Noname manuscript No.
(will be inserted by the editor)

A search direction inspired primal-dual method for saddle point problems

Shengjie Xu

Received: date / Accepted: date

Abstract The primal-dual hybrid gradient algorithm (PDHG), which is indeed the Arrow-Hurwicz method, has been widely used in many application areas especially in image processing. However, the convergence of PDHG was established only under some restrictive conditions in the literature, and it is still missing for the case without extra constraints. In this paper, from a perspective of the variational inequality, we show the original PDHG can provide some ascent directions to the solution set. Such a fact naturally captures the line search method in optimization. Furthermore, inspired by Newton's method, we present a reversible PDHG to solve the saddle point problem. Compared with PDHG, each iteration of the new method needs only two additional minor computations. Moreover, we establish its global convergence and a worst-case $\mathcal{O}(1/t)$ convergence rate. Extensive numerical experiments, including several examples of image inpainting, demonstrate the efficient performance of the novel method.

Keywords Convex optimization · Convergence rate · Image restoration · Primal-dual hybrid gradient method · Saddle point problem · Variational inequality

1 Introduction

We consider a general saddle point problem

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\big\{\Phi(x,y):=\theta_1(x)-y^TAx-\theta_2(y)\big\},\tag{1}$$

where $\mathcal{X}\subset\Re^n$ and $\mathcal{Y}\subset\Re^m$ are closed convex sets, $A\in\Re^{m\times n}$, $\theta_1(x):\Re^n\to\Re$ and $\theta_2(y):\Re^m\to\Re$ are closed convex but not necessarily smooth functions. Throughout this paper, we assume the solution set of (1) is nonempty. There are many problems in different application areas can be reformulated as (1). For example, finding a saddle point of the Lagrangian function of the classic convex minimization problem with linear constraints is a special case of (1). In particular, the model (1) captures a lot of image restoration models with the total variation (TV) regularization introduced in [23], more details can be found in, e.g., [4,25,27].

In fact, as analyzed in [4,5,9,15], the saddle-point problem (1) can be converted to the primal-dual formulation of a nonlinear programming problem. Such a fact has inspired many primal-dual algorithms for TV image restoration problems, we refer the readers to [4,9,11,15, 25,27]. In particular, Chambolle and Pock (abbreviated as C-P) proposed a general primal-dual method in [4]. As described in [16], C-P method has an equivalent form

$$\begin{cases} x^{k+1} = \arg\min\big\{\Phi(x,y^k)+\frac{r}{2}\|x-x^k\|_2^2 \;\big|\; x\in\mathcal{X}\big\},\\ \bar{x}^k \;\;= x^{k+1}+\tau(x^{k+1}-x^k),\\ y^{k+1} = \arg\max\big\{\Phi(\bar{x}^k,y)-\frac{s}{2}\|y-y^k\|_2^2 \;\big|\; y\in\mathcal{Y}\big\},\end{cases}\tag{2}$$

in which $\tau\in[0,1]$ is a combination parameter; $r>0$ and $s>0$ are proximal parameters of the regularization terms. Moreover, $1/r$ and $1/s$ denote the step sizes associated with gradient-type methods for solving the decomposed subproblems in (2) (details can be found in, e.g., [4,25,27]). For case $\tau=1$, a simple convergence proof of the C-P method (2) was given in [15], and the authors interpreted it as an instance of the proximal point algorithm, which is also named the customized proximal point algorithm (CPPA) in [12,20,21]. Moreover, the CPPA is convergent when

Department of Mathematics, Harbin Institute of Technology, China
Department of Mathematics, Southern University of Science and Technology, China
E-mail: xsjnsu@163.com

$rs > \rho(A^TA)$, where $\rho(\cdot)$ denotes the spectral radius of a matrix.

For case $\tau = 0$, the C-P method (2) will be reduced to PDHG proposed in [27], which is essentially the semi-implicit classical Arrow-Hurwicz algorithm in [1] and has an iterative scheme:

$$\begin{cases} x^{k+1} = \arg\min \left\{ \Phi(x, y^k) + \frac{r}{2}\|x - x^k\|_2^2 \;\middle|\; x \in \mathcal{X} \right\}, \\ y^{k+1} = \arg\max \left\{ \Phi(x^{k+1}, y) - \frac{s}{2}\|y - y^k\|_2^2 \;\middle|\; y \in \mathcal{Y} \right\}. \end{cases} \tag{3}$$

The efficiency of PDHG satisfying $rs > \rho(A^TA)$ has been well demonstrated in the imaging literature, we refer the readers to [2,9,25,27]. In [9], the authors proved the PDHG with some restrictions on its step size is convergent for TV image denoising. Later, in [14], He et al established PDHG's convergence when one of functions of the saddle problem (1) is strongly convex. Moreover, by introducing a counterexample, the authors also numerically showed the original PDHG is not necessarily convergent, even though the step sizes $1/r$ and $1/s$ are sufficiently small. Therefore, the convergence of PDHG without extra conditions is still missing.

