An Algorithmic Framework of Generalized Primal-Dual Hybrid Gradient Methods for Saddle Point Problems

Bingsheng He[1] **Feng Ma**[2] **Xiaoming Yuan** [3]

January 30, 2016

Abstract. The primal-dual hybrid gradient method (PDHG) originates from the Arrow-Hurwicz method, and it has been widely used to solve saddle point problems, particularly in image processing areas. With the introduction of a combination parameter, Chambolle and Pock proposed a generalized PDHG scheme with both theoretical and numerical advantages. It has been analyzed that except for the special case where the combination parameter is 1, the PDHG cannot be casted to the proximal point algorithm framework due to the lack of symmetry in the matrix associated with the proximal regularization terms. The PDHG scheme is asymmetric also in the sense that one variable is updated twice while the other only updated once at each iteration. These asymmetry features also explain why more theoretical issues remain challenging for generalized PDHG schemes; for example, the worst-case convergence rate of PDHG measured by the iteration complexity in a nonergodic sense is still missing. In this paper, we further consider how to generalize the PDHG and propose an algorithmic framework of generalized PDHG schemes for saddle point problems. This algorithmic framework allows the output of the PDHG subroutine to be further updated by correction steps with constant step sizes. We investigate the restriction onto these step sizes and conduct the convergence analysis for the algorithmic framework. The algorithmic framework turns out to include some existing PDHG schemes as special cases; and it immediately yields a class of new generalized PDHG schemes by choosing different step sizes for the correction steps. In particular, a completely symmetric PDHG scheme with the golden-ratio step sizes is included. Theoretically, an advantage of the algorithmic framework is that the worst-case convergence rate measured by the iteration complexity in both the ergodic and nonergodic senses can be established.

Keywords. Primal-dual hybrid gradient method, saddle point problem, convex programming, image restoration, convergence rate, variational inequalities.

1 Introduction

We consider the saddle point problem

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

[1]Department of Mathematics, South University of Science and Technology of China, Shenzhen, 518055, China, and Department of Mathematics, Nanjing University, Nanjing, 210093, China. This author was supported by the NSFC Grant 11471156. Email: hebma@nju.edu.cn

[2]College of Communications Engineering, PLA University of Science and Technology, Nanjing, 210007, China. Email: mafengnju@gmail.com

[3]Department of Mathematics, Hong Kong Baptist University, Hong Kong. This author was supported by the General Research Fund from Hong Kong Research Grants Council: HKBU 12300515. Email: xmyuan@hkbu.edu.hk

where $A \in \Re^{m\times n}, b \in \Re^m$, $\mathcal{X} \subseteq \Re^n$ and $\mathcal{Y} \subseteq \Re^m$ are closed convex sets, $\theta_1 : \Re^n \to \Re$ and $\theta_2 : \Re^m \to \Re$ are convex but not necessarily smooth functions. The solution set of (1.1) is assumed to be nonempty throughout our discussion. The model (1.1) captures a variety of applications in different areas. For examples, finding a saddle point of the the Lagrangian function of the canonical convex minimization model with linear equality or inequality constraints is a special case of (1.1). Moreover, a number of variational image restoration problems with the total variation (TV) regularization (see [24]) can be reformulated as special cases of (1.1), see details in, e.g., [4, 26, 27, 28].

Since the work [28], the primal-dual hybrid gradient (PDHG) method has attracted much attention from the image processing field and it is still being widely used for many applications. We start the discussion of PDHG with the scheme proposed in [4]:

$$\left\{ \begin{array}{rcl} x^{k+1} &=& \arg\min\{\Phi(x,y^k) + \frac{r}{2}\|x-x^k\|^2 \mid x \in \mathcal{X}\}, \\ \bar{x}^k &=& x^{k+1} + \tau(x^{k+1}-x^k), \\ y^{k+1} &=& \arg\max\{\Phi(\bar{x}^k,y) - \frac{s}{2}\|y-y^k\|^2 \mid y \in \mathcal{Y}\}, \end{array} \right. \tag{1.2}$$

where $\tau \in [0,1]$ is a combination parameter, $r > 0$ and $s > 0$ are proximal parameters of the regularization terms. The scheme (1.2) splits the coupled term y^TAx in (1.1) and treats the functions θ_1 and θ_2 individually; the resulting subproblems are usually easier than the original problem (1.1). In [4], it was shown that the PDHG scheme (1.2) is closely related to the extrapolational gradient methods in [19, 23], the Douglas-Rachford splitting method in [8, 20] and the alternating direction method of multipliers (ADMM) in [11]. In particular, we refer to [9, 25] for the equivalence between a special case of (1.2) and a linearized version of the ADMM. As analyzed in [4], the convergence of (1.2) can be guaranteed under the condition

$$rs > \|A^TA\|. \tag{1.3}$$

When $\tau = 0$, the scheme (1.2) reduces to the PDHG scheme in [28] which is indeed the Arrow-Hurwicz method in [1]. In addition to the numerical advantages shown in, e.g., [4, 15], the theoretical significance of extending $\tau = 0$ to $\tau \in [0,1]$ was demonstrated in [14]. That is, if $\tau = 0$, then the scheme (1.2) could be divergent even if r and s are fixed at very large values; thus the condition (1.3) is not sufficient to ensure the convergence of (1.2). Note that it could be numerically beneficial to tune the parameters r and s as shown in, e.g., [15, 27]; and it is still possible to investigate the convergence of the PDHG scheme (1.2) with adaptively-adjusting proximal parameters, see, e.g., [3, 9, 12]. Here, to expose our main idea more clearly and to avoid heavy notation, we only focus on the case where both r and s are constant and they satisfy the condition in (1.3) throughout our discussion.

