) \leq \eta^2\sigma_1\|\boldsymbol{u}^k-\tilde{\boldsymbol{u}}^k\|^2. \tag{4.24}$$

On the other hand,

$$\begin{aligned}
&\mathrm{dist}^2(\boldsymbol{u}^{k+1},\mathcal{U}^*)\\
&\leq 2\mathrm{dist}^2(\boldsymbol{u}^{k+1},\tilde{\boldsymbol{u}}^k) + 2\mathrm{dist}^2(\tilde{\boldsymbol{u}}^k,\mathcal{U}^*)\\
&\leq 2\|\boldsymbol{u}^{k+1}-\tilde{\boldsymbol{u}}^k\|^2 + 2\eta^2\sigma_1\|\boldsymbol{u}^k-\tilde{\boldsymbol{u}}^k\|^2\\
&= 2(1-\alpha)^2\|x^k-\tilde{x}^k\|^2 + 2(1-\beta)^2\|y^k-\tilde{y}^k\|^2 + 2\eta^2\sigma_1\|\boldsymbol{u}^k-\tilde{\boldsymbol{u}}^k\|^2\\
&\leq \sigma_2\|\boldsymbol{u}^k-\tilde{\boldsymbol{u}}^k\|^2,
\end{aligned}$$

where $\sigma_2 = \max\{2[(1-\alpha)^2+\eta^2\sigma_1], 2[(1-\beta)^2+\eta^2\sigma_1]\}$. Furthermore,

$$\sigma_3\{\|y^k-\tilde{y}^k\|^2 + \|x^k-\tilde{x}^k\|^2\} \leq \|\boldsymbol{u}^k-\tilde{\boldsymbol{u}}^k\|_G^2,$$

where $\sigma_3 = \lambda_{\min}(G)$. Then, we conclude

$$\mathrm{dist}^2(\boldsymbol{u}^{k+1},\mathcal{U}^*) \leq \sigma\|\boldsymbol{u}^k-\tilde{\boldsymbol{u}}^k\|_G^2. \tag{4.25}$$

by setting $\sigma = \sigma_2/\sigma_3$. Notice that

$$\mathrm{dist}_H^2(\boldsymbol{u}^{k+1},\mathcal{U}^*) \leq \|H\|\mathrm{dist}^2(\boldsymbol{u}^k,\mathcal{U}^*),$$

we obtain that

$$\mathrm{dist}_H^2(\boldsymbol{u}^{k+1},\mathcal{U}^*) \leq \sigma\|H\|\|\boldsymbol{u}^k-\tilde{\boldsymbol{u}}^k\|_G^2. \tag{4.26}$$

Hence, combining (4.11) and (4.26) leads to

$$\begin{aligned}
\|\boldsymbol{u}^{k+1}-\boldsymbol{u}^*\|_H^2 &\leq \|\boldsymbol{u}^k-\boldsymbol{u}^*\|_H^2 - \|\boldsymbol{u}^k-\tilde{\boldsymbol{u}}^k\|_G^2\\
&\leq \|\boldsymbol{u}^k-\boldsymbol{u}^*\|_H^2 - \frac{1}{\sigma\|H\|}\mathrm{dist}_H^2(\boldsymbol{u}^{k+1},\mathcal{U}^*).
\end{aligned} \tag{4.27}$$

Let $\boldsymbol{u}^* \in \mathcal{U}^*$ such that $\mathrm{dist}_H(\boldsymbol{u}^k,\mathcal{U}^*) = \|\boldsymbol{u}^k-\boldsymbol{u}^*\|_H$. Then, it follows from (4.27) that

$$\mathrm{dist}_H^2(\boldsymbol{u}^{k+1},\mathcal{U}^*) \leq \mathrm{dist}_H^2(\boldsymbol{u}^k,\mathcal{U}^*) - \frac{1}{\sigma\|H\|}\mathrm{dist}_H^2(\boldsymbol{u}^{k+1},\mathcal{U}^*).$$

Thus, we obtain the desired inequality (4.23) immediately by rearranging terms. □

Remark 4.11. As mentioned in Remark 4.5, we can recover the convergence result in [14] and obtain the inequality (4.2) without the strong convexity assumption on θ_1. In this case, we can also derive the linear convergence of Algorithm 1 under this error bound condition.