Our aim in this paper is to present a convergent PDHG-based method by a prediction-correction (P-C) way. There has been tremendous interest in developing new algorithms by P-C way, we refer the readers to [12, 15,16,20]. According to the basic line search method in optimization: first, we show PDHG, as a prediction step, can provide some ascent directions to the solution set of (1). Furthermore, inspired by Newton's method and alternating direction method with Gaussian back substitution in [18], we introduce a simple Newton-like search direction; second, we calculate the step size along such a direction by a simple rule; finally, based on the above two points, we propose a new modified PDHG method to solve the saddle point problem.

The novel method needs only an additional simple correction step, though it requires four extra matrix-vector multiplications, it would be cheaper if the sub-problem is more expensive. The new method is convergent just provided $rs > \rho(A^TA)/4$ compared with the original PDHG. In theory, the relaxation of regularization term allows a bigger step size to potentially reduce required convergence iterations (see related work in, e.g., [17,19,26]). Furthermore, we establish its global convergence and a $\mathcal{O}(1/t)$ convergence rate in an ergodic sense. Finally, taking TV image inpainting as an example, we construct some experiments to show the efficient performance of the proposed method.

This paper is organized as follows. In Section 2, from the perspective of variational inequality (VI), we summarize some preliminary results and analyze the VI-structure of the original PDHG. In Section 3, inspired by Newton's direction, we establish our novel method by the basic line search method. Furthermore, we establish its convergence analysis in Section 4. Finally, we check the numerical performance of the new method by some extensive examples.

2 Preliminaries

In this section, we characterize the optimal condition of problem (1) from the perspective of VI. Moreover, we derive the basic VI-structure of the original PDHG (3).

2.1 The optimal condition for problem (1)

The analysis of this paper is based on the following lemma, its proof is elementary and thus omitted here (details can be found in [21]).

Lemma 1 *Suppose $\mathcal{X} \subset \Re^n$ is a closed convex set, $\theta(x) : \mathcal{X} \to \Re$ and $\varphi(x) : \mathcal{X} \to \Re$ are convex functions, and $\varphi(x)$ is differentiable. Assume the solution set of the minimization problem $\min\{\theta(x) + \varphi(x) \mid x \in \mathcal{X}\}$ is nonempty, then*

$$x^* \in \arg\min\{\theta(x) + \varphi(x) \mid x \in \mathcal{X}\}$$

if and only if

$$x^* \in \mathcal{X}, \;\; \theta(x) - \theta(x^*) + (x - x^*)^T \nabla\varphi(x^*) \geq 0, \;\; \forall\, x \in \mathcal{X}.$$

To begin with, we reformulate the optimal condition of the problem (1) into a VI-form. The point (x^*, y^*) is a saddle point of $\Phi(x, y)$ means

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

in other words, we have

$$\begin{cases} x^* \in \arg\min \left\{ \Phi(x, y^*) \;\middle|\; x \in \mathcal{X} \right\}, \\ y^* \in \arg\min \left\{ -\Phi(x^*, y) \;\middle|\; y \in \mathcal{Y} \right\}. \end{cases}$$

By using Lemma 1, it is equivalent to

$$\begin{cases} \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^Ty^*) \geq 0, & \forall\, x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(Ax^*) \geq 0, & \forall\, y \in \mathcal{Y}. \end{cases} \tag{4}$$

Furthermore, by defining

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(u) = \begin{pmatrix} 0 & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{5}$$

$$\mathcal{U} = \mathcal{X} \times \mathcal{Y} \qquad \text{and} \qquad \theta(u) = \theta_1(x) + \theta_2(y),$$

we can rewrite (4) as a compact form:

$$u^* \in \mathcal{U},\ \theta(u)-\theta(u^*)+(u-u^*)^T F(u^*) \geq 0,\ \forall u \in \mathcal{U}. \quad (6)$$

Noting $F(u)$ in (5) is linear and has a skew-symmetric structure, hence we have

$$(u-v)^T(F(u)-F(v)) \equiv 0, \quad (7)$$

which shows F is monotone (see definition in [24]). For these reasons, we also call (6) the mixed monotone variational inequality (abbreviated as MVI). Throughout this paper, we assume the solution set $\mathcal{U}^*$ of (6) is nonempty.

2.2 MVI-structure of the original PDHG

We analyze the MVI-structure of the original PDHG in this subsection. For x-subproblem in (3), according to the basic Lemma 1, $x^{k+1} \in \mathcal{X}$ satisfies

$$\begin{aligned}&\theta_1(x)-\theta_1(x^{k+1})\\ &\quad+(x-x^{k+1})^T\big\{-A^T y^k + r(x^{k+1}-x^k)\big\}\\ &\geq 0, \quad \forall x \in \mathcal{X}.\end{aligned}$$

It can be rewritten as

$$\begin{aligned}&\theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\\ &\quad\big\{-A^T y^{k+1} + r(x^{k+1}-x^k) + A^T(y^{k+1}-y^k)\big\} \quad (8)\\ &\geq 0, \quad \forall x \in \mathcal{X}.\end{aligned}$$

Similarly, for y-subproblem in (3), $y^{k+1} \in \mathcal{Y}$ satisfies

$$\begin{aligned}&\theta_2(y)-\theta_2(y^{k+1})\\ &\quad+(y-y^{k+1})^T\big\{Ax^{k+1} + s(y^{k+1}-y^k)\big\} \quad (9)\\ &\geq 0, \quad \forall y \in \mathcal{Y}.\end{aligned}$$