In [15], we showed that the special case of (1.2) with $\tau = 1$ is an application of the proximal point algorithm (PPA) in [21], and thus the acceleration scheme in [13] can be immediately used to accelerate the PDHG scheme (see Algorithm 4 in [15]). Its numerical efficiency has also been verified therein. This PPA revisit has been further studied in [22], in which a preconditioning version of the PDHG scheme (1.2) was proposed. When $\tau \neq 1$[4], it is shown in [15] (see also (2.7b)) that the matrix associated with the proximal regularization terms in (1.2) is not symmetric and thus the scheme (1.2) cannot be casted to an application of the PPA. The acceleration techniques in [13] are thus not applicable to (1.2). But the convergence can be guaranteed if the output of the PDHG subroutine (1.2) is further corrected by some correction steps (See Algorithms 1 and 2 in [15]). The step sizes of these correction steps, however, must be calculated iteratively and they usually require expensive computation (e.g., multiplications

[4]We analyzed the case $\tau \in [-1,1]$ in [15]; but for simplicity we only focus on $\tau \in [0,1]$ in this paper.

of matrices or matrices and vectors) if high dimensional variables are considered. It is thus natural to ask whether or not we can further correct the output of (1.2) even when $\tau \neq 1$ while the step sizes of the correction steps are iteration-independent (more precisely, constants)? Moreover, it is easy to notice that the PDHG scheme (1.2) is not symmetric also in the sense that the variable x is updated twice while y is updated only once at each iteration. So one more question is whether or not we can update y also right after the PDHG subroutine and modify (1.2) as a completely symmetric version? To answer these questions, we propose the following unified algorithmic framework that allows the variables to be updated both, either, or neither, after the PDHG subroutine (1.2):

$$
\text{(Algorithmic Framework)} \quad \left\{ \begin{array}{l} \left\{ \begin{array}{lcl} \tilde{x}^k &=& \arg\min\{\Phi(x,y^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, \\ \bar{x}^k &=& \tilde{x}^k + \tau(\tilde{x}^k - x^k), \\ \tilde{y}^k &=& \arg\max\{\Phi(\bar{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}, \end{array} \right. \quad (1.4a) \\ \left\{ \begin{array}{lcl} x^{k+1} &=& x^k - \alpha(x^k - \tilde{x}^k), \\ y^{k+1} &=& y^k - \beta(y^k - \tilde{y}^k), \end{array} \right. \quad (1.4b) \end{array} \right.
$$

where the output of the PDHG subroutine (1.2) is denoted by $(\tilde{x}^k, \tilde{y}^k)$, $\alpha > 0$ and $\beta > 0$ are step sizes of the correction steps to further update the variables. This new algorithmic framework is thus a combination of the PDHG subroutine (1.2) with a correction step. We also call $(\tilde{x}^k, \tilde{y}^k)$ generated by (1.4a) a predictor and the new iterate (x^{k+1}, y^{k+1}) updated by (1.4b) a corrector. To see the generality of (1.4), when $\alpha = \beta = 1$, the algorithmic framework (1.4) reduces to the generalized PDHG (1.2) in [4]; when $\tau = 1$ and $\alpha = \beta \in (0, 2)$, the accelerated version of PDHG from the PPA revisit (Algorithm 4 in [15]) is recovered. Moreover, if $\alpha = 1$, we obtain a symmetric generalized PDHG (see (5.9)) that updates both the variables twice at each iteration and even a completely symmetric version with the golden-ratio step sizes (see (5.12)). Therefore, based on this algorithmic framework, some existing PDHG schemes can be recovered and a class of new generalized PDHG schemes can be easily proposed by choosing different values of α, β and τ. We will focus on the case where $\tau \in [0, 1]$ and investigate the restriction onto these constants so that the convergence of the algorithmic framework can be ensured. The analysis is conducted in a unified manner. In addition, the mentioned asymmetry features of the PDHG scheme (1.2) may also be the reason why some critical theoretical issues remain challenging. For example, the worst-case convergence rate measured by the iteration complexity in a nonergodic sense is still open, even when θ_1 or θ_2 is strongly convex, as analyzed in [14]. We will show that for the generalized PDHG algorithmic framework (1.4), the worst-case convergence rate in both the ergodic and nonergodic senses can be derived when the parameters are appropriately restricted. This is a theoretical advantage of the algorithmic framework (1.4).

Finally, we would mention that there are a rich set of literature discussing how to extend the PDHG with more sophisticated analysis to more complicated saddle point models or to more abstract spaces, e.g., [5, 6, 7]. But in this paper we concentrate on the canonical saddle point model (1.1) in a finite dimensional space and the representative PDHG scheme (1.2) to present our idea of algorithmic design clearly.

The rest of this paper is organized as follows. In Section 2, we summarize some understandings on the model (1.1) and the algorithmic framework (1.4) from the variational inequality perspective. Then, we prove the convergence of (1.4) in Section 3 and derive its worst-case convergence rate measured by the iteration complexity in Section 4. In Section 5, we elaborate on the conditions that can ensure the convergence of (1.4); some specific PDHG schemes are thus yielded based on the algorithmic framework (1.4). Finally, we draw some conclusions in Section 6.

2 Variational Inequality Understanding

In this section, we provide the variational inequality (VI) reformulation of the model (1.1) and characterize the algorithmic framework (1.4) via a VI. These VI understandings are the basis of our analysis to be conducted.

2.1 Variational Inequality Reformulation of (1.1)

We first show that the saddle point problem (1.1) can be written as a VI problem. More specifically, if $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ is a solution point of the saddle point problem (1.1), then we have

$$\Phi_{y\in\mathcal{Y}}(x^*, y) \le \Phi(x^*, y^*) \le \Phi_{x\in\mathcal{X}}(x, y^*). \tag{2.1}$$

Obviously, the second inequality in (2.1) implies that

$$x^* \in \mathcal{X}, \;\; \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T y^*) \ge 0, \;\; \forall\, x \in \mathcal{X};$$

and the first one in (2.1) implies that

$$y^* \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(Ax^* - b) \ge 0, \;\; \forall\, y \in \mathcal{Y}.$$

Therefore, finding a solution point (x^*, y^*) of (1.1) is equivalent to solving the VI problem: Find $u^* = (x^*, y^*)$ such that

$$\mathrm{VI}(\Omega, F, \theta) \qquad u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u - u^*)^T F(u^*) \ge 0, \quad \forall\, u \in \Omega, \tag{2.2a}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad F(u) = \begin{pmatrix} -A^T y \\ Ax - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y}. \tag{2.2b}$$

We denote by Ω^* the set of all solution points of $\mathrm{VI}(\Omega, F, \theta)$ (2.2). Notice that Ω^* is convex (see Theorem 2.3.5 in [10] or Theorem 2.1 in [16]).

