A RELAXED PARAMETER CONDITION FOR THE PRIMAL–DUAL HYBRID GRADIENT METHOD FOR SADDLE-POINT PROBLEM

XIAYANG ZHANG
Department of Mathematics and Physics, Nanjing Institute of Technology
Nanjing, 211167, China

YUQIAN KONG
School of Applied Mathematics, Nanjing University of Finance & Economics
Nanjing, 210023, China

SHANSHAN LIU
Hefei 168 Middle School, Hefei, 230031, China

YUAN SHEN*
School of Applied Mathematics, Nanjing University of Finance & Economics
Nanjing, 210023, China

(Communicated by Paulo J. S. Silva)

Abstract. The primal-dual hybrid gradient method and the primal-dual algorithm proposed by Chambolle and Pock are both efficient methods for solving saddle point problem. However, the convergence of both methods depends on some assumptions which can be too restrictive or impractical in real applications. In this paper, we propose a new parameter condition for the primal-dual hybrid gradient method. This improvement only requires either the primal or the dual objective function to be strongly convex. The relaxed parameter condition leads to convergence acceleration. Although counter-example shows that the PDHG method is not necessarily convergent with constant step size, it becomes convergent with our relaxed parameter condition. Preliminary experimental results show that PDHG method with our relaxed parameter condition is more efficient than several state-of-art methods.

1. Introduction. In this paper, we consider solving the saddle-point problem as follows:

$$
\min_{x \in X} \max_{y \in Y} \Phi(x, y) := g(x) + y^T Ax - f^*(y),
$$

where \( f : \mathbb{R}^n \to \mathbb{R} \) is a proper, closed convex function, \( f^* : \mathbb{R}^m \to \mathbb{R} \) is the Fenchel conjugate of \( f \), \( g : \mathbb{R}^n \to \mathbb{R} \) is a strongly convex function w.r.t. \( x \) with parameter \( \alpha > 0 \), and \( A \in \mathbb{R}^{m \times n} \) is a given matrix.

2020 Mathematics Subject Classification. Primary: 90C30, 65K05; Secondary: 65K15.

Key words and phrases. Primal-dual hybrid gradient method, Chambolle-pock first-order method, saddle-point problem, parameter condition.

The first author is supported by the National Natural Science Foundation of China Grant 71571096 and Nanjing Institute of Technology Research Start-up Foundation Grant YKJ201738. The second author is supported by the National Social Science Foundation of China under Grants 19AZD018, 20BGL028, 19BGL205.

*Corresponding author: Yuan Shen.
The saddle-point problem (1.1) can be regarded as the primal-dual formulation of the following primal problem, e.g., see [9]:

$$\min_{x \in X} g(x) + f(Ax),$$

which arises in various applications such as image and signal processing, see e.g. [9, 16, 29, 32, 34, 36]. However, the non-smoothness of the objective (e.g., $g(x)$ can be a L1 regularizer) makes (1.1) computationally challenging, so there is considerable interest in designing efficient algorithms for solving (1.1).

For solving (1.1), Zhu and Chan proposed a primal-dual hybrid gradient (PDHG) algorithm [36] which treats the primal and dual variables separately. Some other primal-dual methods has been developed thereafter. The iterative schemes of some PDHG based methods can be unified as below:

$$\begin{cases}
x^{k+1} = \arg \min \left\{ \Phi(x, y^k) + \frac{1}{2\tau} \|x - x^k\|^2 \mid x \in X \right\}, \\
y^{k+1} = \arg \max \left\{ \Phi(x^{k+1} + \alpha(x^{k+1} - x^k), y) - \frac{1}{2\sigma} \|y - y^k\|^2 \mid y \in Y \right\},
\end{cases} \quad (1.2)$$

where $\tau$ and $\sigma$ are proximal parameters and $\alpha$ is a relaxation parameter. As its variables $x$ and $y$ are decoupled, we are able to take advantage of the special structure of $f$ and $g$, hence the subproblems of PDHG can be simple enough to have closed-form solution.

The PDHG method has its roots in the classical Arrow-Hurwicz method [1]. It is a first order method which only requires functional and gradient evaluations, hence it is suitable for large-scale problems. Taking the total-variation image denoising problem as an example, there have been a bunch of efficient algorithms for solving it, such as Chambolle’s method [8], the method of Chan, Golub and Mulet (CGM) [12], fast total variation denoising (FTVd) [31] and split Bregman [17]. The PDHG was reported to outperform the aforementioned algorithms in some numerical experiments with a wide variety of problem settings, e.g., see [2, 14, 34, 36].

When we take $\alpha = 0$ in (1.2), i.e., the extrapolation step $\alpha(x^{k+1} - x^k)$ vanishes, (1.2) reduces to PDHG method. Despite its satisfactory numerical performance, without additional assumptions, the convergence properties of the PDHG were unknown. Its convergence properties have been established by imposing additional restrictions such that the step sizes $\tau$ and $\sigma$ are small enough [14]. Furthermore, He, You and Yuan [19] showed that the PDHG is indeed convergent provided one of the objective functions is strongly convex w.r.t. $x$ (or $y$). Later in Section 3 we will show that the parameter condition in [19] is a special case of our parameter condition.