By adding inequalities (8) and (9) and combing with the basic notations in (5), we have

MVI-structure of the original PDHG

$$\begin{aligned}&u^{k+1} \in \mathcal{U}, \quad \theta(u)-\theta(u^{k+1})\\ &\quad+(u-u^{k+1})^T\big\{F(u^{k+1}) + Q(u^{k+1}-u^k)\big\} \quad (10)\\ &\geq 0, \quad \forall u \in \mathcal{U},\end{aligned}$$

where

$$Q = \begin{pmatrix} rI & A^T \\ 0 & sI \end{pmatrix}. \quad (11)$$

2.3 A basic assumption

To ensure the convergence of the new method, we assume the matrix Q defined in (11) satisfies $Q^T+Q \succ 0$, in other words, the proximal parameters r and s satisfy

$$r>0, \quad s>0 \quad \text{and} \quad rs > \rho(A^TA)/4.$$

In such a case, it will provide a wider choice of the regularization parameter r and s compared with the condition $rs > \rho(A^TA)$ for CPPA and PDHG.

3 Algorithm

In this section, our goal is to propose an acceletable convergence method under the framework of prediction-correction method in [21], the original PDHG will be regarded as a prediction step in our new algorithm. Moreover, inspired by Newton's method and alternating direction method with Gaussian back substitution, we establish our novel scheme by a standard line search method in optimization.

3.1 Motivation

As analyzed in [4,15], CPPA has a good structure and it is convergent provided $rs > \rho(A^TA)$. Naturally, we want to design a acceletable CPPA-based algorithm by a P-C way. To begin with, we regard CPPA as a prediction step, which has a form

$$\begin{cases} \tilde{x}^k = \arg\min\{\Phi(x,y^k) + \frac{r}{2}\|x-x^k\|_2^2 \mid x \in \mathcal{X}\},\\ \tilde{y}^k = \arg\max\{\Phi(2\tilde{x}^k-x^k,y) - \frac{s}{2}\|y-y^k\|_2^2 \mid y \in \mathcal{Y}\}. \end{cases} \quad (12)$$

By the similar analysis in section 2.2, the predictor $\tilde{u}^k$ satisfies the following variational inequality:

$$\theta(u)-\theta(\tilde{u}^k)+(u-\tilde{u}^k)^T\big\{F(\tilde{u}^k)+H(\tilde{u}^k-u^k)\big\} \geq 0, \quad (13)$$

where

$$H = \begin{pmatrix} rI & A^T \\ A & sI \end{pmatrix}.$$

Suppose $u^* \in \mathcal{U}^*$ is a solution point, thanks to H is symmetric and positive definite when $rs > \rho(A^TA)$, then by the norm equivalence principle in $\Re^n$, the saddle point problem (1) equals

$$\min\big\{\frac{1}{2}\|u-u^*\|_H^2 \mid u \in \mathcal{U}\big\}. \quad (14)$$

If $u^k = \tilde{u}^k$, according to (6) and (13), u^k would be a solution point, and hence we define a convergence indicator

$$e(u^k, \tilde{u}^k) = u^k - \tilde{u}^k, \tag{15}$$

which has a same effect of determining convergence with $\nabla f(x^k)$ for unconstrained convex optimization

$$\min\big\{f(x)\big\}.$$

Therefore, in the following sections, we assume $u^k \neq \tilde{u}^k$. It is well known that Newton's method has a second-order convergence rate, inspired such a fact, we want to propose a Newton-like correction step. In general, the classical Newton's method (see details in, e.g., [3]) has an iterative scheme

$$u^{k+1} = u^k - \alpha_k(\nabla^2 f(u^k))^{-1}\nabla f(u^k),$$

where α_k denotes the step size. Noting that

$$\nabla^2(\frac{1}{2}\|u - u^*\|_H^2) = H$$

in (14), and thus we want to propose the following Newton-like correction step

$$u^{k+1} = u^k - \alpha_k H^{-T} e(u^k, \tilde{u}^k).$$

However, it is expensive to calculate H^{-1}. Recalling Q defined in (11), is an upper triangular matrix and its diagonal parts are rI and sI, and hence we have

$$Q^{-T} = \begin{pmatrix} \frac{1}{r}I & 0 \\ -\frac{1}{rs}A & \frac{1}{s}I \end{pmatrix}.$$

Because of the simple realization of Q^{-T}, in this paper, we propose the following weakly Newton-like correction step

$$u^{k+1} = u^k - \alpha_k Q^{-T} e(u^k, \tilde{u}^k). \tag{16}$$

In fact, (16) can be rewritten as

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

which enjoys a similar correction step with the alternating direction method with Gaussian back substitution proposed in [18]. Therefore, we also regard the Newton-like correction step (16) as a special Gaussian back substitution step of the original PDHG.

