CrossMark

# A primal–dual prediction–correction algorithm for saddle point optimization

**Hongjin He[1] · Jitamitra Desai[2] · Kai Wang[2]**

**Abstract** In this paper, we introduce a new primal–dual prediction–correction algorithm for solving a saddle point optimization problem, which serves as a bridge between the algorithms proposed in Cai et al. (J Glob Optim 57:1419–1428, 2013) and He and Yuan (SIAM J Imaging Sci 5:119–149, 2012). An interesting byproduct of the proposed method is that we obtain an easily implementable projection-based primal–dual algorithm, when the primal and dual variables belong to simple convex sets. Moreover, we establish the worst-case $\mathcal{O}(1/t)$ convergence rate result in an ergodic sense, where $t$ represents the number of iterations.

**Keywords** Saddle point problem · Primal–dual algorithm · Prediction–correction algorithm · Projection method · Convergence rate

## 1 Introduction

Consider the following saddle-point optimization problem:

$$\min_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} \ \Phi(x, y) := \langle y, Ax \rangle + \frac{\nu}{2} \|By - b\|^2, \tag{1.1}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are two nonempty, closed and convex subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively; $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{n \times n}$ are two given matrices; $b \in \mathbb{R}^n$ is a given (observation) vector; $\nu > 0$ serves as a tuning parameter; $\langle \cdot, \cdot \rangle$ denotes the standard inner product of vectors; and $\| \cdot \|$ is the Euclidean norm. It is well documented in the literature that the saddle point problem (1.1) has many applications in diverse areas such as constrained optimization duality, zero-sum games, and image processing (see [4–6, 8–10, 13]), amongst others.

✉ Jitamitra Desai
  jdesai@ntu.edu.sg

[1] Department of Mathematics, School of Science, Hangzhou Dianzi University, Hangzhou 310018, China

[2] School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore

To solve the saddle point problem (1.1), Chambolle and Pock [4] introduced the so-called first-order primal–dual algorithm (FOPDA), which has been successfully applied to solve many image restoration problems. In recent years, primal–dual methods have been extensively studied for solving convex optimization and saddle point problems arising in signal/image processing, computer vision, and machine learning (see [12] for a recent overview on such primal–dual algorithmic advancements). As pointed out in [4], the primal–dual hybrid gradient (PDHG) method [11,15] as well as the Arrow–Hurwicz–Uzawa method [1] are special cases of the FOPDA method. Although the conditions for global convergence of FOPDA are weaker than those in [1,15], the numerical results reported in [2,10] show that FOPDA [4] performs better in practice when its parameters are set beyond their theoretically guaranteed limits. Motivated by these results, He and Yuan [10] introduced a new class of primal–dual based prediction–correction algorithms where the range of the involved parameters in [4] was extended to further augment the performance of FOPDA. Subsequently, Cai et al. [2] presented an improved version of He and Yuan's algorithm with a new correction step, which further weakens the conditions assumed in [10] and also improves its computational efficiency.

In this paper, we propose another primal–dual algorithm, which aims to bridge the gap between the algorithms in [2,10]. As a byproduct, we also obtain a *partially linearized* primal–dual algorithm, which only requires two projections during the prediction stage. This resulting linearized version of our algorithm is relatively easier to implement as compared to the algorithms in [2,4,10] as long as the projections of $\mathcal{X}$ and $\mathcal{Y}$ are readily available as closed-form solutions or are simple enough to compute numerically. Recently, Chambolle and Pock [3] established the $\mathcal{O}(1/t)$ convergence rate of the FOPDA method in an ergodic sense. In a similar fashion, in the third contribution of this work, we also establish the $\mathcal{O}(1/t)$ convergence rate for the proposed algorithm, which further serves to provide an interesting complement to the methods introduced in [2,10].

The remainder of this paper is organized as follows. In Sect. 2, we begin by summarizing some basic notations and definitions, and follow this up by deriving an equivalent variational characterization of the saddle point problem (1.1). In Sect. 3, we propose the algorithmic framework of the primal–dual based prediction–correction method and briefly address its connections to existing primal–dual algorithms. More importantly, we describe a simplified projection-based primal–dual method obtained by virtue of appropriately chosen parameters. In Sect. 4, we establish the $\mathcal{O}(1/t)$ convergence rate of the proposed method in an ergodic sense, and finally, Sect. 5 summarizes the contributions of this paper and presents extensions for future research.