When we take $\alpha = 1$ in (1.2), we obtain the first order primal-dual algorithm proposed by Chambolle and Pock in [9]. In fact, the algorithm 1 in [9] is referred as the CP method in many literatures. The CP method has been proved to be convergent when $\alpha = 1$ with constant stepsize $\tau \sigma < 1/\|A^T A\|$ as in Algorithm 1 of [9]. Better rate can be achieved in algorithm 2 of [9] when either $f$ or $g$ is strongly convex with respect to the underlying variables. The acceleration technique in algorithm 2 adjust the parameters $\alpha \tau$ and $\sigma$ in each iteration for $\alpha_k = 1/\sqrt{1 + 2\tau_k}$, $\tau_{k+1} = \alpha_k \tau_k$, and $\sigma$. Such technique improves the convergence rate from $O(1/K)$ in algorithm 1 to $O(1/K^2)$.

In [14], it was shown that the CP method is related to the inexact Uzawa method [1] and its convergence was proved for certain variational imaging models with some specific step size sequences. In [20, 28], the CP method is shown as an application of the proximal point algorithm (PPA), and it is related to the classical Douglas-Rachford splitting method [25] under some special cases. Moreover, the relation between CP method and linearized ADMM was also clarified as shown in [11].
Goldstein et al. improved the CP method by taking a special adaptive step size which enables faster convergence. Incorporating the extragradient strategy, Bonettini and Ruggiero established the convergence of the PDHG method for nonsmooth convex optimization problems, and showed the convergence when the step length parameters are a priori selected sequences. Recently, some inertial variants of the CP method were proposed by introducing a relaxation/inertial step on both primal and dual variables at the end of each iteration, which enables better convergence rate and numerical behaviour, e.g., see [11, 30, 25].

Another way to improve the existing primal-dual methods is to add on a correction step. He and Yuan studied some existing primal-dual methods from the perspective of contraction perspective, simplified the existing convergence analysis, and showed that the domain of parameters (such as $\alpha$) can be significantly enlarged if some simple correction steps are applied [20]. Some other correction strategies were also proposed in [7, 18] for primal-dual methods. Recently, Zhang et al. proposed a simple primal-dual method in [35] for total-variation image restoration problem.

The satisfactory performances of primal-dual methods inspire us to improve their performance further. As illustrated in literatures (see e.g. [2, 9, 34, 36]), $\tau$ and $\sigma$ are the step sizes in its subproblems, and the numerical efficiency of PDHG could highly rely on their settings. In [2], the convergence of a variation of PDHG method was established with some asymptotical conditions on the step size sequence. In particular, the step size rules of PDHG method in [36] was well explained as they satisfy the conditions shown in [2, 14, 9, 20]. In general, the following restriction is imposed on the both PDHG and CP method in order to guarantee its convergence:

$$\tau \sigma < \frac{1}{\|A^T A\|}. \quad (1.3)$$

In fact, the parameter condition (1.3) is usually too conservative as the value of $\|A^T A\|$ can be large in practice. Even if (1.3) holds, the primal-dual method (1.2) is not necessarily convergent with constant stepsie if $\alpha \neq 1$, see appendix for a counter example. To overcome this issue, some efficient variations of PDHG with special variate step size strategy have been proposed, e.g., see [16], however, this step size rule might be too complicated in some certain applications.

The main contribution of our paper is proposing a new relaxed parameter condition for $(\tau, \sigma, \alpha)$, which gives us more flexibility in choosing parameters for the PDHG method. More specifically, to the best of our knowledge, the convergence results of PDHG with $\alpha \in (0, 1)$ and constant stepsize has not yet been established, while we derive the convergence properties for the PDHG method with $\alpha \in (0, 1)$ provided the proposed parameter condition holds. In addition, we achieve an "optimal" parameter condition such that the extension parameter $\alpha$ is chosen properly to maximize the domain of step sizes $\tau$ and $\sigma$. Preliminary experimental results show that the PDHG method with "optimal" parameter condition outperforms some state-of-art methods while solving LASSO and TV-l2 denoising problems.

The rest part of this paper is organized as follows. In Section 2, we summarize some useful preliminaries and notations. In Section 3, the proposed parameter conditions is shown. In Section 4 the contractive properties of the iterating sequence generated by PDHG method with the proposed conditions is studied under mild assumptions hence the convergence of PDHG method with relaxed parameter condition (PDHG-R) is proved. We carried out numerical experiments on LASSO and TV-l2 denoising problems in section 5. Finally, we draw some conclusions in Section 6.
2. Preliminaries. In this section, we summarize some basic concepts and preliminaries that will be used in the convergence analysis, and introduce some useful notations.

According to the first-order optimality condition, the saddle-point problem (1.1) can be recasted as the following variational inequality (VI): Find \( u^* \in \Omega \) such that

\[
\begin{aligned}
\text{VI}(\Omega, \theta, M) \quad u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u - u^*)^T M u^* \geq 0, \quad \forall u \in \Omega,
\end{aligned}
\]

where

\[
\begin{aligned}
u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = g(x) + f^*(y), \quad M = \begin{pmatrix} 0 & -A^T \\ A & 0 \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y}.
\end{aligned}
\]

(2.1a)

The solution set of \( \text{VI}(\Omega, \theta, M) \) is denoted by \( \Omega^* \), and it is always assumed to be nonempty.