3.2 Two intrinsic PDHG-based ascent directions

According to (10), the original PDHG, as the prediction step, it satisfies the following variational inequality:

$$\theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T\big\{F(\tilde{u}^k) + Q(\tilde{u}^k - u^k)\big\} \geq 0. \tag{17}$$

Substituting $u = u^*$ into inequality (17) and combing with the structure of F in (5), we have

$$\begin{aligned}
&(\tilde{u}^k - u^*)^T Q(u^k - \tilde{u}^k) \\
&\geq \theta(\tilde{x}^k) - \theta(x^*) + (\tilde{u}^k - u^*)^T F(\tilde{u}^k) \\
&= \theta(\tilde{x}^k) - \theta(x^*) + (\tilde{u}^k - u^*)^T F(u^*) \\
&\geq 0.
\end{aligned}$$

Using the fact $\tilde{u}^k - u^* = u^k - u^* - (u^k - \tilde{u}^k)$, it follows that

$$\begin{aligned}
&\langle u^k - u^*, Q(u^k - \tilde{u}^k)\rangle \\
&= \big\langle \nabla(\frac{1}{2}\|(u - u^*)\|_2^2)\Big|_{u=u^k}, Q(u^k - \tilde{u}^k)\big\rangle \\
&\geq \langle u^k - \tilde{u}^k, Q(u^k - \tilde{u}^k)\rangle \\
&= \frac{1}{2}\|u^k - \tilde{u}^k\|_{Q^T+Q}^2,
\end{aligned} \tag{18}$$

where $\langle\cdot\rangle$ denotes the inner product. This shows that $Q(u^k - \tilde{u}^k)$ is an ascent direction of $\frac{1}{2}\|u - u^*\|_2^2$ at the point u^k. In fact, there has been some work based on such a direction, we refer the readers to [15], see details in Algorithm 1 (in which $\theta = 0$, $H = I$ and $M = Q$), and we denote such a method as PDHG-HeY. In this paper, we consider its 'twins-direction'

$$d(u^k, \tilde{u}^k) := Q^{-T}(u^k - \tilde{u}^k). \tag{19}$$

Taking the nonsingularity of Q into consideration, we use the QQ^T-norm, which is defined by

$$\|x - y\|_{QQ^T} := \|Q^T(x - y)\|_2 = \sqrt{(x - y)^T QQ^T(x - y)},$$

to replace the Euclidean norm. In such a case, the above inequality (18) can be rewritten as

$$\begin{aligned}
&\langle u^k - u^*, Q(u^k - \tilde{u}^k)\rangle \\
&= \langle QQ^T(u^k - u^*), Q^{-T}(u^k - \tilde{u}^k)\rangle \\
&= \big\langle \nabla(\frac{1}{2}\|Q^T(u - u^*)\|^2)\Big|_{u=u^k}, d(u^k, \tilde{u}^k)\big\rangle \\
&\geq \frac{1}{2}\|u^k - \tilde{u}^k\|_{Q^T+Q}^2.
\end{aligned}$$

This indicates the direction $d(u^k, \tilde{u}^k)$ (19) is an ascent direction of $\frac{1}{2}\|u - u^*\|_{QQ^T}^2$ at the point u^k.

Two intrinsic PDHG-based ascent directions

- Direction 1: $Q(u^k - \tilde{u}^k)$.
- Direction 2: $Q^{-T}(u^k - \tilde{u}^k)$.

3.3 The reversible primal-dual hybrid gradient method

Based on the ascent direction (19), now we establish our new algorithm by a standard line search method.

According to the line search method in numerical optimization, see details in, e.g., [22]. In order to determine the step size, we do a linear search (16) to maximize the quadratic contraction measure

$$\|u^k - u^*\|^2_{QQ^T} - \|u^{k+1} - u^*\|^2_{QQ^T}$$

along the descend direction $-d(u^k, \tilde{u}^k)$. Due to

$$\begin{aligned}
&\|u^{k+1} - u^*\|^2_{QQ^T} \\
&= \|u^k - \alpha Q^{-T}(u^k - \tilde{u}^k) - u^*\|^2_{QQ^T} \\
&= \|u^k - u^*\|^2_{QQ^T} - 2\alpha(u^k - u^*)^T Q(u^k - \tilde{u}^k) \\
&\quad + \alpha^2\|u^k - \tilde{u}^k\|^2_2 \\
&\overset{(18)}{\leq} \|u^k - u^*\|^2_{QQ^T} - \alpha\|u^k - \tilde{u}^k\|^2_{Q^T+Q} \\
&\quad + \alpha^2\|u^k - \tilde{u}^k\|^2_2.
\end{aligned}$$

Therefore, we obtain the inequality

$$\begin{aligned}
&\|u^k - u^*\|^2_{QQ^T} - \|u^{k+1} - u^*\|^2_{QQ^T} \\
&\geq q(\alpha) := \alpha\|u^k - \tilde{u}^k\|^2_{Q^T+Q} - \alpha^2\|u^k - \tilde{u}^k\|^2_2.
\end{aligned}$$

By maximizing the lower bound quadratic contraction function $q(\alpha)$, it follows that

$$\alpha_k^* = \frac{\|u^k - \tilde{u}^k\|^2_{Q^T+Q}}{2\|u^k - \tilde{u}^k\|^2_2}.$$