Clearly, for the mapping F given in (2.2b), we have

$$(u - v)^T(F(u) - F(v)) = 0, \quad \forall\, u, v \in \Omega. \tag{2.3}$$

Moreover, under the condition (1.3), the matrix

$$\begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \tag{2.4}$$

is positive definite. The positive definiteness of this matrix plays a significant role in analyzing the convergence of the PDHG scheme (1.2), see e.g. [4, 9, 15, 27].

2.2 Variational Inequality Characterization of (1.4)

Then, we rewrite the algorithmic framework (1.4) also in the VI form. Note that the PDHG subroutine (1.4a) can be written as

$$\begin{cases} \tilde{x}^k = \arg\min\{\theta_1(x) - (y^k)^T(Ax-b) + \frac{r}{2}\|x-x^k\|^2 \mid x \in \mathcal{X}\}, & (2.5a) \\ \tilde{y}^k = \arg\max\{\theta_2(y) + y^T(A\bar{x}^k - b) + \frac{s}{2}\|y-y^k\|^2 \mid y \in \mathcal{Y}\}. & (2.5b) \end{cases}$$

Thus, the optimality conditions of (2.5a) and (2.5b) are

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x-\tilde{x}^k)^T\big(-A^Ty^k + r(\tilde{x}^k - x^k)\big) \ge 0, \;\; \forall\, x \in \mathcal{X},$$

and

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y-\tilde{y}^k)^T\big((A\bar{x}^k - b) + s(\tilde{y}^k - y^k)\big) \ge 0, \;\; \forall\, y \in \mathcal{Y},$$

respectively. Recall

$$\bar{x}^k = \tilde{x}^k + \tau(\tilde{x}^k - x^k).$$

We get the following VI containing only (x^k, y^k) and $(\tilde{x}^k, \tilde{y}^k)$:

$$\begin{aligned} &(\tilde{x}^k, \tilde{y}^k) \in \mathcal{X} \times \mathcal{Y}, \quad (\theta_1(x) - \theta_1(\tilde{x}^k)) + (\theta_2(y) - \theta_2(\tilde{y}^k)) \\ &+ \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{y}^k \\ A\tilde{x}^k - b \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) + A^T(\tilde{y}^k - y^k) \\ \tau A(\tilde{x}^k - x^k) + s(\tilde{y}^k - y^k) \end{pmatrix} \right\} \ge 0, \; \forall (x,y) \in \mathcal{X} \times \mathcal{Y}. \quad (2.6) \end{aligned}$$

Using the notations in (2.2b), the PDHG subroutine (1.4a) can be written compactly as follows.

PDHG subroutine Step (Prediction):

$$\tilde{u}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T\{F(\tilde{u}^k) + Q(\tilde{u}^k - u^k)\} \ge 0, \quad \forall\, u \in \Omega, \qquad (2.7a)$$

where

$$Q = \begin{pmatrix} rI_n & A^T \\ \tau A & sI_m \end{pmatrix} \qquad (2.7b)$$

Accordingly, the correction step (1.4b) can be rewritten as the following compact form.

Correction Step

$$u^{k+1} = u^k - M(u^k - \tilde{u}^k), \qquad (2.8a)$$

where

$$M = \begin{pmatrix} \alpha I_n & 0 \\ 0 & \beta I_m \end{pmatrix}. \qquad (2.8b)$$

3 Convergence

In this section, we investigate the convergence of the algorithmic framework (1.4). Recall that (1.4) can be rewritten as (2.7)-(2.8). We prove the convergence of (1.4) under the condition

$$\text{(Convergence Condition)} \qquad \begin{cases} H := QM^{-1} \succ 0, & (3.1a) \\ G := Q^T + Q - M^THM \succ 0, & (3.1b) \end{cases}$$

where the matrices Q and M are given in (2.7b) and (2.8b), respectively. Thus, it suffices to check the positive definiteness of two matrices to ensure the convergence of the algorithmic framework (1.4). The condition (3.1) also implies some specific restrictions onto the involved parameters in (1.4) to be discussed in Section 5.

Because of the VI reformulation (2.2) of the model (1.1), we are interested in using the VI characterization (2.7)-(2.8) to discern how accurate an iterate generated by the algorithmic framework (1.4) is to a solution point of $\mathrm{VI}(\Omega, F, \theta)$. This analysis is summarized in the following theorem.

Theorem 3.1. *Let $\{u^k\}$ be the sequence generated by the algorithmic framework (1.4). Under the condition* (3.1)*, we have*

$$\theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T F(\tilde{u}^k) \geq \frac{1}{2}\big(\|u - u^{k+1}\|_H^2 - \|u - u^k\|_H^2\big) + \frac{1}{2}\|u^k - \tilde{u}^k\|_G^2, \quad \forall u \in \Omega. \tag{3.2}$$

Proof. It follows from (3.1a) that $Q = HM$. Recall that $M(u^k - \tilde{u}^k) = (u^k - u^{k+1})$ in (2.8a). Thus, (2.7a) can be written as

$$\theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T F(\tilde{u}^k) \geq (u - \tilde{u}^k)^T H(u^k - u^{k+1}), \ \forall u \in \Omega. \tag{3.3}$$

Applying the identity

$$(a - b)^T H(c - d) = \frac{1}{2}\{\|a - d\|_H^2 - \|a - c\|_H^2\} + \frac{1}{2}\{\|c - b\|_H^2 - \|d - b\|_H^2\}$$

to the right-hand side of (3.3) with

$$a = u, \quad b = \tilde{u}^k, \quad c = u^k, \quad \text{and} \quad d = u^{k+1},$$

we obtain

$$(u - \tilde{u}^k)^T H(u^k - u^{k+1}) = \frac{1}{2}\big(\|u - u^{k+1}\|_H^2 - \|u - u^k\|_H^2\big) + \frac{1}{2}\big(\|u^k - \tilde{u}^k\|_H^2 - \|u^{k+1} - \tilde{u}^k\|_H^2\big). \tag{3.4}$$