Let \( \partial g(x) \) and \( \partial f^*(y) \) be the sub-differentials of \( g(x) \) and \( f^*(y) \), respectively. Then, \( \text{VI}(\Omega, F) \) can be reformulated as: Find \( u^* \in \Omega \), such that

\[
\begin{aligned}
\text{VI}(\Omega, F) : \quad (u - u^*)^T F(u^*) \geq 0, \quad \forall u \in \Omega,
\end{aligned}
\]

where

\[
\begin{aligned}
F(u) := \begin{pmatrix} \xi_x - A^T y \\ \eta_y + A x \end{pmatrix} \quad \text{with} \quad \xi_x \in \partial g(x) \quad \text{and} \quad \eta_y \in \partial f^*(y).
\end{aligned}
\]

(2.2a)

Note the mapping \( F \) is monotone due to the convexity of \( \theta(u) \) w.r.t. \( u \) and the skew-symmetricity of matrix \( M \).

Let \( \| \cdot \| \) denote the standard definition of \( l_p \)-norm, and particularly, and let \( \| \cdot \|_2 := \| \cdot \|_2 \) denote the Euclidean norm. For a positive definite matrix \( H \), we denote by \( \| \cdot \|_H \) the \( H \)-norm, i.e., \( \| x \|_H := \sqrt{x^T H x} \).

3. New parameter conditions. In this section, we present the relaxed parameter conditions for PDHG method, and derive an “optimal” parameter condition among the proposed conditions. Before introducing the parameter condition, we need to impose the following strongly convexity assumption.

**Assumption:**

- \( g(x) \) is strongly convex w.r.t. \( x \) with modulus \( \gamma > 0 \), i.e.,

\[
\begin{aligned}
g(\hat{x}) - g(x) \geq (\hat{x} - x)^T \xi_x + \frac{\gamma}{2} \| \hat{x} - x \|^2, \quad \forall x, \hat{x} \in \mathcal{X}.
\end{aligned}
\]

(3.1)

Indeed, the assumption (3.1) can be met in many practical applications, see the examples shown in [17, 4, 5, 6, 23, 33, 19]. The new parameter condition for PDHG method is as follows:

- 

\[
\begin{aligned}
\tau^{-1} \sigma^{-1} \geq \left( \frac{(1 - \alpha)^2}{\gamma} + \alpha^2 \right) \| A^T A \|,
\end{aligned}
\]

(3.2)

where \( \| \cdot \| \) denotes the spectral norm of a matrix.

**Remarks:**

- Note (3.2) implies the following relation between \( \sigma \) and \( \alpha \):

\[
\begin{aligned}
\sigma^{-1} \geq \frac{(1 - \alpha)^2 \| A^T A \|}{\gamma},
\end{aligned}
\]

(3.3)

which will be useful later.

- With each fixed \( \alpha \), (3.2) becomes a relation between \( \tau \) and \( \sigma \), hence (3.2) can also be recognized as a category of parameter conditions (for \( \tau \) and \( \sigma \)).
Indeed, many applications of (1.1) meet the strongly convexity assumption (3.1). For example, the application of the linearized Bregman scheme in [17] to some sparse or low-rank optimization models (e.g., [4, 5, 6]); and various splitting versions of the augmented Lagrangian method in [22, 26] to some convex minimization models with linear constraints (e.g., [4, 23, 33]). In [19], the standard TV-denosing model was proved to satisfy the assumption (3.1) too.

Compared with the parameter condition \( \tau \sigma < \frac{1}{\|A^T A\|} \) (see (1.3)) which is essential to the convergence of CP method and many other primal-dual methods, (3.2) is usually more relaxed which enables larger step size and potentially faster convergence. To see this, we take \( \alpha = 0.5 \) in (3.2) for example, then the upper bound of \( \sigma \tau \) becomes \( 4/(1 + \gamma^{-1} \tau^{-1}) \|A^T A\| \) which is close to \( 4/\|A^T A\| \) with relatively large \( \tau \). This means the product of step sizes \( \sigma \) and \( \tau \) can be almost quadrupled compared with that in classical CP method.

There are two special cases of (3.2): When we take \( \alpha = 1 \), the primal-dual framework (1.2) reduces to the CP method and the new parameter condition (3.2) reduces to \( \tau \sigma < 1/\|A^T A\| \) which is exactly the convergence requirement in [9]; When \( \alpha \) is set to be 0, the primal-dual framework (1.2) becomes to the PDHG method, and the new parameter condition (3.2) reduces to \( \sigma < \gamma/\|A^T A\| \) which is exactly the convergence requirement for PDHG method as shown in [19]. Based on the above analysis, our parameter condition is a generalization of some existing conditions. We also observe that the right hand side of (3.2) is strongly convex w.r.t. \( \alpha \), so if it is minimized over \( \alpha \), we can further extend the domain of \( (\tau, \sigma) \). By simple calculation, we find that the minimizer of the right hand side of (3.2) is
\[
\alpha^* = \frac{1}{1 + \gamma \tau}.
\] (3.4)

Plugging \( \alpha = \alpha^* \) into (3.2), we obtain the “optimal” parameter condition as follows:
\[
\sigma \leq \frac{(1 + \gamma \tau)}{\|A^T A\|}.
\] (3.5)

It is trivial that (3.5) is always more relaxed than \( \tau \sigma < 1/\|A^T A\| \), and the domain of \( (\tau, \sigma) \) is strictly extended. We plot the boundary curves of parameter domain with “optimal” \( \alpha \) and several fixed \( \alpha \) in Figure 1.