## 2 Preliminaries

In this section, we begin by summarizing some notations and definitions, and moreover, we reformulate the saddle point problem (1.1) as a variational inequality, which serves to facilitate the subsequent convergence analysis.

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 2-norm of an arbitrary matrix $N$.

Let $\Omega$ be a convex subset of $\mathbb{R}^n$. We denote $P_\Omega(\cdot)$ to be the projection of $(\cdot)$ onto $\Omega$ under the Euclidean norm, that is,

$$P_\Omega(v) := \arg\min \{\|u - v\| \mid u \in \Omega\}, \quad \forall v \in \mathbb{R}^n.$$

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

$$\max_{x \in \mathcal{X}} \left\{ \langle x, A^\top y^* \rangle + \frac{\nu}{2} \|By^* - b\|^2 \right\} \leq \langle x^*, A^\top y^* \rangle + \frac{\nu}{2} \|By^* - b\|^2$$
$$\leq \min_{y \in \mathcal{Y}} \left\{ \langle x^*, A^\top y \rangle + \frac{\nu}{2} \|By - b\|^2 \right\},$$

which reduces to the following pair of variational inequalities:

$$\begin{cases} \langle x - x^*, -A^\top y^* \rangle \geq 0, & \forall x \in \mathcal{X}, \\ \langle y - y^*, Ax^* + \nu B^\top (By^* - b) \rangle \geq 0, & \forall y \in \mathcal{Y}. \end{cases}$$

The above variational characterization can be compactly rewritten as a problem of finding $\boldsymbol{u}^* \in \Omega$, such that

$$\langle \boldsymbol{u} - \boldsymbol{u}^*, F(\boldsymbol{u}^*) \rangle \geq 0, \quad \forall \boldsymbol{u} \in \Omega, \tag{2.1a}$$

where

$$\boldsymbol{u} := \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(\boldsymbol{u}) := \begin{pmatrix} -A^\top y \\ Ax + \nu B^\top (By - b) \end{pmatrix}, \quad \text{and} \quad \Omega := \mathcal{X} \times \mathcal{Y}. \tag{2.1b}$$

It is easy to verify that the underlying mapping $F$ defined in (2.1b) is monotone, i.e.,

$$\langle \boldsymbol{u}_1 - \boldsymbol{u}_2, F(\boldsymbol{u}_1) - F(\boldsymbol{u}_2) \rangle \geq 0, \quad \forall \boldsymbol{u}_1, \boldsymbol{u}_2 \in \Omega.$$

Throughout this paper, our convergence analysis is based on the variational characterization (2.1) and the associated theory of variational inequalities.

## 3 Algorithm

To begin with, define:

$$M := \begin{pmatrix} \frac{1}{\tau} I & A^\top \\ \theta A & \frac{1}{\sigma} I + 2\rho B^\top B \end{pmatrix} \quad \text{and} \quad H := \begin{pmatrix} \frac{1}{\tau} I & 0 \\ 0 & \frac{1}{\sigma} I + 2\rho B^\top B \end{pmatrix}, \tag{3.1}$$

where $\tau, \theta, \sigma, \rho$ are given constants such that both $M$ and $H$ are positive definite, denoted using standard notation as $M \succ 0$ and $H \succ 0$.

We are now ready to formally present the algorithmic framework of the primal–dual prediction–correction algorithm where, for the sake of notational convenience, we have represented $\boldsymbol{u} := (x, y)$ and $\lambda_{\max}(B^\top B)$ denotes the maximum eigenvalue of $B^\top B$.

*Remark 3.1* When we set $\rho := 0$ in (3.2c), Algorithm 1 reduces to the first algorithm proposed in [10]. Besides, if we set $\rho := \nu$, while taking $H$ as the identity matrix and

$$\alpha_k^* := \frac{\langle \boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k, \widetilde{M}(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k) \rangle}{\|M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)\|^2}, \quad \text{where} \quad \widetilde{M} := \begin{pmatrix} \frac{1}{\tau} I & A^\top \\ \theta A & \frac{1}{\sigma} I + \nu B^\top B \end{pmatrix}, \tag{3.5}$$

as the step size [instead of $\alpha_k$ in (3.3)], the iterative scheme then coincides with the algorithm introduced in [2]. Clearly, Algorithm 1 can be regarded as an extension of the respective methods proposed in [2,10].