For the last term of the right-hand side of (3.4), we have

$$\begin{aligned}
&\|u^k - \tilde{u}^k\|_H^2 - \|u^{k+1} - \tilde{u}^k\|_H^2 \\
&\quad = \quad \|u^k - \tilde{u}^k\|_H^2 - \|(u^k - \tilde{u}^k) - (u^k - u^{k+1})\|_H^2 \\
&\quad \overset{(3.1a)}{=} \quad \|u^k - \tilde{u}^k\|_H^2 - \|(u^k - \tilde{u}^k) - M(u^k - \tilde{u}^k)\|_H^2 \\
&\quad = \quad 2(u^k - \tilde{u}^k)^T HM(u^k - \tilde{u}^k) - (u^k - \tilde{u}^k)^T M^T HM(u^k - \tilde{u}^k) \\
&\quad = \quad (u^k - \tilde{u}^k)^T (Q^T + Q - M^T HM)(u^k - \tilde{u}^k) \\
&\quad \overset{(3.1b)}{=} \quad \|u^k - \tilde{u}^k\|_G^2.
\end{aligned} \tag{3.5}$$

Substituting (3.4) and (3.5) into (3.3), we obtain the assertion (3.2). □

With the assertion in Theorem 3.1, we can show that the sequence $\{u^k\}$ generated by the algorithmic framework (1.4) is strictly contractive with respect to the solution set Ω^*. We prove this property in the following theorem.

Theorem 3.2. *Let $\{u^k\}$ be the sequence generated by the algorithmic framework (1.4). Under the condition* (3.1)*, we have*

$$\|u^{k+1} - u^*\|_H^2 \leq \|u^k - u^*\|_H^2 - \|u^k - \tilde{u}^k\|_G^2, \quad \forall u^* \in \Omega^*. \tag{3.6}$$

Proof. Setting $u = u^*$ in (3.2), we get

$$\|u^k - u^*\|_H^2 - \|u^{k+1} - u^*\|_H^2 \geq \|u^k - \tilde{u}^k\|_G^2 + 2\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{u}^k - u^*)^T F(\tilde{u}^k)\}. \quad (3.7)$$

Then, using (2.3) and the optimality of u^*, we have

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{u}^k - u^*)^T F(\tilde{u}^k) = \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{u}^k - u^*)^T F(u^*) \geq 0$$

and thus

$$\|u^k - u^*\|_H^2 - \|u^{k+1} - u^*\|_H^2 \geq \|u^k - \tilde{u}^k\|_G^2. \quad (3.8)$$

The assertion (3.6) follows directly. □

Theorem 3.2 shows that the sequence $\{u^k\}$ is Fèjer monotone and the convergence of $\{u^k\}$ to a $u^* \in \Omega^*$ in H-norm is immediately implied, see, e.g., [2].

4 Convergence Rate

In this section, we establish the worst-case $O(1/t)$ convergence rate measured by the iteration complexity for the algorithmic framework (1.4), where t is the iteration counter. We discuss the convergence rates in both the ergodic and nonergodic senses.

4.1 Convergence Rate in the Ergodic Sense

First, let us characterize an approximate solution of VI(Ω, F, θ) (2.2). Because of (2.3), we have

$$(u - u^*)^T F(u^*) = (u - u^*)^T F(u).$$

Thus, (2.2a) can be rewritten as

$$u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u - u^*)^T F(u) \geq 0, \quad \forall\, u \in \Omega. \quad (4.1)$$

For given $\epsilon > 0$, $\tilde{u} \in \Omega$ is called an ϵ-approximate solution of VI(Ω, F, θ) if it satisfies

$$\theta(u) - \theta(\tilde{u}) + (u - \tilde{u})^T F(u) \geq -\epsilon, \;\; \forall\; u \in \mathcal{D}_{(\tilde{u})}, \quad (4.2)$$

where

$$\mathcal{D}_{(\tilde{u})} := \{u \in \Omega \,|\, \|u - \tilde{u}\| \leq 1\}.$$

We refer the reader to [18] (see (2.5) therein) for the rational of defining an approximate solution onto the set $\mathcal{D}_{(\tilde{u})}$.

In the following we show that for given $\epsilon > 0$, based on t iterations of the algorithmic framework (1.4), we can find $\tilde{u} \in \Omega$ such that

$$\tilde{u} \in \Omega \quad \text{and} \quad \sup_{u \in \mathcal{D}_{(\tilde{u})}} \big\{\theta(\tilde{u}) - \theta(u) + (\tilde{u} - u)^T F(u)\big\} \leq \epsilon, \quad (4.3)$$

i.e., an ϵ-approximate solution of VI(Ω, F, θ). Theorem 3.1 will be used in the analysis; similar techniques can be found in [16].

Theorem 4.1. *Let $\{u^k\}$ be the sequence generated by the algorithmic framework (1.4) under the conditions* (1.3) *and* (3.1)*. For any integer $t > 0$, let $\tilde{u}_t$ be defined as*

$$\tilde{u}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{u}^k. \quad (4.4)$$

Then we have

$$\tilde{u}_t \in \Omega, \quad \theta(\tilde{u}_t) - \theta(u) + (\tilde{u}_t - u)^T F(u) \leq \frac{1}{2(t+1)} \|u - u^0\|_H^2, \quad \forall u \in \Omega. \quad (4.5)$$

Proof. First, we know $\tilde{u}^k \in \Omega$ for all $k \geq 0$. Because of the convexity of $\mathcal{X}$ and $\mathcal{Y}$, it follows from (4.4) that $\tilde{u}_t \in \Omega$. Using (2.3) and (3.2), we have

$$\theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T F(u) + \frac{1}{2}\|u - u^k\|_H^2 \geq \frac{1}{2}\|u - u^{k+1}\|_H^2, \ \forall u \in \Omega. \tag{4.6}$$

Summarizing the inequality (4.6) over $k = 0, 1, \ldots, t$, we obtain

$$(t+1)\theta(u) - \sum_{k=0}^{t} \theta(\tilde{u}^k) + \Big((t+1)u - \sum_{k=0}^{t} \tilde{u}^k\Big)^T F(u) + \frac{1}{2}\|u - u^0\|_H^2 \geq 0, \quad \forall u \in \Omega.$$

Since $\tilde{u}_t$ is defined in (4.4), we have

$$\frac{1}{t+1}\sum_{k=0}^{t} \theta(\tilde{u}^k) - \theta(u) + (\tilde{u}_t - u)^T F(u) \leq \frac{1}{2(t+1)}\|u - u^0\|_H^2, \quad \forall u \in \Omega. \tag{4.7}$$

Further, because of the convexity of $\theta(u)$, we get

$$\theta(\tilde{u}_t) \leq \frac{1}{t+1}\sum_{k=0}^{t} \theta(\tilde{u}^k).$$

Substituting it into (4.7), the assertion of this theorem follows directly. $\Box$

The conclusion (4.5) indicates that $\tilde{u}_t$ defined in (4.4) is an approximate solution of $\mathrm{VI}(\Omega, F, \theta)$ with an accuracy of $O(1/t)$. Since $\tilde{u}_t$ is given by the average of t iterations of (1.4), the worst-case $O(1/t)$ convergence rate in the ergodic sense is established for the the algorithmic framework (1.4). Finally, we would remark that the condition of $G \succ 0$ in (3.1) can be relaxed to $G \succeq 0$ for deriving the convergence rate result in this subsection.