![Figure 1: Comparison of parameter domain with different setting of \( \alpha \).](image)

We see from Figure 1 that the domain of new parameter condition (3.2) with any fixed \( \alpha \) is larger than that of original parameter condition (1.3) when \( \gamma \) is larger than...
some threshold. Furthermore we observe from Figure 1 that the domain of optimal parameter condition is always larger than that of new condition. More specifically, the boundary curve of parameter domain with optimal parameter condition is tangent to that with new parameter condition and fixed $\alpha$. i.e., the “optimal” boundary curve is the envelop of all boundary curves of parameter domain with fixed $\alpha$, hence it is optimal.

4. Theoretical analysis. In this section, we analyze the contractive property of PDHG method with relaxed parameter condition under mild assumptions, and then establish the convergence property of PDHG method with relaxed parameter condition.

We first prove that when $g$ satisfies the assumption (3.1), the generated sequence of PDHG method (1.2) is convergent to the solution set $\Omega^*$ provided (3.2) hold.

Lemma 4.1. With given $u^k$, let $u^{k+1}$ be generated by PDHG method (1.2). $\theta(u)$ and $\Omega$ are defined in (2.1b). Then we have

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

where

$$Q = \begin{pmatrix} \tau^{-1} I_n & A^T \\ \alpha A & \sigma^{-1} I_m \end{pmatrix}. \quad (4.7)$$

Proof. Invoking the first-order optimality conditions of subproblems in (1.2) we get

$$x^{k+1} \in \mathcal{X}, \quad g(x) - g(x^{k+1}) + (x - x^{k+1})^T \{ - A^T y^k + \tau^{-1} (x^{k+1} - x) \} \geq 0, \quad \forall x \in \mathcal{X},$$

and

$$y^{k+1} \in \mathcal{Y}, \quad f^* (y) - f^* (y^{k+1}) + (y - y^{k+1})^T \{ A x^{k+1} + \alpha A (x^{k+1} - x) + \sigma^{-1} (y^{k+1} - y^k) \} \geq 0, \quad \forall y \in \mathcal{Y}. \quad (4.8)$$

Summing the above two inequalities, we can obtain

$$u^{k+1} = (x^{k+1}, y^{k+1}) \in \Omega \quad \text{and} \quad \theta(u) - \theta(u^{k+1}) + \left( \frac{x - x^{k+1}}{\tau} \right)^T \left\{ - A^T y^{k+1} \right\} + \left( \frac{y - y^{k+1}}{\sigma} \right)^T \left\{ A x^{k+1} \right\} \geq 0, \quad \forall u \in \Omega. \quad (4.9)$$

Using the definitions of the matrices $M$ and $Q$ (see (2.1b) and (4.7)), the above inequality can be reformulated as

$$u^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T M u^{k+1} \geq (u - u^{k+1})^T Q (u^k - u^{k+1}), \quad \forall u \in \Omega. \quad (4.10)$$

The assertion is thus proved. \hfill \square

Lemma 4.2. With given $u^k$, let $u^{k+1}$ be generated by PDHG method (1.2). If the assumption (3.1) and parameter condition (3.2) hold, then we have

$$(u^{k+1} - u)^T H (u^k - u^{k+1}) \geq (u^{k+1} - u)^T F(u) - \frac{(1 - \alpha)^2}{2\gamma} \| A^T (y^k - y^{k+1}) \|^2, \quad \forall u \in \Omega, \quad (4.11)$$

where

$$H = \begin{pmatrix} \tau^{-1} I_n & \alpha A^T \\ \alpha A & \sigma^{-1} I_m \end{pmatrix}. \quad (4.12)$$