On the one hand, we consider two norms $\|\cdot\|_{Q^T+Q}$ and $\|\cdot\|_2$, according to the principle of norm equivalence in $\Re^n$, we know there exists a positive number α satisfies $\alpha_k^* > \alpha$, it will ensure the lower-boundary of the step size α_k^*. On the other hand, noting $q(\alpha)$ is a lower bounded quadratic contraction function, similar with [21], we introduce a relaxation factor $\gamma \in (0, 2)$. Therefore, we obtain the following P-C iterative scheme:

Prediction step of RPDHG

$$\begin{cases} \tilde{x}^k = \arg\min\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|_2^2 \mid x \in \mathcal{X}\}, \\ \tilde{y}^k = \arg\max\{\Phi(\tilde{x}^k, y) - \frac{s}{2}\|y - y^k\|_2^2 \mid y \in \mathcal{Y}\}. \end{cases} \tag{20}$$

Correction step of RPDHG

$$u^{k+1} = u^k - \gamma\alpha_k^* Q^{-T}(u^k - \tilde{u}^k), \tag{21}$$

where

$\gamma \in (0, 2)$,

$$\alpha_k^* = \frac{\|u^k - \tilde{u}^k\|^2_{Q^T+Q}}{2\|u^k - \tilde{u}^k\|_2^2} \quad \text{and} \quad Q^{-T} = \begin{pmatrix} \frac{1}{r}I & 0 \\ -\frac{1}{rs}A & \frac{1}{s}I \end{pmatrix}.$$

Convergence condition of RPDHG

$r > 0, \quad s > 0 \quad \text{and} \quad rs > \rho(A^T A)/4.$

As can been seen, Q^{-T} has a simple structure, whose components I and A are same as the matrix Q, and hence it can be regarded as a reversible version of Q. Therefore, we name the new method (20)-(21) as reversible primal-dual hybrid gradient method (RPDHG). Generally speaking, we suggest to take the relaxation factor $\gamma \in (0.8, 1.5)$ in the real computations.

4 Convergence

In this section, we establish the convergence analysis of proposed method (20)-(21). Similar with [7,8], we show the new method has a global convergence and a worst-case $\mathcal{O}(1/t)$ convergence rate in an ergodic sense.

4.1 Global convergence

The convergence analysis is based on the following lemma.

Lemma 2 *Let $\{u^k\}$ be the sequence generated by* (20) *and* (21) *for problem* (1)*, and $\gamma \in (0, 2)$ is a relaxation factor. Then there exists $\sigma > 0$, it satisfies the Féjer monotone property:*

$$\begin{aligned}
&\|u^k - u^*\|^2_{QQ^T} - \|u^{k+1} - u^*\|^2_{QQ^T} \\
&\geq \sigma\gamma(2 - \gamma)\|u^k - \tilde{u}^k\|^2_{Q^T+Q}.
\end{aligned} \tag{22}$$

Proof By the definition of u^{k+1} in (21), we have

$$\begin{aligned}
&\|u^{k+1}-u^*\|^2_{QQ^T}\\
&=\|u^k-\gamma\alpha_k^*Q^{-T}(u^k-\tilde{u}^k)-u^*\|^2_{QQ^T}\\
&=\|u^k-u^*\|^2_{QQ^T}-2\gamma\alpha_k^*(u^k-u^*)^TQ(u^k-\tilde{u}^k)\\
&\quad+\gamma^2(\alpha_k^*)^2\|u^k-\tilde{u}^k\|_2^2\\
&\overset{(18)}{\leq}\|u^k-u^*\|^2_{QQ^T}-\gamma\alpha_k^*\|u^k-\tilde{u}^k\|^2_{Q^T+Q}\\
&\quad+\gamma^2(\alpha_k^*)^2\|u^k-\tilde{u}^k\|_2^2\\
&=\|u^k-u^*\|^2_{QQ^T}+\frac{\gamma(\gamma-2)}{2}\alpha_k^*\|u^k-\tilde{u}^k\|^2_{Q^T+Q}.
\end{aligned}\tag{23}$$

Moreover, since $\|\cdot\|_{Q^T+Q}$ and $\|\cdot\|_2$ are two norms. By the principle of norm equivalence in $\Re^n$, there exists $2\sigma>0$ such that

$$\alpha_k^*=\frac{\|u^k-\tilde{u}^k\|^2_{Q^T+Q}}{2\|u^k-\tilde{u}^k\|_2^2}\geq 2\sigma.\tag{24}$$

Plug (24) back into (23), which completes the proof of Lemma 2. □

Based on Lemma 2, we have the following global convergence theorem.