4.2 Convergence Rate in a Nonergodic Sense

In this subsection, we derive the worst-case $O(1/t)$ convergence rate in a nonergodic sense for the algorithmic framework (1.4). Note that a nonergodic worst-case convergence rate is generally stronger than its ergodic counterparts. As mentioned, it seems still open whether or not the PDHG scheme (1.2) has any nonergodic worst-case convergence rate. We first prove a lemma.

Lemma 4.2. *Let $\{u^k\}$ be the sequence generated by the algorithmic framework (1.4). Under the condition* (3.1)*, we have*

$$(u^k - \tilde{u}^k)^T M^T H M\{(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\} \geq \frac{1}{2}\|(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\|_{(Q^T+Q)}^2. \tag{4.8}$$

Proof. Setting $u = \tilde{u}^{k+1}$ in (2.7a), we get

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{u}^{k+1} - \tilde{u}^k)^T F(\tilde{u}^k) \geq (\tilde{u}^{k+1} - \tilde{u}^k)^T Q(u^k - \tilde{u}^k). \tag{4.9}$$

Note that (2.7a) is also true for $k := k+1$. Thus, it holds that

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (u - \tilde{u}^{k+1})^T F(\tilde{u}^{k+1}) \geq (u - \tilde{u}^{k+1})^T Q(u^{k+1} - \tilde{u}^{k+1}), \forall u \in \Omega.$$

Then, setting $u = \tilde{u}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{u}^k - \tilde{u}^{k+1})^T F(\tilde{u}^{k+1}) \geq (\tilde{u}^k - \tilde{u}^{k+1}) Q(u^{k+1} - \tilde{u}^{k+1}). \tag{4.10}$$

Combining (4.9) and (4.10), and using the monotonicity of F, we have

$$(\tilde{u}^k - \tilde{u}^{k+1})^T Q\{(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\} \geq 0. \tag{4.11}$$

Adding the term

$$\{(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\}^T Q\{(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\}$$

to both sides of (4.11), and using $v^T Q v = \frac{1}{2} v^T (Q^T + Q) v$, we obtain

$$(u^k - u^{k+1})^T Q\{(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\} \geq \frac{1}{2}\|(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\|^2_{(Q^T+Q)}.$$

Substituting $(u^k - u^{k+1}) = M(u^k - \tilde{u}^k)$ into the left-hand side of the last inequality and using $Q = HM$, we obtain (4.8) and the lemma is proved. □

Now, we establish the worst-case $O(1/t)$ convergence rate in a nonergodic sense for the algorithmic framework (1.4). Similar techniques can be found in [17].

Theorem 4.3. *Let $\{u^k\}$ be the sequence generated by the algorithmic framework (1.4). Under the condition* (3.1)*, then we have*

$$\|M(u^t - \tilde{u}^t)\|^2_H \leq \frac{1}{(t+1)c_0}\|u^0 - u^*\|^2_H, \tag{4.12}$$

where $c_0 > 0$ is a constant.

Proof. First, setting $a = M(u^k - \tilde{u}^k)$ and $b = M(u^{k+1} - \tilde{u}^{k+1})$ in the identity

$$\|a\|^2_H - \|b\|^2_H = 2a^T H(a-b) - \|a-b\|^2_H,$$

we obtain

$$\begin{aligned} &\|M(u^k - \tilde{u}^k)\|^2_H - \|M(u^{k+1} - \tilde{u}^{k+1})\|^2_H \\ &= \; 2(u^k - \tilde{u}^k)^T M^T H M[(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})] - \|M[(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})]\|^2_H. \end{aligned}$$

Inserting (4.8) into the first term of the right-hand side of the last equality, we obtain

$$\begin{aligned} &\|M(u^k - \tilde{u}^k)\|^2_H - \|M(u^{k+1} - \tilde{u}^{k+1})\|^2_H \\ &\geq \; \|(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\|^2_{(Q^T+Q)} - \|M[(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})]\|^2_H \\ &= \; \|(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\|^2_G \geq 0. \end{aligned}$$

The last inequality holds because the matrix $(Q^T + Q) - M^T H M = G$ and $G \succeq 0$. We thus have

$$\|M(u^{k+1} - \tilde{u}^{k+1})\|_H \leq \|M(u^k - \tilde{u}^k)\|_H, \quad \forall\, k > 0. \tag{4.13}$$

Recall that the matric M is positive definite. Thus, the sequence $\{\|M(u^k - \tilde{u}^k)\|^2_H\}$ is monotonically non-increasing. Then, it follows from $G \succ 0$ and Theorem 3.2 there is a constant $c_0 > 0$ such that

$$\|u^{k+1} - u^*\|^2_H \leq \|u^k - u^*\|^2_H - c_0\|M(u^k - \tilde{u}^k)\|^2_H, \quad \forall u^* \in \Omega^*. \tag{4.14}$$

Furthermore, it follows from (4.14) that

$$\sum_{k=0}^{\infty} c_0\|M(u^k - \tilde{u}^k)\|^2_H \leq \|u^0 - u^*\|^2_H, \quad \forall\, u^* \in \Omega^*. \tag{4.15}$$

Therefore, we have

$$(t+1)\|M(u^t - \tilde{u}^t)\|_H^2 \le \sum_{k=0}^{t} \|M(u^k - \tilde{u}^k)\|_H^2. \tag{4.16}$$

The assertion (4.12) follows from (4.15) and (4.16) immediately. □

Let $d := \inf\{\|u^0 - u^*\|_H \mid u^* \in \Omega^*\}$. Then, for any given $\epsilon > 0$, Theorem 4.3 shows that the algorithmic framework (1.4) needs at most $\lfloor d^2/c_0\epsilon \rfloor$ iterations to ensure that $\|M(u^k - \tilde{u}^k)\|_H^2 \le \epsilon$. Recall that u^k is a solution of $\mathrm{VI}(\Omega, F, \theta)$ if $\|M(u^k - \tilde{u}^k)\|_H^2 = 0$ (see (2.7a) and due to $Q = HM$). A worst-case $O(1/t)$ convergence rate in a nonergodic sense is thus established for the algorithmic framework (1.4).