and $F(u)$ is defined in (2.2b).
Proof. We deal with the right-hand side of (4.6) first. Recalling the assumption (3.1) and the convexity of $f^*(y)$ w.r.t. $y$, for any $\xi \in \partial g(x)$ and $\eta \in \partial f^*(y)$, we have
\[
g(x^{k+1}) - g(x) \geq \frac{\gamma}{2} ||x^{k+1} - x||^2 + (x^{k+1} - x)^T \xi_x,
\]
and
\[
f^*(y^{k+1}) - f^*(y) \geq (y^{k+1} - y)^T \eta_y.
\]
In other words, for any $\xi \in \partial g(x)$ and $\eta \in \partial f^*(y)$, we have
\[
\theta(u^{k+1}) - \theta(u) \geq \left( \begin{array}{c} x^{k+1} - x \\ y^{k+1} - y \end{array} \right)^T \left( \begin{array}{c} \xi_x \\ \eta_y \end{array} \right) + \frac{\gamma}{2} ||x^{k+1} - x||^2. \tag{4.13}
\]
Since the matrix $M$ is skew-symmetric, we have
\[
(u^{k+1} - u)^T M u^{k+1} = (u^{k+1} - u)^T M u. \tag{4.14}
\]
Combining (4.13) and (4.14), and using the notation of $F(u)$ (see (2.2b)), we obtain
\[
\theta(u^{k+1}) - \theta(u) + (u^{k+1} - u)^T M u^{k+1} \geq (u^{k+1} - u)^T F(u) + \frac{\gamma}{2} ||x^{k+1} - x||^2.
\]
Substituting it into the right-hand side of (4.6), we get
\[
(u^{k+1} - u)^T Q(u^k - u^{k+1}) \geq (u^{k+1} - u)^T F(u) + \frac{\gamma}{2} ||x^{k+1} - x||^2, \quad \forall u \in \Omega.
\]
Using the notations of $H$ and $Q$, it follows from the last inequality that
\[
(u^{k+1} - u)^T H(u^k - u^{k+1}) \geq (u^{k+1} - u)^T F(u) + \frac{\gamma}{2} ||x^{k+1} - x||^2
\]
\[
- (1 - \sigma)(Ax^{k+1} - Ax)^T (y^k - y^{k+1}). \tag{4.15}
\]
Furthermore, using the Cauchy-Schwarz inequality, we get
\[
- (1 - \sigma)(Ax^{k+1} - Ax)^T (y^k - y^{k+1}) \geq - \frac{\gamma}{2} ||x^{k+1} - x||^2 - \frac{(1 - \sigma)\gamma}{2\sigma^2} ||A^T (y^k - y^{k+1})||^2.
\]
Substituting it into the right-hand side of (4.15), we obtain
\[
(u^{k+1} - u)^T H(u^k - u^{k+1}) \geq (u^{k+1} - u)^T F(u) - \frac{(1 - \sigma)^2}{2\sigma^2} ||A^T (y^k - y^{k+1})||^2.
\]
The assertion (4.11) is proved. \qed

Lemma 4.3. With given $u^k$, let $u^{k+1}$ be generated by PDHG method (1.2). If the assumption (3.1) and parameter condition (3.2) are satisfied, then we have
\[
(u - u^{k+1})^T F(u) \geq \frac{1}{2} \left( ||u - u^{k+1}||_H^2 - ||u - u^k||_H^2 \right) + \frac{1}{2} ||u^{k+1} - u||_G^2, \quad \forall u \in \Omega, \tag{4.16}
\]
where the matrix $H$ is defined by (4.12) and
\[
G = \begin{pmatrix} \tau^{-1} I_n & \alpha A^T \\ \alpha A & (\sigma^{-1} \frac{(1 - \alpha)\gamma}{\gamma^2} ||A^T A||) I_m \end{pmatrix}. \tag{4.17}
\]
Proof. First, note that the matrix $G$ is positive semi-definite due to (3.2) and (3.3). This conclusion will be used in (4.4). It follows from (4.11) that
\[
(u - u^{k+1})^T F(u) \geq (u - u^{k+1})^T H(u^k - u^{k+1}) - \frac{(1 - \alpha)^2 ||A^T A||}{2\gamma} ||y^k - y^{k+1}||^2, \quad \forall u \in \Omega. \tag{4.18}
\]
Applying the identity
\[
b^T H(b - a) = \frac{1}{2} \left( ||b||_H^2 - ||a||_H^2 + ||a - b||_H^2 \right),
\]
to the first term in the right-hand side of (4.18) with
\[ a = u - u^k \quad \text{and} \quad b = u - u^{k+1}, \]
we have
\[
(u - u^{k+1})^T F(u) \geq \frac{1}{2} (\|u - u^{k+1}\|_H^2 - \|u - u^k\|_H^2) + \frac{1}{2} \|u^k - u^{k+1}\|_H^2 - \frac{(1 - \alpha)^2 \|A^T A\|}{2\gamma} \|y^k - y^{k+1}\|^2.
\] (4.19)

Using the notation of matrix \( G \), the last two terms in the right-hand side of (4.19) lead to
\[
\|u^k - u^{k+1}\|_H^2 - \frac{(1 - \alpha)^2 \|A^T A\|}{2\gamma} \|y^k - y^{k+1}\|^2 = \|u^k - u^{k+1}\|_G^2.
\] (4.20)

Substituting it in (4.19), we proved the assertion. \( \square \)

Based on the above lemmas, we are able to establish the contractive property of the sequence generated by PDHG method with relaxed parameter condition.

**Theorem 4.4.** With given \( u^k \), let \( u^{k+1} \) be generated by PDHG method. If the assumption (3.1) and parameter condition (3.2) are satisfied, then the sequence \( \{u^k\} \) converges to an arbitrary solution of original problem (1.1).

**Proof.** According to (4.16), it holds that
\[
\|u - u^k\|_H^2 - \|u - u^{k+1}\|_H^2 - \|u^k - u^{k+1}\|_G^2 \geq 2(u^{k+1} - u)^T F(u), \quad \forall u \in \Omega,
\]
where the matrices \( H \) and \( G \) are defined in (4.12) and (4.17), respectively.