Theorem 1 *Let $\{u^k\}$ be the sequence generated by* (20) *and* (21) *for problem* (1)*, then for any relaxation factor $\gamma\in(0,2)$, we have $u^k\to u^\infty\in\mathcal{U}^*$.*

Proof According to inequality (22), the generated sequence $\{u^k\}$ satisfies

$$\begin{aligned}
&\|u^{k+1}-u^*\|^2_{QQ^T}\\
&\leq\|u^k-u^*\|^2_{QQ^T}-\sigma\gamma(2-\gamma)\|u^k-\tilde{u}^k\|^2_{Q^T+Q}\\
&\leq\|u^k-u^*\|^2_{QQ^T}\\
&\leq\cdots\\
&\leq\|u^0-u^*\|^2_{QQ^T}.
\end{aligned}$$

Consequently, the generated sequence $\{u^k\}$ is bounded. By adding (22) from $k=0,1,\cdots,\infty$, it follows that

$$\begin{aligned}
&\sum_{k=0}^{\infty}\sigma\gamma(2-\gamma)\|u^k-\tilde{u}^k\|^2_{Q^T+Q}\\
&\leq\sum_{k=0}^{\infty}(\|u^k-u^*\|^2_{QQ^T}-\|u^{k+1}-u^*\|^2_{QQ^T})\\
&\leq\|u^0-u^*\|^2_{QQ^T},
\end{aligned}$$

and hence we obtain

$$\lim_{k\to\infty}\|u^k-\tilde{u}^k\|^2_{Q^T+Q}=0.\tag{25}$$

This indicates the sequence $\{\tilde{u}^k\}$ is also bounded. By the compactness of the closed bounded set in $\Re^n$, there exists a subsequence $\{\tilde{u}^{k_j}\}$ converges to u^∞. Therefore, combing with (25), we have

$$\theta(u)-\theta(u^\infty)+(u-u^\infty)^TF(u^\infty)\geq 0,\quad\forall\,u\in\mathcal{U},$$

in (17), which means $u^\infty\in\mathcal{U}^*$ is a solution point. Then by (22), we also have

$$\begin{aligned}
&\|u^k-u^\infty\|^2_{QQ^T}-\|u^{k+1}-u^\infty\|^2_{QQ^T}\\
&\geq\sigma\gamma(\gamma-2)\|u^k-\tilde{u}^k\|^2_{Q^T+Q}.
\end{aligned}\tag{26}$$

Meanwhile, according to Féjer monotonicity (26), the sequence $\{\tilde{u}^k\}$ cannot have any other accumulation point. Therefore, by (25), we have $\lim\limits_{k\to\infty}u^k=\lim\limits_{k\to\infty}\tilde{u}^k=u^\infty$. This completes the proof. □

4.2 Convergence rate

Similar with [7,8], first, we need to define a suitable approximate-solution of the variational inequality (6). It is easy to see that if $\tilde{u}$ satisfies

$$\tilde{u}\in\mathcal{U},\ \ \theta(u)-\theta(\tilde{u})+(u-\tilde{u})^TF(\tilde{u})\geq 0,\ \ \forall\,u\in\mathcal{U},$$

then $\tilde{u}$ will be a solution point of (6). Recall the operator $F(\cdot)$ in (5) is linear with a skew-symmetric structure, we have

$$(u-\tilde{u})^TF(\tilde{u})\equiv(u-\tilde{u})^TF(u),$$

and consequently the solution point $\tilde{u}$ can be characterized by

$$\tilde{u}\in\mathcal{U},\ \ \theta(u)-\theta(\tilde{u})+(u-\tilde{u})^TF(u)\geq 0,\ \ \forall\,u\in\mathcal{U}.$$

For a given $\epsilon>0$, as described in [10], we say $\tilde{u}$ is a ϵ-approximate solution of (6) means

$$\tilde{u}\in\mathcal{U},\ \theta(u)-\theta(\tilde{u})+(u-\tilde{u})^TF(u)\geq-\epsilon,\ \forall\,u\in\mathcal{D}_{\tilde{u}},\tag{27}$$

where

$$\mathcal{D}_{\tilde{u}}=\{u\in\mathcal{U}\ \big|\ \|u-\tilde{u}\|\leq 1\}.$$

Multiplying both sides by -1 in (27), we also have

$$\tilde{u}\in\mathcal{U},\quad\theta(\tilde{u})-\theta(u)+(\tilde{u}-u)^TF(u)\leq\epsilon,\quad\forall\,u\in\mathcal{D}_{\tilde{u}}.$$

Next, for the proposed scheme (20)-(21), our goal is to show after t iteration times, we can find an approximate solution $\tilde{u}_t\in\mathcal{U}$ satisfies

$$\sup_{u\in\mathcal{D}_{\tilde{u}_t}}\big\{\theta(\tilde{u}_t)-\theta(u)+(\tilde{u}_t-u)^TF(u)\big\}\leq\mathcal{O}(1/t).$$

Lemma 3 *Let $\{u^k\}$ be the sequence generated by* (20) *and* (21) *for problem* (1)*. Then for any relaxation factor $\gamma \in (0,2)$, we have*

$$\begin{aligned}&\alpha_k^*\{\theta(u)-\theta(\tilde{u}^k)+(u-\tilde{u}^k)^T F(\tilde{u}^k)\}\\ &\geq \frac{1}{2\gamma}\{\|u-u^{k+1}\|^2_{QQ^T}-\|u-u^k\|^2_{QQ^T}\}\\ &\quad+\frac{2-\gamma}{4}\alpha_k^*\|u^k-\tilde{u}^k\|^2_{Q^T+Q}.\end{aligned}\tag{28}$$

Proof At first, according to (17) and (21), we have

$$\begin{aligned}&\theta(u)-\theta(\tilde{u}^k)+(u-\tilde{u}^k)^T F(\tilde{u}^k)\\ &\geq \frac{1}{\gamma\alpha_k^*}(u-\tilde{u}^k)^T QQ^T(u^k-u^{k+1}),\end{aligned}$$