---

**Algorithm 1** A primal-dual prediction-correction algorithm.

---

1: Select $\tau > 0$, $\sigma > 0$, $\theta \in [-1, 1]$, and $\rho > -1/(2\sigma\lambda_{\max}(B^\top B))$.

2: **for** $k = 0, 1, 2, \cdots$ **do**

3: Generate a pair of predictors $\widetilde{\boldsymbol{u}}^k := (\widetilde{x}^k, \widetilde{y}^k)$ via

$$
\begin{cases}
\widetilde{x}^k = \arg\max_{x \in \mathcal{X}} \left\{ \tau\Phi(x, y^k) - \frac{1}{2}\|x - x^k\|^2 \right\}, & \text{(3.2a)} \\[2mm]
\bar{x}^k = \widetilde{x}^k + \theta(\widetilde{x}^k - x^k), & \text{(3.2b)} \\[2mm]
\widetilde{y}^k = \arg\min_{y \in \mathcal{Y}} \left\{ \sigma\Phi(\bar{x}^k, y) + \frac{1}{2}\|y - y^k\|^2 + \sigma\rho\|B(y - y^k)\|^2 \right\}. & \text{(3.2c)}
\end{cases}
$$

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

$$
\boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \gamma\alpha_k H^{-1}M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k), \quad \gamma \in (0, 2), \tag{3.3}
$$

where $H$ and $M$ are defined by (3.1) and

$$
\alpha_k := \frac{\langle \boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k, M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)\rangle}{\|H^{-1}M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)\|_H^2}. \tag{3.4}
$$

5: **end for**

---

Note that the $y$-related subproblem (3.2c) is essentially a constrained minimization problem if $\mathcal{Y}$ is a proper subset of $\mathbb{R}^n$. In this situation, Algorithm 1 fails to be easily implemented when $\rho := 0$ (as is the case in [2,4,10]). Indeed, observe that, when taking $\rho := -\nu/2$, the iterative schemes (3.2a) and (3.2c) amount to two projection steps, given by,

$$
\begin{cases}
\widetilde{x}^k = P_{\mathcal{X}}\left(x^k + \tau A^\top y^k\right), & \text{(3.6a)} \\[2mm]
\widetilde{y}^k = P_{\mathcal{Y}}\left(y^k - \sigma A\bar{x}^k + \sigma\nu B^\top(By^k - b)\right). & \text{(3.6b)}
\end{cases}
$$

Clearly, the projection steps (3.6a) and (3.6b) are computationally easier to implement than the original projection operators introduced in [2,4,10], as long as the sets $\mathcal{X}$ and $\mathcal{Y}$ are relatively simple. It is worthwhile to note that such a linearization strategy has appeared in a recent work [14], where the authors only considered the case of a fixed $\theta = 1$ in (3.2b), and showed that this variant yields significant computational benefits.

## 4 Convergence results

In this section, we first prove that Algorithm 1 is globally convergent to a solution of (1.1). Our secondary goal is to establish the worst-case $\mathcal{O}(1/t)$ convergence rate result in an ergodic sense, which is not estimated in [2,10]. In our analysis, we impose the same assumptions on the parameters $\tau$, $\sigma$, and $\delta$, as stated in [10].

**Lemma 4.1** *Let $\{\boldsymbol{u}^k := (x^k, y^k)\}$ be generated by Algorithm 1. Then, we have*

$$
\langle \boldsymbol{u} - \widetilde{\boldsymbol{u}}^k, F(\widetilde{\boldsymbol{u}}^k) + M(\widetilde{\boldsymbol{u}}^k - \boldsymbol{u}^k)\rangle \geq 0, \quad \forall \boldsymbol{u} \in \Omega, \tag{4.1}
$$

*where $M$ is as given in (3.1).*

*Proof* By invoking the first-order optimality conditions of (3.2a) and (3.2c), we have

$$\begin{cases} \langle x - \widetilde{x}^k, -A^\top y^k + \frac{1}{\tau}(\widetilde{x}^k - x^k) \rangle \geq 0, \quad \forall x \in \mathcal{X}, \\ \langle y - \widetilde{y}^k, A\bar{x}^k + \nu B^\top(B\widetilde{y}^k - b) + \left(\frac{1}{\sigma}I + 2\rho B^\top B\right)(\widetilde{y}^k - y^k) \rangle \geq 0, \quad \forall y \in \mathcal{Y}. \end{cases}$$