Substituting \( u = u^* \) in the above inequality and using the fact \( (u^{k+1} - u^*)^T F(u^*) \geq 0 \), we get
\[
\|u^k - u^*\|_H^2 - \|u^{k+1} - u^*\|_H^2 - \|u^k - u^{k+1}\|_G^2 \geq 0, \quad \forall u^* \in \Omega^*.
\] (4.21)
Take \( k = 1 \) to (4.21) and sum the inequalities up, and note that the matrix \( G \) is positive semi-definite due to (3.2) and (3.3), we obtain that the sequence \( \{u^k\} \) is bounded and
\[
\lim_{k \to \infty} \|u^k - u^{k+1}\| = 0.
\] (4.22)

Let \( u^\infty \) be a cluster point of \( \{u^k\} \) and \( \{u^{k_j}\} \) be the subsequence converging to \( u^\infty \). Taking \( k = k_j - 1 \) in (4.10) and letting \( j \to \infty \), noticing (4.22) and the continuity of \( \theta(u) \), we obtain
\[
u^\infty \in \Omega, \quad \theta(u) - \theta(u^\infty) + (u - u^\infty)^T M u^\infty \geq 0, \quad \forall u \in \Omega,
\]
which indicates that \( u^\infty \) is a solution of VI(\( \Omega, \theta, M \)). Hence (4.21) is also valid when \( u^* \) is replaced with \( u^\infty \). Combining (4.21) with \( \lim_{j \to \infty} u^{k_j} = u^\infty \), we arrive at
\[
\lim_{k \to \infty} \|u^k - u^\infty\|_H = 0
\]
hence the proof is complete. \( \square \)

5. **Numerical experiments.** In this section, numerical experiments were carried out to validate the efficiency of PDHG method with relaxed condition. We applied PDHG method with relaxed parameter condition(PDHG-R) to solve the LASSO model and TV-l2 model, and compare it with the classic PDHG method, CP method, CP method with acceleration(CP-A) and FISTA.

All algorithms were coded and implemented in Matlab 2016a, and all experiments were carried out on a desktop computer with a 4.0 GHz intel Core i7 6700k CPU and 8 GB of memory.
5.1. LASSO. The LASSO model was first introduced in [29] to solve variable selec-
tion regression problems. Given a sample matrix \( A \in \mathbb{R}^{m \times n} \) and the response \( b \in \mathbb{R}^m \), the LASSO model learns the linear regression coefficient \( x \in \mathbb{R}^n \) which is a sparse vector with \( k \) nonzero entries by solving

\[
\min_{x \in X} \beta \|x\|_1 + \frac{1}{2} \|Ax - b\|_2^2
\]

for some scaling parameter \( \beta \). By the definition of conjugate function, (5.23) can be reformulated as the following minmax problem:

\[
\min_{x \in X} \max_y \beta \|x\|_1 + y^T Ax - \frac{1}{2} \|y + b\|_2^2,
\]

where the last term of (5.24) is strongly convex w.r.t. \( y \) with \( \gamma = 1 \).

The data in our test were generated in the same fashion as that in Section 11.1 of [3]. Each element of the sample matrix \( A \) is drawn from an \( N(0, 1) \) distribution. A true \( x^{true} \) is generated with \( s \) non-zero entries at random position. each element is sampled from an \( N(0, 1) \) distribution. The constant \( b \) is computed as \( b = Ax^{true} + v \) where \( v \) is Gaussian noise.

Some preliminary experiments show that PDHG, CP and PDHG-R methods perform well for \( \tau = 0.2 \), \( \sigma = 5/\|A^T A\| \). Also PDHG-R converges the fastest with \( \alpha = a/\gamma \). Therefore unless specified otherwise, the parameter settings are \( \tau = 0.2 \), \( \sigma = 5/\|A^T A\| \), sample size \( m \) is 100, feature number \( n \) is 10000, \( \beta = 0.2\beta_{max} \), number of non-zero elements \( s \) in \( x^{true} \) is 10. Here \( \beta_{max} = \|A^T b\|_{\infty} \) is a value such that any \( \beta \) beyond the value leads to a zero solution.

First we run all the algorithms for fixed 100 iterations and present the iteration progress of relative errors. The relative error is defined in the following way: we first run the PDHG method for as many as 10000 iterations and take the final objective value as the optimal value \( \Phi^* \) approximately; then the error is calculated by \( \|\Phi^k - \Phi^*\|/\|\Phi^*\| \) where \( \Phi^k \) is the objective value at \( k \)-th iteration. We omit the CPU time comparison since the per-iteration cost of different algorithms are almost identical. Figure 2 and Table 1 illustrate the experimental results for random problems under different duplet \( (m, n) \), i.e. sample size and feature number. We observe that PDHG-R with the optimal parameter condition converges faster than other competing algorithms in most cases.

To have deeper insight of the performance of our algorithm, we do further miscellaneous tests. Figure 3 and table 2 depict the experimental results with different sparsity levels \( s \) and that with different balancing parameter \( \beta \). The result is consistent with previous results showing that PDHG method with relaxed condition is most efficient in most cases. The only exceptions is that the accelerated CP(CP-A) can be more efficient in certain circumstances. However CP-A is very sensitive to the parameter choice and its performance can be very undesirable in other cases as shown in Figure 3 and table 2. For example, CP-A reached the tolerance level of \( 10^{-6} \) in 134 iterations when \( s = 5 \). however it takes 3450 iterations, more than 20 times slower than that in last case and 7 times slower than any other methods, to reach the same accuracy when \( s = 20 \).

The above experimental results confirm that PDHG method with relaxed condition is most efficient in most cases and this performance advantage is invulnerable to various problem settings.