5 Numerical Experiment

This section focuses on studying the numerical behavior of Algorithm 1. The algorithm was coded by MATLAB R2016a, and all experiments were executed on a Lenovo laptop with Windows 10 system and Inter(R) Core(TM) i7-7500 (2.70GH) CPU processor with a 16GB memory.

5.1 Example 1

Firstly, we consider the image inpainting model tested in [4], which takes the form as follows:

$$\min_{x} \max_{y\in\mathcal{C}} : \left\{ \langle y, Ax\rangle + \frac{\lambda}{2}\|Sx-g\|^2 \right\}, \tag{5.1}$$

where g is a noisy input image, $S \in \mathbb{R}^{N\times N}$ is a mask operator, and the convex set $\mathcal{C}$ is defined as $\mathcal{C} = \{y \mid \|y\|_\infty \le 1\}$. It is noticed that this problem can be seen as a special case of (1.1). Then, two color images: House.png (256×256) and Lena.png (512×512) were processed in our numerical simulation. To generate corrupted images, we enforce a character mask operator S for the first image House.png so that about 15% of pixels are missed, and adopt a line mask representing operator S for the second image Lena.png so that a part of pixels at rows are retained and the other (sdr= 75%) pixels are missed. Moreover, we add zero-mean Gaussian noise with the standard deviation 0.02 to the incomplete images. The original and degraded images are displayed in Figure 1.

Furthermore, we set $\lambda = 45$ in model (5.1) and employ the following stopping criterion:

$$\text{Rer} := \frac{\|x^{k+1}-x^k\|}{\|x^{k+1}\|} < \text{Tol}, \tag{5.2}$$

where $\{x^k\}$ is the sequence generated by Algorithm 1, and Tol is the error tolerance and set as 10^{-5}. The maximum number of iterations is set as 150. All algorithms start their iterations with the degraded images. 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. To investigate the sensitivity of the parameters in Algorithm 1, six sets of the parameters (τ, α, β) were tested in our numerical experiment. The related parameters are tuned by Assumption 4.1 and numerical simulation. The proximal parameters r and s are selected as $(1/3, 25)$ by the strategy in [16].

Figure 2 draws the evolution curves of SNR and Relative error (Rer) with respect to iterations. In Figure 2, we notice that Algorithm 1 can quickly achieve higher SNR values for the cases of $\tau = 1, \alpha = 1.8, \beta = 1.8$ and $\tau = 0.95, \alpha = 1.7, \beta = 1.7/0.95$ as compared to the other cases. Furthermore, Algorithm 1 can reach better relative error values in most of cases except that $\tau = 1, \alpha = 1, \beta = 1$ and $\tau = 0.8, \alpha = 1.5, \beta = 1.5/0.8$. Hence, it implies that Algorithm 1 is feasible and efficient when dealing with these image inpainting problems, and is also sensitive about the values of τ, α and β. Moreover, the evolution curves in Figure 2 further support the linear convergence behaviors of Algorithm 1.

Figure 1: From left to right: the original images, the degraded images (sdr= 75%), the restored images.