Since Q is nonsingular, we know QQ^T is symmetric and positive definite. By using a basic identity

$$\begin{aligned}&(a-b)QQ^T(c-d)\\ &=\frac{1}{2}\{\|a-d\|^2_{QQ^T}-\|a-c\|^2_{QQ^T}\}\\ &\quad+\frac{1}{2}\{\|c-b\|^2_{QQ^T}-\|d-b\|^2_{QQ^T}\}\end{aligned}$$

and set

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

we obtain

$$\begin{aligned}&(u-\tilde{u}^k)^T QQ^T(u^k-u^{k+1})\\ &=\frac{1}{2}\{\|u-u^{k+1}\|^2_{QQ^T}-\|u-u^k\|^2_{QQ^T}\}\\ &\quad+\frac{1}{2}\{\|u^k-\tilde{u}^k\|^2_{QQ^T}-\|u^{k+1}-\tilde{u}^k\|^2_{QQ^T}\}.\end{aligned}\tag{29}$$

For the second term of (29), we have

$$\begin{aligned}&\frac{1}{2}\{\|u^k-\tilde{u}^k\|^2_{QQ^T}-\|u^{k+1}-\tilde{u}^k\|^2_{QQ^T}\}\\ &=\frac{1}{2}\|u^k-\tilde{u}^k\|^2_{QQ^T}\\ &\quad-\frac{1}{2}\|u^k-\gamma\alpha_k^*Q^{-T}(u^k-\tilde{u}^k)-\tilde{u}^k\|^2_{QQ^T}\\ &=\frac{1}{2}\{-(\gamma\alpha_k^*)^2\|u^k-\tilde{u}^k\|_2^2+\gamma\alpha_k^*\|u^k-\tilde{u}^k\|^2_{Q^T+Q}\}\\ &=\frac{1}{2}\Big\{\frac{(2-\gamma)\gamma}{2}\alpha_k^*\|u^k-\tilde{u}^k\|^2_{Q^T+Q}\Big\},\end{aligned}\tag{30}$$

and thus we have

$$\begin{aligned}&\theta(u)-\theta(\tilde{u}^k)+(u-\tilde{u}^k)^T F(\tilde{u}^k)\\ &\overset{(29)}{\geq}\frac{1}{\gamma\alpha_k^*}\big[\frac{1}{2}\{\|u-u^{k+1}\|^2_{QQ^T}-\|u-u^k\|^2_{QQ^T}\}\\ &\quad+\frac{1}{2}\{\|u^k-\tilde{u}^k\|^2_{QQ^T}-\|u^{k+1}-\tilde{u}^k\|^2_{QQ^T}\}\big]\\ &\overset{(30)}{=}\frac{1}{\gamma\alpha_k^*}\Big\{\frac{1}{2}\{\|u-u^{k+1}\|^2_{QQ^T}-\|u-u^k\|^2_{QQ^T}\}\\ &\quad+\frac{2-\gamma}{4}\gamma\alpha_k^*\|u^k-\tilde{u}^k\|^2_{Q^T+Q}\Big\}\\ &=\frac{1}{2\gamma\alpha_k^*}\{\|u-u^{k+1}\|^2_{QQ^T}-\|u-u^k\|^2_{QQ^T}\}\\ &\quad+\frac{2-\gamma}{4}\|u^k-\tilde{u}^k\|^2_{Q^T+Q}.\end{aligned}$$

This completes the proof of Lemma 3. □

According to the inequality (28) and the monotonicity of the operator F, we have

$$\begin{aligned}&\alpha_k^*\{\theta(u)-\theta(\tilde{u}^k)+(u-\tilde{u}^k)^T F(u)\}\\ &\geq\alpha_k^*\{\theta(u)-\theta(\tilde{u}^k)+(u-\tilde{u}^k)^T F(\tilde{u}^k)\}\\ &\geq\frac{1}{2\gamma}\{\|u-u^{k+1}\|^2_{QQ^T}-\|u-u^k\|^2_{QQ^T}\}.\end{aligned}\tag{31}$$

Based on the inequality (31), we have the following theorem.

Theorem 2 *Let $\{u^k\}$ be the sequence generated by* (20) *and* (21) *for problem* (1)*. Then for any positive integer t and relaxation factor $\gamma\in(0,2)$, we have*

$$\theta(\tilde{u}_t)-\theta(u)+(\tilde{u}_t-u)^T F(u)\leq\frac{1}{2\gamma\Upsilon_t}\|u-u_0\|^2_{QQ^T},\tag{32}$$

where

$$\tilde{u}_t=\frac{1}{\Upsilon_t}(\sum_{k=0}^{t}\alpha_k^*\tilde{u}^k)\quad\textit{and}\quad\Upsilon_t=\sum_{i=0}^{t}\alpha_k^*.\tag{33}$$

Proof To begin with, by adding the inequality (31) from $k=0,\ \cdots,\ t$, it follows that

$$\sum_{k=0}^{t}\alpha_k^*\{\theta(\tilde{u}^k)-\theta(u)+(\tilde{u}^k-u)^T F(u)\}\leq\frac{1}{2\gamma}\|u-u^0\|^2_{QQ^T}.$$