5.2. TV-l2 image denoising. Variational models have been extremely successful in a wide variety of image restoration problems, and remain one of the most active areas of research in mathematical image processing and computer vision. The most
fundamental image restoration problem is perhaps denoising. It forms a significant preliminary step in many machine vision tasks such as object detection and recognition. Total variation based image restoration models were first introduced by Rudin, Osher, and Fatemi (ROF) in their pioneering work \cite{ROF}. It was designed with the explicit goal of preserving sharp discontinuities (edges) in an image while
removing noise and other unwanted fine scale detail. In case of deblurring problem, it was discretized and formulated as the following minimization problem:

$$\min_{x \in X} \|Ax\|_1 + \frac{\lambda}{2}\|x - f\|_2^2,$$

(5.25)
where $A$ is the discrete gradient operator and $f$ is the observed image. It can also be written as the (1.1) form:

$$\min_{x \in X} \max_{y \in B_\infty} \frac{\lambda}{2} \|x - f\|^2_2 + y^T Ax. \quad (5.26)$$

The problem settings are similar with that in Section 6.2 of [9]. We compare the methods by restoring images 'sunflower.png(900 × 824)' and 'boat.png(512 × 512)' corrupted by Gaussian noise with standard deviation 0.05. We set the stepsize parameters $\tau = \tau_0 = \sigma = 1/\sqrt{\|A^T A\|}$, $\gamma = 0.35\lambda$ and $\tau_n, \sigma_n = 1/\|A^T A\|$. We use $\lambda = 8$ and $\lambda = 16$ for image restoration as suggested in Section 6.2 of [9]. To show the effect of restored images, the original, noisy and recovered images are presented in figure 4 and 5 when $\lambda = 8$.

![Image](image.png)

**Figure 4.** From left to right, up to bottom: the original figure 'sunflower.png(900 × 824)', noisy figure with additive Gaussian noise (Standard deviation=0.05), figure recovered by PDHG method ($\lambda = 8$) and figure recovered by PDHG-R ($\lambda = 8$).

We use peak signal-to-noise ratio (PSNR) to measure the effect of image recovery. PSNR is plotted against iteration No. in all figures. As we can see in figure 6, PDHG-R outperform the rest methods for both image restorations. This result is consistent when the balancing parameter $\lambda$ of the ROF model varies between 8 and 16 as recommended in section 6.2 of [9].

Since most primal-dual methods are sensitive to the stepsize parameter, we now investigate the influence of stepsize on the methods by restoring “cameraman.png(256×256)”. In figure 7, we find that PDHG-R remains the most efficient for image restoration with various stepsize parameters. In figure 7 FISTA is used as a benchmark as it does not contain the parameter $\tau$ and $\sigma$. We find that CP and PDHG method are evidently faster than FISTA when $\tau = 0.1$ or 1 but are evidently slower than FISTA when $\tau = 10$. Therefore PDHG-R is more stable than PDHG and CP methods.

Both experiments demonstrate that our relaxed condition makes the PDHG method the fastest as well as the most stable one among the methods in comparison.
Figure 5. From left to right, up to bottom: the original figure 'boat.png (512 × 512)', noisy figure with additive Gaussian noise (Standard deviation=0.05), figure recovered by PDHG method ($\lambda = 8$) and figure recovered by PDHG-R ($\lambda = 8$).

Figure 6. Results of image denoising with different images and $\lambda$
6. Conclusions. In this paper, we proposed a new parameter condition with constant stepsize for PDHG method to solve saddle-point problems. The contractive property of the sequence generated by PDHG-R was derived under some mild assumptions hence the convergence of the method was proved. Numerical experiments show that PDHG method with our relaxed parameter condition outperforms some state-of-art methods under various problem settings.

Acknowledgment. We would like to give special thanks to Bingsheng He for his valuable advice, guidance and constructive comments.

Appendix. In this section, we show by a toy example that the sequence generated by PDHG method (1.2) with $\alpha \neq 1$ is not necessarily convergent. We consider the following saddle point problem:

\[
\min_x \max_y L(x, y) = x - y(x - 1),
\]

defined on $\mathbb{R}_+ \times \mathbb{R}_+$. It is easy to verify that $(x^*_r, y^*_r) = (1, 1)$ is the unique saddle point.

Finding the saddle point of $L(x, y)$ falls into the category of the model (1.1) with $g(x) = x$, $f^*(y) = -y$, $A = 1$, $\mathcal{X} = \mathcal{Y} = \mathbb{R}_+$. When PDHG method (1.2) is applied to solve (6.27) with $\tau = \sigma = 2$, $\alpha = 0.5$, the scheme (1.2) reduces to

\[
\begin{align*}
    x^{k+1} &= \arg \min_x \{ L(x, y^k) + \frac{1}{2\tau} \| x - x^k \|^2 \mid x \geq 0 \} \\
    &= \max \{ (x^k + 2y^k - 2), 0 \}, \\
    y^{k+1} &= \arg \max_y \{ L(x^{k+1}, y) - \frac{1}{2\sigma} \| y - y^k \|^2 \mid y \geq 0 \} \\
    &= \max \{ (y^k - 3x^{k+1} + x^k + 2), 0 \}. \\
\end{align*}
\]