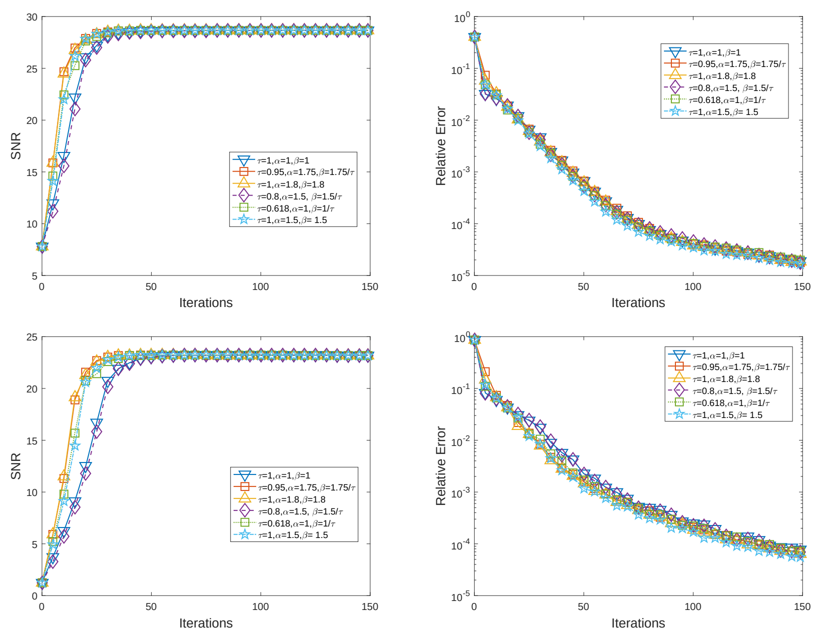

Figure 2: Evolution of SNRs and Relative errors defined by (5.2) with respect to iterations. The first row and second row correspond to image House.png and Lena.png (75%), respectively.

5.2 Example 2

Next, we deal with the matrix completion problem in [3]:

$$\min_{X} : \{\|X\|_* \;\mid\; X_{ij} = M_{ij},\ (i,j) \in \Omega\}\,, \tag{5.3}$$

where $X \in \mathbb{R}^{m\times n}$ is the unknown matrix to be completed,

$$\Omega = \{(i,j)|i \in \{1,2,\ldots,m\}, j \in \{1,2,\ldots,n\}\}$$

is an index set with cardinality p, M_{ij} represent the sampled (known) entries of M and $\|X\|_*$ is the nuclear norm. The data is generated by the way in [7]. By introducing a Lagrangian multiplier $Z_\Omega \in \mathbb{R}^{m\times n}$, the model (5.3) can be equivalent to the following saddle point problem:

$$\min_{X}\max_{Z_\Omega} : \{\|X\|_* - \langle Z_\Omega, X_\Omega - M_\Omega\rangle\}\,,$$

which can be seen as a special case of (1.1). Then, we use Algorithm 1 to solve the above matrix completion problem and the tolerance of the stopping criterion is set as 10^{-5}. The proximal parameters r and s are chosen as 0.005 and 202 by the strategy in [7, 14]. Here, we test five sets of the parameters: $(\tau,\alpha,\beta) \in \{(1,1.5,1.5),\ (0.618,1,1/0.618),\ (1,1,1),\ (0.95,1.7,1.7/0.95),\ (0.9,1.5,1.5/0.9)\}$, and plot the evolution curves of Relative error with respect to iterations and computing time (Second) in Figure 3. The results demonstrate that Algorithm 1 still performs stable for different dimensions. Comparatively, Algorithm 1 with $(0.618,1,1/0.618)$ performs better than the other cases. Moreover, the curves shown in Figure 3 further support the linear convergence behaviors of Algorithm 1.