Using the notation in (2.1), together with (3.2b), yields the desired inequality (4.1). □

**Lemma 4.2** *Let $\boldsymbol{u}^* := (x^*, y^*)$ be an arbitrary solution of (1.1). Then, the sequences $\{\boldsymbol{u}^k\}$ and $\{\widetilde{\boldsymbol{u}}^k\}$ generated by Algorithm 1 satisfy the following property:*

$$\langle \boldsymbol{u}^k - \boldsymbol{u}^*, M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k) \rangle \geq \langle \boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k, M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k) \rangle. \tag{4.3}$$

*Proof* Note that the inequality (4.1) holds for any $\boldsymbol{u} \in \Omega$. As a consequence, we have

$$\langle \boldsymbol{u}^* - \widetilde{\boldsymbol{u}}^k, F(\widetilde{\boldsymbol{u}}^k) + M(\widetilde{\boldsymbol{u}}^k - \boldsymbol{u}^k) \rangle \geq 0, \quad \forall \boldsymbol{u}^* \in \Omega. \tag{4.4}$$

On the other hand, it follows from (2.1) that

$$\langle \widetilde{\boldsymbol{u}}^k - \boldsymbol{u}^*, F(\boldsymbol{u}^*) \rangle \geq 0. \tag{4.5}$$

Adding (4.4) and (4.5) yields

$$\langle \widetilde{\boldsymbol{u}}^k - \boldsymbol{u}^*, F(\boldsymbol{u}^*) - F(\widetilde{\boldsymbol{u}}^k) \rangle \geq \langle \boldsymbol{u}^* - \widetilde{\boldsymbol{u}}^k, M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k) \rangle,$$

which together with the monotonicity of $F$ implies that

$$\langle \widetilde{\boldsymbol{u}}^k - \boldsymbol{u}^*, M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k) \rangle \geq 0.$$

Thus, we conclude that

$$\langle \boldsymbol{u}^k - \boldsymbol{u}^*, M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k) \rangle \geq \langle \boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k, M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k) \rangle,$$

and the assertion of this lemma is obtained. □

*Remark 4.1* When $\rho = \nu$, note that the matrix $M$ can be expressed as the sum of two parts

$$M = \widetilde{M} + N, \quad \text{where} \quad N := \begin{pmatrix} 0 & 0 \\ 0 & \rho B^\top B \end{pmatrix},$$

and $\widetilde{M}$ is as previously defined in (3.5). Consequently, the term on the right hand side of (4.3) satisfies

$$\begin{aligned} \langle \boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k, M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k) \rangle &= \langle \boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k, \widetilde{M}(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k) \rangle + \rho \|B(y^k - \widetilde{y}^k)\|^2 \\ &\geq \langle \boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k, \widetilde{M}(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k) \rangle, \end{aligned}$$

and this validates the reasonableness of the choice of step size used in [2].

It is easy to verify that the matrix $H$, defined in (3.1), is always positive definite under the condition that $\rho > -1/(2\sigma\lambda_{\max}(B^\top B))$. Therefore, it follows that

$$\langle \boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k, M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k) \rangle \geq \frac{\delta}{1+\delta} \|\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k\|_H^2 > 0, \quad \forall \boldsymbol{u}^k \neq \widetilde{\boldsymbol{u}}^k. \tag{4.6}$$

A formal proof of this result is presented below, following a similar line of reasoning as used in [10,11].

**Lemma 4.3** *Let H and M be defined as* (3.1). *If the parameters $\tau$ and $\sigma$ satisfy*

$$\tau\sigma\frac{(1+\theta)^2}{4} < \frac{1}{\|A^\top A\|},$$

*we have*

$$\langle u - \widetilde{u}, M(u - \widetilde{u})\rangle \geq \frac{\delta}{1+\delta}\|u - \widetilde{u}\|_H^2 > 0, \quad \forall u \neq \widetilde{u},$$

*where*

$$\delta := \frac{2}{1+\theta}\sqrt{\frac{1}{\tau\sigma\|A^\top A\|}} - 1 > 0.$$

*Proof* Under the assumed condition that $\delta > 0$, it holds that

$$\tau(1+\delta)\|A^\top A\|\frac{(1+\theta)^2}{4} = \frac{1}{\sigma(1+\delta)}.$$