Figure 7. Results of image denoising with different stepsize $\tau$ and $\sigma$. 
Starting with \((\frac{x_0^k}{y_0^k}) = (0_0^0)\), the sequence \((\frac{x_k^k}{y_k^k})\) generated by (6.28) is:

\[
\begin{align*}
(\frac{x^0}{y^0}) &= (0 \ 0), \quad (\frac{x^1}{y^1}) = (0 \ 2), \quad (\frac{x^2}{y^2}) = (2 \ 0), \quad (\frac{x^3}{y^3}) = (0 \ 4), \\
(\frac{x^4}{y^4}) &= (6 \ 0), \quad (\frac{x^5}{y^5}) = (4 \ 0), \quad (\frac{x^6}{y^6}) = (2 \ 0) = (\frac{x^2}{y^2}), \quad \ldots.
\end{align*}
\]

This is actually a cyclic sequence with four distinct iterates:

\[
(\frac{x^{k+4}}{y^{k+4}}) = (\frac{x^k}{y^k}), \quad \forall k \geq 2.
\]

Hence, the sequence \(\{(x_k, y_k)\}\) generated by PDHG method (1.2) with \(\tau = \sigma = 2, \alpha = 0.5\) does not converge to the saddle-point \((1, 1)\).

**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] 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), 236–253.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed optimization and statistical learning via ADMM, *Found. Trends Mach. Learn.*, 3 (2010), 1–122.

[4] J. Cai, E. J. Candès and Z. Shen, A singular value thresholding algorithm for matrix completion, *SIAM J. Optim.*, 20 (2010), 1956–1982.

[5] J. Cai, S. Osher and Z. Shen, Linearized Bregman iterations for compressed sensing, *Math. Comput.*, 78 (2009), 1515–1536.

[6] J. Cai, S. Osher and Z. Shen, Linearized Bregman iteration for frame based image deblurring, *SIAM J. Imaging Sci.*, 2 (2009), 226–252.

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

[8] A. Chambolle, An algorithm for total variation minimization and applications, *J. Math. Imaging Vis.,* 20 (2004), 89–97.

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

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

[11] R. Chan, S. Ma and J. Yang, Inertial primal dual algorithms for structured convex optimization, *SIAM J. Imag. Sci.*, 8 (2015), 2239–2267.

[12] T. Chan, G. Golub and P. Mulet, A nonlinear primal-dual method for total variation-based image restoration, *SIAM J. Sci. Comput.*, 20 (1999), 1964–1977.

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

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

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

[16] T. Goldstein, M. Li and X. Yuan, Adaptive primal-dual splitting methods for statistical learning and image processing, *Adv. Neural Inform. Process. Syst.*, (2015), 2089–2097.

[17] T. Goldstein and S. Osher, The split Bregman method for L1-regularized problems, *SIAM J. Imag. Sci.*, 2 (2009), 323–343.

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

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

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

[21] H. He, J. Desai and K. Wang, A primal-dual prediction-correction algorithm for saddle point optimization, *J. Global Optim.*, 66 (2016), 573–583.
[22] M. Hestenes, Multiplier and gradient methods, *J. Optim. Theory Appl.*, 4 (1969), 303–320.
[23] N. Higham, Computing the nearest correlation matrix —A problem from finance, *IMA J. Numer. Anal.*, 22 (2002), 329–343.
[24] Y. Nesterov, Gradient methods for minimizing composite objective function, *Math. Program.*, 140 (2013), 125–161.
[25] J. Pesquet and N. Pustelnik, A parallel inertial proximal optimization method, *Pac. J. Optim.*, 8 (2012), 273–306.
[26] M. Powell, A method for nonlinear constraints in minimization problems, In *Optimization edited by R. Fletcher*, Academic Press, (1969), 283–298.
[27] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D*, 60 (1992), 259–268.
[28] Y. Shen, Q. Li and J. Wu, A variable step-size primal-dual algorithm based on proximal point algorithm (in Chinese), *Math. Numer. Sinica.*, 40 (2018), 85–95.
[29] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. Roy. Statist. Soc. Ser. B*, 58 (1996), 267–288.
[30] T. Valkonen, Inertial, corrected, primal-dual proximal splitting, *SIAM J. Optim.*, 30 (2020), 1391–1420.
[31] Y. Wang, J. Yang, W. Yin and Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, *SIAM J. Imaging Sci.*, 1 (2008), 248–272.
[32] P. Weiss, L. Blanc-Feraud and G. Aubert, Efficient schemes for total variation minimization under constraints in image processing, *SIAM J. Sci. Comput.*, 31 (2009), 2047–2080.
[33] B. Zhang, Z. Zhu and S. Wang, A simple primal-dual method for total variation image restoration, *J. Vis. Commun. Image R.*, 38 (2016), 814–823.
[34] H. Zhang, J. Cai, L. Cheng and J. Zhu, Strongly convex programming for exact matrix completion and robust principal component analysis, *Inver. Prob. Imaging*, 6 (2012), 357–372.
[35] X. Zhang, M. Burger and S. Osher, A unified primal-dual algorithm framework based on Bregman iteration, *J. Sci. Comput.*, 46 (2011), 20–46.
[36] M. Zhu and T. Chan, An Efficient Primal-dual Hybrid Gradient Algorithm for Total Variation Image Restoration, CAM Report 08-34, UCLA, USA, 2008.

Received February 2021; revised August 2021; early access January 2022.

E-mail address: 369318324@qq.com
E-mail address: 229369689@qq.com.
E-mail address: 769767821@qq.com
E-mail address: ocsiban@126.com