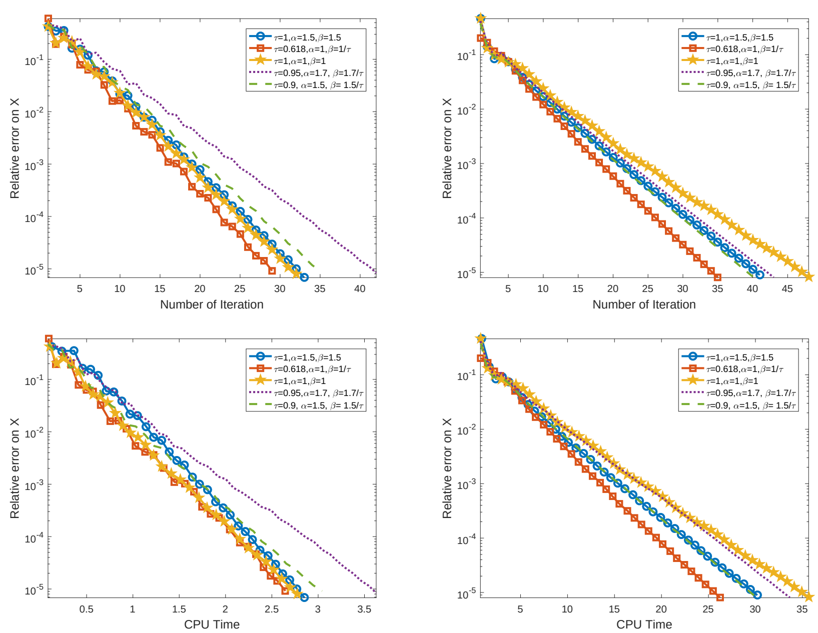

Figure 3: Sensitivity analysis on parameters of Algorithm 1 for matrix completion problems. Left: Dimension−500, Right: Dimension−1000.

6 Conclusion

In this paper, we studied the linear convergence of the generalized primal-dual hybrid gradient algorithm for the saddle point problem. This research project was completed by using two strategies, the first one is that one of the objective functions is strongly convex with Lipschitz continuous gradient, and the second one is that the problem possesses some error bound conditions. Some computational results illustrate the feasibility and efficient of this method.

References

[1] K.J. Arrow, L. Hurwicz and H. Uzawa, With contributions by H.B. Chenery, S.M. Johnson, S. Karlin, T. Marschak, and R.M. Solow*, Studies in Linear and Non-Linear Programming*, volume II of *Stanford Mathematical Studies in the Social Science*, Stanford Unversity Press, Stanford, California, 1958.

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

[3] E.J. Candès and B. Recht, Exacr matrix completion via convex optimization, *Found. Comput. Math.* 9 (2008) 717–772.

[4] 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.

[5] 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.

[6] X.K. Chang and J.F. Yang, A golden ratio primal–dual algorithm for structured convex optimization, *J. Sci. Comput.* 87 (2021)(47), https://doi.org/10.1007/s10915-021-01452-9.

[7] C.H. Chen, B.S. He and X.M. Yuan, Matrix completion via an alternating direction method, *IMA J. Numer. Anal* 32 (2012) 227–245.

[8] P.J. Chen, J.G. Huang and X.Q. Zhang, A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration, *Inverse Problems* 29 (2013): 025011.

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

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

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

[12] D.R. Han, D.F. Sun and L.W. Zhang, Linear rate convergence of the alternating direction method of multipliers for convex composite programming, *Math. Oper. Res.* 43 (2017) 622–637.

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

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

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

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

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

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

[19] F. Jiang, X.J. Cai, Z.M. Wu and H.C. Zhang, A first-order inexact primal-dual algorithm for a class of convex-concave saddle point problems, *Numer. Algor.* (2021), https://doi.org/10.1007/s11075-021-01069-x.

[20] Y.L. Liu, Y.B. Xu and W.T. Yin, Acceleration of primal–dual methods by preconditioning and simple subproblem procedures, *J. Sci. Comput.* 86 (2021): Article number 21.

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

[22] 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.

[23] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer, Boston, 2003.

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

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

[26] K. Wang and D.R. Han, On the linear convergence of the general first order primal-dual algorithm, *J. Ind. Manag. Optim.* (2021), doi:10.3934/jimo.2021134.

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

[28] M. Yan, A new primal–dual algorithm for minimizing the sum of three functions with a linear operator, *J. Sci. Comput.* 76 (2018) 1698–1717.

[29] W.H. Yang and D.R. 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.

[30] Y.D. Yu, T.Y. Lin, E. Mazumdar and M.I. Jordan, Fast distributionally robust learning with variance reduced min–max optimization, 2021, arXiv:2104.13326v1.

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

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

Manuscript received 8 June 2021
revised 21 January 2022
accepted for publication 28 February 2022

Kai Wang
School of Mathematics and Statistics
Nanjing University of Science and Technology
Nanjing, 210094, China
E-mail address: wangkaihawk@njust.edu.cn

Qun Wang
School of Data Sciences
Zhejiang University of Finance & Economics
Hangzhou, 310018, China
E-mail address: wangqun@zufe.edu.cn