Using the Cauchy–Schwarz inequality yields

$$(1+\theta)\langle y - \widetilde{y}, A(x - \widetilde{x})\rangle$$

$$\geq -(\tau(1+\delta)\|A^\top A\|)\frac{(1+\theta)^2}{4}\|y - \widetilde{y}\|^2 - \frac{1}{\tau(1+\delta)\|A^\top A\|}\|A(x - \widetilde{x})\|^2$$

$$= -\frac{1}{\sigma(1+\delta)}\|y - \widetilde{y}\|^2 - \frac{1}{\tau(1+\delta)\|A^\top A\|}\|A(x - \widetilde{x})\|^2$$

$$\geq -\frac{1}{1+\delta}\left(\frac{1}{\tau}\|x - \widetilde{x}\|^2 + \frac{1}{\sigma}\|y - \widetilde{y}\|^2\right).$$

On the other hand,

$$\|u - \widetilde{u}\|_H^2 = \frac{1}{\tau}\|x - \widetilde{x}\|^2 + \frac{1}{\sigma}\|y - \widetilde{y}\|^2 + 2\rho\|y - \widetilde{y}\|_{B^\top B}^2$$

$$\geq \frac{1}{\tau}\|x - \widetilde{x}\|^2 + \frac{1}{\sigma}\|y - \widetilde{y}\|^2.$$

Combining the above inequalities leads to

$$\frac{1}{1+\delta}\|u - \widetilde{u}\|_H^2 + (1+\theta)\langle y - \widetilde{y}, A(x - \widetilde{x})\rangle \geq 0.$$

For any $u \neq \widetilde{u}$, we get

$$\langle u - \widetilde{u}, M(u - \widetilde{u})\rangle$$

$$= \|u - \widetilde{u}\|_H^2 + (1+\theta)\langle y - \widetilde{y}, A(x - \widetilde{x})\rangle$$

$$= \frac{\delta}{1+\delta}\|u - \widetilde{u}\|_H^2 + \frac{1}{1+\delta}\|u - \widetilde{u}\|_H^2 + (1+\theta)\langle y - \widetilde{y}, A(x - \widetilde{x})\rangle$$

$$\geq \frac{\delta}{1+\delta}\|u - \widetilde{u}\|_H^2 > 0.$$

Hence, the assertion of this lemma is proved.                                        $\square$

Therefore, it follows from Lemma 4.3 that the inequality (4.6) holds. Then, an immediate consequence to (4.3) and (4.6) is that,

$$\langle H(\boldsymbol{u}^k - \boldsymbol{u}^*), H^{-1}M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)\rangle \geq \frac{\delta}{1+\delta}\|\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k\|_H^2 > 0, \quad \forall \boldsymbol{u}^k \neq \widetilde{\boldsymbol{u}}^k.$$

Obviously, the above inequality indicates that, at any iteration $k$ of the algorithm, $-H^{-1}M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)$ is a *descent direction* of the distance function $\frac{1}{2}\|\boldsymbol{u} - \boldsymbol{u}^*\|_H^2$, which also ensures that the correction step (3.3) is well defined.

**Theorem 4.1** *Let $\boldsymbol{u}^*$ be an arbitrary solution of* (1.1). *Then, there exists a constant $c > 0$ such that the sequence $\{\boldsymbol{u}^k\}$ generated by Algorithm 1 satisfies*

$$\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^*\|_H^2 \leq \|\boldsymbol{u}^k - \boldsymbol{u}^*\|_H^2 - c\|\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k\|_H^2. \tag{4.7}$$

*Proof* Recalling the iterative scheme (3.3), we have

$$
\begin{aligned}
\|\boldsymbol{u}^{k+1} &- \boldsymbol{u}^*\|_H^2 \\
&= \|\boldsymbol{u}^k - \gamma\alpha_k H^{-1}M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k) - \boldsymbol{u}^*\|_H^2 \\
&= \|\boldsymbol{u}^k - \boldsymbol{u}^*\|_H^2 - 2\gamma\alpha_k\langle \boldsymbol{u}^k - \boldsymbol{u}^*, M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)\rangle + \gamma^2\alpha_k^2\|H^{-1}M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)\|_H^2 \\
&\leq \|\boldsymbol{u}^k - \boldsymbol{u}^*\|_H^2 - 2\gamma\alpha_k\langle \boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k, M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)\rangle + \gamma^2\alpha_k^2\|H^{-1}M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)\|_H^2 \\
&= \|\boldsymbol{u}^k - \boldsymbol{u}^*\|_H^2 - \gamma(2-\gamma)\alpha_k\langle \boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k, M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)\rangle \\
&\leq \|\boldsymbol{u}^k - \boldsymbol{u}^*\|_H^2 - \gamma(2-\gamma)\alpha_k \frac{\delta}{1+\delta}\|\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k\|_H^2.
\end{aligned}
\tag{4.8}
$$