5 How to Ensure the Condition (3.1)

In previous sections, we have analyzed the convergence for the algorithmic framework (1.4) under the condition (3.1). Now, we discuss how to appropriately choose the parameters τ, α and β to ensure the condition (3.1). Different choices of these parameters also specify the algorithmic framework (1.4) as a class of specific PDHG schemes for the saddle point model (1.1).

5.1 General Study

Recall the definitions of Q and M in (2.7b) and (2.8b), respectively. Let us take a closer look at the matrices H and G defined in (3.1).

First, we have

$$H = QM^{-1} = \begin{pmatrix} rI_n & A^T \\ \tau A & sI_m \end{pmatrix} \begin{pmatrix} \frac{1}{\alpha} I_n & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} = \begin{pmatrix} \frac{r}{\alpha} I_n & \frac{1}{\beta} A^T \\ \frac{\tau}{\alpha} A & \frac{1}{\beta} s I_m \end{pmatrix}.$$

Thus, we set $\beta = \frac{\alpha}{\tau}$ to ensure that H is symmetric. With this restriction, we have

$$H = \begin{pmatrix} \frac{r}{\alpha} I_n & \frac{\tau}{\alpha} A^T \\ \frac{\tau}{\alpha} A & \frac{\tau}{\alpha} s I_m \end{pmatrix} = \frac{1}{\alpha} \begin{pmatrix} rI_n & \tau A^T \\ \tau A & \tau s I_m \end{pmatrix}.$$

Then, it is clear that under the condition (1.3), we can choose $\tau \in (0, 1]$ to ensure the positive definiteness of H. Overall, to ensure the positive definiteness and symmetry of H, we pose the restriction:

$$\tau \in (0, 1] \qquad \text{and} \qquad \beta = \frac{\alpha}{\tau}. \tag{5.1}$$

Second, let us deal with the matric G in (3.1). Since $HM = Q$, we have

$$\begin{aligned} Q^T + Q - M^T H M &= Q^T + Q - M^T Q \\ &= \begin{pmatrix} 2rI_n & (1+\tau)A^T \\ (1+\tau)A & 2sI_m \end{pmatrix} - \begin{pmatrix} \alpha I_n & 0 \\ 0 & \frac{\alpha}{\tau} I_m \end{pmatrix} \begin{pmatrix} rI_n & A^T \\ \tau A & sI_m \end{pmatrix} \\ &= \begin{pmatrix} 2rI_n & (1+\tau)A^T \\ (1+\tau)A & 2sI_m \end{pmatrix} - \begin{pmatrix} \alpha r I_n & \alpha A^T \\ \alpha A & \frac{\alpha}{\tau} s I_m \end{pmatrix} \\ &= \begin{pmatrix} (2-\alpha) r I_n & (1+\tau-\alpha) A^T \\ (1+\tau-\alpha) A & (2 - \frac{\alpha}{\tau}) s I_m \end{pmatrix}. \end{aligned} \tag{5.2}$$

Therefore, we need to ensure

$$G = \begin{pmatrix} (2-\alpha)rI_n & (1+\tau-\alpha)A^T \\ (1+\tau-\alpha)A & (2-\frac{\alpha}{\tau})sI_m \end{pmatrix} \succ 0. \tag{5.3}$$

We present a strategy to ensure (5.3) in the following theorem.

Theorem 5.1. *For given $\tau \in (0,1]$, under the condition (1.3), the assertion (5.3) holds when*

$$\tau = 1 \qquad \text{and} \qquad \alpha \in (0,2), \tag{5.4}$$

or

$$\tau \in (0,1), \qquad \text{and} \qquad \alpha \le (1+\tau) - \sqrt{1-\tau}. \tag{5.5}$$

Proof. First, if $\tau = 1$, it follows from (5.2) that

$$G = (2-\alpha)\begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}.$$

Under the condition (1.3), we have $G \succ 0$ for all $\alpha \in (0,2)$.

Now, we consider $\tau \in (0,1)$. For $G \succ 0$, we need to ensure

$$(2-\alpha) > 0, \qquad (2-\frac{\alpha}{\tau}) > 0 \qquad \text{and} \qquad (2-\alpha)(2-\frac{\alpha}{\tau})rs > (1+\tau-\alpha)^2\|A^TA\|.$$

Thus, under the condition (1.3), we need to guarantee

$$(2-\alpha) > 0, \qquad (2-\frac{\alpha}{\tau}) > 0, \tag{5.6a}$$

and

$$(2-\alpha)(2-\frac{\alpha}{\tau}) \ge (1+\tau-\alpha)^2. \tag{5.6b}$$

By a manipulation, the inequality (5.6b) is equivalent to

$$\Big(\frac{1}{\tau}-1\Big)\alpha^2 + 2\Big((1+\tau)-(1+\frac{1}{\tau})\Big)\alpha + \big(4-(1+\tau)^2\big) \ge 0,$$

and thus

$$\Big(\frac{1-\tau}{\tau}\Big)\alpha^2 + 2\Big(\frac{\tau^2-1}{\tau}\Big)\alpha + (3+\tau)(1-\tau) \ge 0.$$

Multiplying the positive factor $\dfrac{\tau}{1-\tau}$, we get

$$\alpha^2 - 2(1+\tau)\alpha + \tau(3+\tau) \ge 0.$$

Because the small root of the equation

$$\alpha^2 - 2(1+\tau)\alpha + \tau(3+\tau) = 0$$

is $(1+\tau)-\sqrt{1-\tau}$, the condition (5.6b) is satisfied when (5.5) holds.