It follows from the definition of $\alpha_k$, defined by (3.4), that

$$
\begin{aligned}
\alpha_k &:= \frac{(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)^\top M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)}{\|H^{-1}M(\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k)\|_H^2} \\
&\geq \frac{\delta\|\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k\|_H^2}{(1+\delta)\|M^\top H^{-1}M\|\|\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k\|^2} \\
&\geq \frac{\delta\lambda_{\min}(H)}{(1+\delta)\|M^\top H^{-1}M\|} \\
&= \frac{\delta \cdot \min\left\{\frac{1}{\tau}, \frac{1}{\sigma} + 2\rho\lambda_{\min}(B^\top B)\right\}}{(1+\delta)\|M^\top H^{-1}M\|} \\
&:= \alpha_{\min} > 0,
\end{aligned}
$$

where $\lambda_{\min}(\bullet)$ denotes the minimum eigenvalue of a matrix $(\bullet)$.

Therefore, we conclude from (4.8) that

$$\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^*\|_H^2 \leq \|\boldsymbol{u}^k - \boldsymbol{u}^*\|_H^2 - \left(\frac{\gamma(2-\gamma)\delta\alpha_{\min}}{1+\delta}\right)\|\boldsymbol{u}^k - \widetilde{\boldsymbol{u}}^k\|_H^2.$$

Setting $c := (\gamma(2-\gamma)\delta\alpha_{\min})/(1+\delta)$ completes the proof. □

The above theorem proves that the sequence $\{\boldsymbol{u}^k\}$ generated by Algorithm 1 is Fejér monotone with respect to the solution set of (2.1). Following the line of reasoning presented in [10, Theorem 3.7] (see also [2,4]), we can also prove that Algorithm 1 enjoys global convergence. Therefore, we only state the following theorem and refer the reader to the aforementioned references for a detailed proof of this result.

**Theorem 4.2** *The sequence* $\{u^k\}$ *generated by Algorithm 1 is globally convergent to a solution point of* (1.1).

Next, we prove that Algorithm 1 has a worst-case $\mathcal{O}(1/t)$ convergence rate in an ergodic sense. We begin this analysis with a fundamental inequality proved in the following lemma.

**Lemma 4.4** *Let the sequences* $\{u^k\}$ *and* $\{\widetilde{u}^k\}$ *be generated by Algorithm 1. Then, the following inequality*

$$\langle u - \widetilde{u}^k, F(u)\rangle + \frac{1}{2\gamma\alpha_k}\left(\|u - u^k\|_H^2 - \|u - u^{k+1}\|_H^2\right) \geq \frac{(2-\gamma)\delta}{2(1+\delta)}\|u^k - \widetilde{u}^k\|_H^2,$$

(4.9)

*holds for all* $u \in \Omega$, *where* $H$ *is as defined in* (3.1).

*Proof* Rearranging the terms in (4.1), we get

$$\langle u - \widetilde{u}^k, F(\widetilde{u}^k)\rangle \geq \langle u - \widetilde{u}^k, M(u^k - \widetilde{u}^k)\rangle.$$

(4.10)

It follows from the monotonicity of $F$ that

$$\langle u - \widetilde{u}^k, F(u)\rangle \geq \langle u - \widetilde{u}^k, F(\widetilde{u}^k)\rangle, \quad \forall u \in \Omega.$$

Clearly, from (4.10), we have

$$\langle u - \widetilde{u}^k, F(u)\rangle \geq \langle u - \widetilde{u}^k, M(u^k - \widetilde{u}^k)\rangle, \quad \forall u \in \Omega.$$

(4.11)

Note that the correction step (3.3) can be rewritten as

$$u^k - \widetilde{u}^k = \frac{1}{\gamma\alpha_k}M^{-1}H(u^k - u^{k+1}),$$

and consequently, we have

$$\langle u - \widetilde{u}^k, M(u^k - \widetilde{u}^k)\rangle = \frac{1}{\gamma\alpha_k}\langle u - \widetilde{u}^k, H(u^k - u^{k+1})\rangle.$$