In the following we show that (5.6a) is satisfied when (5.5) holds. It follows from (5.5) directly $\alpha < 2$ for all $\tau \in (0,1)$. In addition, because

$$(1-\tau)-\sqrt{1-\tau} < 0, \ \forall\, \tau \in (0,1),$$

we have

$$(1+\tau) - \sqrt{1-\tau} < 2\tau, \quad \forall \tau \in (0,1).$$

Consequently, using (5.5), we get

$$\frac{\alpha}{\tau} \leq \frac{(1+\tau) - \sqrt{1-\tau}}{\tau} < 2, \quad \forall \tau \in (0,1).$$

The proof is complete. $\Box$

In the following theorem, we summarize the restriction onto the involved parameters that can ensure the convergence of the algorithmic framework (1.4).

Theorem 5.2. *Under the condition* (1.3)*, if the parameters* τ, α *and* β *in* (1.4) *are chosen such that either*

$$\tau = 1, \qquad \text{and} \qquad \alpha = \beta \in (0,2), \tag{5.7a}$$

or

$$\tau \in (0,1), \qquad \alpha \leq 1 + \tau - \sqrt{1-t} \qquad \text{and} \qquad \beta = \frac{\alpha}{\tau}, \tag{5.7b}$$

then the matrices Q and M defined respectively in (2.7b) *and* (2.8b) *satisfy the convergence condition* (3.1)*. Thus, the sequence generated by the algorithmic framework (1.4) is convergent to a solution point of* VI(Ω, F, θ)*.*

5.2 Special Cases

In this subsection, we specify the general restriction posed in Theorem 5.2 on the parameters of the algorithmic framework (1.4) and discuss some special cases of this framework.

5.2.1 Case I: $\tau = 1$

According to (5.7a), we can take $\alpha = \beta \in (0,2)$ and thus the algorithmic framework (1.4) can be written as

$$\left\{ \begin{array}{l} \left\{ \begin{array}{rcl} \tilde{x}^k &=& \arg\min\{\Phi(x,y^k) + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, \\ \bar{x}^k &=& 2\tilde{x}^k - x^k, \\ \tilde{y}^k &=& \arg\max\{\Phi(\bar{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \end{array} \right. \quad (5.8a) \\ \begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ y^k \end{pmatrix} - \alpha \begin{pmatrix} x^k - \tilde{x}^k \\ y^k - \tilde{y}^k \end{pmatrix}. \quad (5.8b) \end{array} \right.$$

This is exactly Algorithm 4 in [15] which can be regarded as an accelerated version of (1.2) using the technique in [13]. Clearly, choosing $\alpha = \beta = 1$ in (5.8) recovers the special case of (1.2) with $\tau = 1$.

5.2.2 Case II: $\alpha = 1$

Now let $\alpha = 1$ be fixed. In this case, the algorithmic framework (1.4) can be written as

$$\left\{ \begin{array}{rcl} x^{k+1} &=& \arg\min\{\Phi(x,y^k) + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, \\ \bar{x}^k &=& x^{k+1} + \tau(x^{k+1} - x^k), \\ \tilde{y}^k &=& \arg\max\{\Phi(\bar{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}, \\ y^{k+1} &=& y^k + \beta(\tilde{y}^k - y^k), \end{array} \right. \tag{5.9}$$

which is symmetric in the sense that it updates both the variables x and y twice at each iteration. For this special case, the restriction in Theorem 5.2 can be further specified. We present it in the following theorem.

Theorem 5.3. *If $\alpha = 1$; the parameters β and τ satisfy*

$$\tau \in \Big[\frac{\sqrt{5}-1}{2}, 1\Big], \qquad \textit{and} \qquad \beta = \frac{1}{\tau}, \tag{5.10}$$

then the conditions (5.7a) *or* (5.7b) *is satisfied and thus the PDHG scheme (5.9) is convergent.*

Proof. If $\alpha = 1$ and $\tau = 1$, it follows from (5.10) that the condition (5.7a) is satisfied. Now we consider $\alpha = 1$ and $\tau \in (0, 1)$. In this case the condition (5.7b) becomes

$$\tau \in (0,1) \qquad \text{and} \qquad 1 \le 1 + \tau - \sqrt{1-\tau},$$

and thus

$$\tau \in (0,1) \qquad \text{and} \qquad \tau^2 + \tau - 1 \ge 0.$$

The above condition is satisfied for all $\tau \in \big[\frac{\sqrt{5}-1}{2}, 1\big)$. The proof is complete. $\Box$

Therefore, with the restriction in (5.10), the PDHG scheme (5.9) can be written as

$$\left\{\begin{array}{rcl} x^{k+1} &=& \arg\min\{\Phi(x,y^k) + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, \\ \bar{x}^k &=& x^{k+1} + \tau(x^{k+1} - x^k), \\ \tilde{y}^k &=& \arg\max\{\Phi(\bar{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}, \\ y^{k+1} &=& y^k + \frac{1}{\tau}(\tilde{y}^k - y^k), \end{array}\right. \tag{5.11}$$

where $\tau \in \Big[\frac{\sqrt{5}-1}{2}, 1\Big]$. Indeed, we can further consider choosing

$$\tau_0 = \frac{\sqrt{5}-1}{2}.$$

Then, we have

$$\frac{1}{\tau_0} = 1 + \tau_0,$$

and it follows from (5.11) that

$$y^k + \frac{1}{\tau_0}(\tilde{y}^k - y^k) = y^k + (1+\tau_0)(\tilde{y}^k - y^k) = \tilde{y}^k + \tau_0(\tilde{y}^k - y^k).$$

Thus, the scheme (5.11) can be further specified as

$$\left\{\begin{array}{rcl} x^{k+1} &=& \arg\min\{\Phi(x,y^k) + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, \\ \bar{x}^k &=& x^{k+1} + \tau_0(x^{k+1} - x^k), \\ \tilde{y}^k &=& \arg\max\{\Phi(\bar{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}, \\ y^{k+1} &=& \tilde{y}^k + \tau_0(\tilde{y}^k - y^k), \end{array}\right. \tag{5.12}$$

where $\tau_0 = \frac{\sqrt{5}-1}{2}$. It is a completely symmetric PDHG scheme in the sense that both the variables are updated twice at each iteration and the additional update steps both use the golden-ratio step sizes.

5.2.3 Case III: $\beta = 1$

We can also fix $\beta = 1$. In this case, the algorithmic framework (1.4) reduces to

$$\left\{ \begin{array}{l} \left\{ \begin{array}{rcl} \tilde{x}^k &=& \arg\min\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, \\ \bar{x}^k &=& \tilde{x}^k + \tau(\tilde{x}^k - x^k), \\ y^{k+1} &=& \arg\max\{\Phi(\bar{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}. \end{array} \right. \quad (5.13a) \\ x^{k+1} \;=\; x^k - \alpha(x^k - \tilde{x}^k). \quad (5.13b) \end{array} \right.$$

For this case, the restriction on the involved parameters in Theorem 5.2 can be further specified as the following theorem.

Theorem 5.4. *If $\beta = 1$; the parameters α and τ satisfy*

$$\alpha = \tau \in (0, 1], \tag{5.14}$$

then the convergence conditions (5.3) *are satisfied.*

Proof. First, in this case, the condition (5.1) is satisfied. The matrix G in (5.3) becomes

$$G = \begin{pmatrix} (2-\alpha)rI_n & A^T \\ A & sI_m \end{pmatrix},$$

which is positive definite for all $\alpha \in (0, 1]$ under the assumption (1.3). □

Therefore, considering the restriction in (5.14), we can specify the scheme (5.13) as

$$\left\{ \begin{array}{l} \left\{ \begin{array}{rcl} \tilde{x}^k &=& \arg\min\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, \\ \bar{x}^k &=& \tilde{x}^k + \tau(\tilde{x}^k - x^k), \\ y^{k+1} &=& \arg\max\{\Phi(\bar{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}. \end{array} \right. \quad (5.15a) \\ x^{k+1} \;=\; x^k - \tau(x^k - \tilde{x}^k). \quad (5.15b) \end{array} \right.$$

where $\tau \in \big(0, 1\big]$.

6 Conclusions

We propose an algorithmic framework of generalized primal-dual hybrid gradient (PDHG) methods for saddle point problems and study its convergence. This algorithmic framework includes some existing PDHG schemes as special cases, and it can yield a class of new generalized PDHG schemes. It also possesses some theoretical advantages such as the worst-case convergence rate measured by the iteration complexity in a nonergodic sense. Our analysis provides a unified perspective to the study of some PDHG schemes for saddle point problems. It is interesting to know if our analysis could be extended to some special nonconvex cases with strong application backgrounds in image processing such as some semiconvex variational image restoration models. We leave it as a further research topic.

References

[1] K. J. Arrow, L. Hurwicz and H. Uzawa, *Studies in linear and non-linear programming*, With contributions by H. B. Chenery, S. M. Johnson, S. Karlin, T. Marschak, R. M. Solow. Stanford Mathematical Studies in the Social Science, Vol. II. Stanford University Press, Stanford, Calif., 1958.

[2] E. Blum and W. Oettli, Mathematische Optimierung *Grundlagen und Verfahren. Ökonometrie und Unternehmensforschung*, Springer-Verlag, Berlin-Heidelberg-New York, 1975.

[3] S. Bonettini and V. Ruggiero, *On the convergence of primal-dual hybrid gradient algorithms for total variation image restoration*, J. Math. Imaging Vision, 44 (2012), pp. 236-253.

[4] A. Chambolle and T. Pock, *A first-order primal-dual algorithms for convex problem with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120-145.

[5] A. Chambolle and T. Pock, *On the ergodic convergence rates of a first-order primal-dual algorithm*, Math. Program., to appear.

[6] A. Chambolle and T. Pock, *A remark on accelerated block coordinate descent for computing the proximity operators of a sum of convex functions*, SMAI J. Comput. Math., to appear.

[7] L. Condat, *A primaldual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appli., 158 (2013), pp. 460-479.

[8] J. Douglas and H. H. Rachford, *On the numerical solution of the heat conduction problem in 2 and 3 space variables*, Trans. Amer. Math. Soc., 82(1956), pp. 421-439.

[9] E. Esser, X. Zhang and T. F. Chan, *A general framework for a class of first order primal-dual algorithms for TV minimization*, SIAM J. Imaging Sci., 3 (2010), pp. 1015-1046.

[10] F. Facchinei and J. S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, Vol. I. Springer Series in Operations Research, Springer Verlag, New York, 2003.

[11] R. Glowinski and A. Marrocco, *Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problèmes non linéaires*, R.A.I.R.O., R2, 1975, pp. 41-76.

[12] T. Goldstein, M. Li, X. M. Yuan, E. Esser and R. Baraniuk, *Adaptive primal-dual hybrid gradient methods for saddle-point problems*, NIPS, 2015.

[13] E. G. Gol'shtein and N. V. Tret'yakov, *Modified Lagrangian in convex programming and their generalizations*, Math. Program. Study, 10(1979), pp. 86-97.

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

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

[16] B. S. He and X. M. Yuan, *On the $O(1/n)$ convergence rate of Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700-709.

[17] B. S. He and X.M. Yuan, *On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers*, Numerische Mathematik, 130 (2015), pp. 567-577.

[18] Y. E. Nesterov, *Gradient methods for minimizing composite objective function*, Math. Prog., Ser. B, 140 (2013), pp. 125-161.

[19] G. M. Korpelevich, *The extragradient method for finding saddle points and other problems*, Ekonomika i Matematchskie Metody, 12 (1976), pp. 747-756.

[20] P. L. Lions, B. Mercier, *splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16(1979), pp. 964-979.

[21] B. Martinet, *Regularization d'inequations variationelles par approximations successives*, Revue Francaise d'Informatique et de Recherche Opérationelle, 4(1970), pp. 154-159.

[22] T. Pock and A. Chambolle, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, IEEE International Conference on Computer Vision, 2011, pp. 1762-1769.

[23] L. D. Popov, *A modification of the Arrow-Hurwitz method of search for saddle points*, Mat. Zametki, 28(5) (1980), pp. 777-784.

[24] L. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 227-238.

[25] R. Shefi, *Rate of convergence analysis for convex nonsmooth optimization algorithms*, PhD Thesis, Tel Aviv University, Israel, 2015.

[26] P. Weiss, L. Blanc-Feraud and G. Aubert, *Efficient schemes for total variation minimization under constraints in image processing*, SIAM J. Sci. Comput., 31(3) (2009), pp. 2047-2080.

[27] X. Zhang, M. Burger and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (1) (2010), pp. 20-46.

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