(4.12)

An application, on the right hand side of (4.12), of the following equality

$$\langle a - b, H(c - d)\rangle = \frac{1}{2}\left(\|a - d\|_H^2 - \|a - c\|_H^2\right) + \frac{1}{2}\left(\|c - b\|_H^2 - \|d - b\|_H^2\right)$$

with settings $a := u$, $b := \widetilde{u}^k$, $c := u^k$, and $d := u^{k+1}$, yields

$$\langle u - \widetilde{u}^k, H(u^k - u^{k+1})\rangle$$
$$= \frac{1}{2}\left(\|u - u^{k+1}\|_H^2 - \|u - u^k\|_H^2\right) + \frac{1}{2}\left(\|u^k - \widetilde{u}^k\|_H^2 - \|u^{k+1} - \widetilde{u}^k\|_H^2\right).$$

(4.13)

The last term of (4.13) satisfies

$$\frac{1}{2}\left(\|u^k - \widetilde{u}^k\|_H^2 - \|u^{k+1} - \widetilde{u}^k\|_H^2\right)$$
$$= \frac{1}{2}\left(\|u^k - \widetilde{u}^k\|_H^2 - \|u^k - \gamma\alpha_k H^{-1}M(u^k - \widetilde{u}^k) - \widetilde{u}^k\|_H^2\right)$$
$$= \frac{1}{2}\gamma(2-\gamma)\alpha_k\langle u^k - \widetilde{u}^k, M(u^k - \widetilde{u}^k)\rangle$$
$$\geq \frac{\delta\gamma(2-\gamma)\alpha_k}{2(1+\delta)}\|u^k - \widetilde{u}^k\|_H^2,$$

(4.14)

where the first equality comes from (3.3) and the second inequality follows from (4.6). Combining (4.11), (4.12), (4.13), and (4.14) leads to

$$\langle u - \widetilde{u}^k, F(u) \rangle \geq \frac{1}{2\gamma\alpha_k} \left( \|u - u^{k+1}\|_H^2 - \|u - u^k\|_H^2 \right) + \frac{(2-\gamma)\delta}{2(1+\delta)} \|u^k - \widetilde{u}^k\|_H^2,$$

and the desired result is obtained by rearranging the terms in the above inequality. $\quad\square$

**Theorem 4.3** *For any integer $t > 0$, we denote*

$$\bar{u}_t = \frac{1}{\Upsilon_t} \sum_{k=0}^{t} \alpha_k \widetilde{u}^k, \quad with \quad \Upsilon_t = \sum_{k=0}^{t} \alpha_k, \tag{4.15}$$

*where $\widetilde{u}^k$ ($k = 0, 1, \cdots, t$) are generated by Algorithm 1 and $\alpha_k$ is given by (3.4). Then, we have $\bar{u}_t \in \Omega$ and*

$$\langle \bar{u}_t - u, F(u) \rangle \leq \frac{1}{2\gamma\alpha_{\min}(t+1)} \|u - u^0\|_H^2.$$

*Proof* First, as a direct consequence of (4.9), we have

$$\alpha_k \langle u - \widetilde{u}^k, F(u) \rangle + \frac{1}{2\gamma} \|u - u^k\|_H^2 \geq \frac{1}{2\gamma} \|u - u^{k+1}\|_H^2, \quad \forall u \in \Omega.$$

Summing the above inequality over $k = 0, 1, \cdots, t$ results in

$$\left\langle \Upsilon_t u - \sum_{k=0}^{t} \alpha_k \widetilde{u}^k, F(u) \right\rangle + \frac{1}{2\gamma} \|u - u^0\|_H^2 \geq 0, \quad \forall u \in \Omega.$$

As $\alpha_k \geq \alpha_{\min}$, we have $\Upsilon_t \geq (t+1)\alpha_{\min}$ and

$$\frac{1}{\Upsilon_t} \leq \frac{1}{(t+1)\alpha_{\min}}.$$

Therefore,

$$\left\langle \frac{1}{\Upsilon_t} \sum_{k=0}^{t} \alpha_k \widetilde{u}^k - u, F(u) \right\rangle \leq \frac{1}{2\gamma\Upsilon_t} \|u - u^0\|_H^2 \leq \frac{1}{2\gamma\alpha_{\min}(t+1)} \|u - u^0\|_H^2,$$

and the assertion of this theorem is obtained. $\quad\square$

The foregoing theorem shows that for a given arbitrary compact set $\mathcal{D} \subset \Omega$, after $t$ iterations of Algorithm 1, we obtain an approximate solution $\bar{u}_t$, given by (4.15), for the variational inequality (2.1) [and equivalently for the saddle point problem (1.1)] such that

$$\sup \{\langle \bar{u}_t - u, F(u) \rangle\} \leq \epsilon,$$

where $\epsilon := \frac{d^2}{2\gamma\alpha_{\min}(t+1)}$, with $d := \sup\{\|u - u^0\|_H \mid u \in \Omega\}$. Hence, the $\mathcal{O}(1/t)$ convergence rate of Algorithm 1 is established in an ergodic sense.

## 5 Conclusion

In this paper, we introduced a primal–dual prediction–correction algorithm for solving a saddle point optimization problem. The proposed algorithm enjoys an algorithmic framework that bridges the gap between the approaches described in [2] and [10]. As a byproduct, we also obtain a projection-based primal–dual algorithm by choosing an appropriate proximal parameter, which is comparatively easier to implement than the methods described in [2,4,10], as long as the projections of both $\mathcal{X}$ and $\mathcal{Y}$ are simple enough to be computed. Furthermore, we also established that our proposed algorithm is globally convergent with a worst-case $\mathcal{O}(1/t)$ convergence rate in an ergodic sense, which supports the numerical performance reported in [2,10] from a theoretical perspective.

In conclusion, note that the computational experience reported in several previous works in the literature has indicated that the speed of convergence of the PDHG method is highly sensitive with respect to the choice of the values of $\tau$ and $\sigma$. Most recently, the authors [7] proposed an adaptive version of PDHG to accelerate its numerical performance. Taking a cue from this approach, we can also gainfully employ an adaptive strategy for the acceleration of our method, and the results of this research are forthcoming.

## References

1. Arrow, K., Hurwicz, L., Uzawa, H.: With contributions by H.B. Chenery, S.M. Johnson, S. Karlin, T. Marschak, and R.M. Solow. Studies in Linear and Non-Linear Programming. Stanford Mathematical Studies in the Social Sciences, vol. II. Stanford Unversity Press, Stanford (1958)
2. Cai, X., Han, D., Xu, L.: An improved first-order primal-dual algorithm with a new correction step. J. Glob. Optim. **57**, 1419–1428 (2013)
3. Chambolle, A., Pock, T.: On the ergodic convergence rates of a first order primal dual algorithm. Math. Program. Ser. A. doi:10.1007/s10107-015-0957-3
4. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**, 120–145 (2011)
5. Chen, Y., Lan, G., Ouyang, Y.: Optimal primal dual methods for a class of saddle point problems. SIAM J. Optim. **24**, 1779–1814 (2014)
6. Esser, E., Zhang, X., Chan, T.: A general framework for a class of first-order primal-dual algorithms for convex optimization in imaging sciences. SIAM J. Imaging Sci. **3**, 1015–1046 (2010)
7. Goldstein, T., Li, M., Yuan, X., Esser, E., Baraniuk, R.: Adaptive primal-dual hybrid gradient methods for saddle point problems (2015). ArXiv:1305.0546v2
8. Gu, G., He, B., Yuan, X.: Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a uniform approach. Comput. Optim. Appl. **59**, 135–161 (2014)
9. Han, D., Xu, W., Yang, H.: An operator splitting method for variational inequalities with partially unknown mappings. Numer. Math. **111**, 207–237 (2008)
10. He, B., Yuan, X.: Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIAM J. Imaging Sci. **5**, 119–149 (2012)
11. He, B.S., You, Y., Yuan, X.M.: On the convergence of primal dual hybrid gradient algorithm. SIAM J. Imaging Sci. **7**, 2526–2537 (2015)
12. Komodakis, N., Pesquet, J.C.: Playing with duality: an overview of recent primal dual approaches for solving large scale optimization problems. IEEE Signal Process Mag. **32**(6), 31–54 (2015)
13. Nemirovski, A.: 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**, 229–251 (2004)

14. Tian, W., Yuan, X.: Linearized primal-dual methods for linear inverse problems with total variation regularization and finite element discretization, Working Paper (2015). http://www.math.hkbu.edu.hk/~xmyuan/Paper/LPD-TV-June19.pdf
15. Zhu, M., Chan, T.: An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration. CAM Reports 08-34, UCLA (2008)