Noting the notations in (33), by dividing Υ_t both sides, we have

$$\begin{aligned}&\Big\{\frac{1}{\Upsilon_t}\sum_{k=0}^{t}\alpha_k^*\theta(\tilde{u}^k)\Big\}-\theta(u)+(\tilde{u}^k-u)^T F(u)\\ &\leq\frac{1}{2\gamma\Upsilon_t}\|u-u^0\|^2_{QQ^T}.\end{aligned}\tag{34}$$

Meanwhile, according to the convexity of $\theta(u)$, we know

$$\theta(\tilde{u}_t) \leq \frac{1}{\Upsilon_t}\Big\{\sum_{k=1}^{t}\alpha_k^*\theta(\tilde{u}^k)\Big\}.$$

By substituting it in (34), the assertion of this theorem follows directly. □

With the help of the preceding lemma and theorem, now we can show the convergence rate of the proposed scheme (20)-(21). According to the conclusion in Lemma 2, we know $\alpha_k^* \geq 2\sigma$. Therefore, for the right term in (32), we have

$$\begin{aligned}\frac{1}{2\gamma\Upsilon_t}\|u-u_0\|^2_{QQ^T} &= \frac{1}{2\gamma\sum_{k=0}^{t}\alpha_k^*}\|u-u_0\|^2_{QQ^T}\\ &\leq \frac{1}{2\gamma(t+1)2\sigma}\|u-u_0\|^2_{QQ^T}.\end{aligned}$$

If we set $b = \sup_{u\in\mathcal{D}_{\tilde{u}_t}}\{\|u-u_0\|^2_{QQ^T}\}$, it is easy to see that

$$\theta(\tilde{u}_t)-\theta(u)+(\tilde{u}_t-u)^TF(u) \leq \frac{1}{2\gamma(t+1)2\sigma}b = \mathcal{O}(\frac{1}{t}),$$

and consequently a $\mathcal{O}(1/t)$ convergence rate in an ergodic sense is established.

5 Numerical experiments

We conduct some numerical experiments to evaluate the performance of the novel method (20)-(21) in this section. It contains three examples of image inpainting.

5.1 Image restoration problems

As analyzed in [4,6,23,27], some classical image restoration problems with TV regularization can be reformulated as the following min-max problem:

$$\min_x\max_{y\in\mathcal{Y}}\big\{\Phi(x,y) = y^T\nabla x + \frac{\mu}{2}\|Bx-z\|_2{}^2\big\},$$

where $\mu>0$ is a constant balancing the data-fidelity and TV regularization terms; $\mathcal{Y}$ is the Cartesian product of some unit balls in $\Re^2$; ∇ denotes a discrete gradient operator and B is the matrix representation of image processing operator. For example, if $B=I$, it will correspond to TV image denosing. At first, we need to solve two decomposed subproblems in RPDHG scheme (20)-(21). For x-subproblem in (20),

$$\begin{aligned}\tilde{x}^k &= \arg\min_x\{\Phi(x,y^k)+\frac{r}{2}\|x-x^k\|_2^2\}\\ &= \arg\min_x\{(y^k)^T\nabla x+\frac{\mu}{2}\|Bx-z\|_2{}^2+\frac{r}{2}\|x-x^k\|_2^2\}\\ &= (\mu B^TB+rI)^{-1}(\mu B^TBz+rx^k-\nabla^Ty^k).\end{aligned}$$

It equals to solving a linear equations system. For image denosing and deblurring, as described in [13], the matrix μB^TB+rI would be a block circulant with circulant blocks matrix (BCCB) under the image periodic boundary condition, we can solve it by fast Fourier transform. For image inpainting model, μB^TB+rI would be a diagonal matrix. It remains to solve the y-subproblem in (20),

$$\begin{aligned}\tilde{y}^k &= \arg\max_{y\in\mathcal{Y}}\big\{\Phi(\tilde{x}^k,y)-\frac{s}{2}\|y-y^k\|_2^2\big\}\\ &= \arg\min_{y\in\mathcal{Y}}\big\{-\Phi(\tilde{x}^k,y)+\frac{s}{2}\|y-y^k\|_2^2\big\}\\ &= \mathbf{P}_{\mathcal{Y}}(y^k+\frac{1}{s}\nabla\tilde{x}^k),\end{aligned}$$

where $\mathbf{P}_{\mathcal{Y}}(\cdot)$ is a projection operator on set $\mathcal{Y}$. Therefore, the original PDHG (3) has an iterative scheme:

$$\begin{cases}x^{k+1} = (\mu B^TB+rI)^{-1}(\mu B^TBz+rx^k-\nabla^Ty^k),\\ y^{k+1} = \mathbf{P}_{\mathcal{Y}}(y^k+\frac{1}{s}\nabla x^{k+1}).\end{cases}$$

And the reversible PDHG (20)-(21) thus has a form:

$$\begin{cases}\begin{cases}\tilde{x}^k &= (\mu B^TB+rI)^{-1}(\mu B^TBz+rx^k-\nabla^Ty^k),\\ \tilde{y}^k &= \mathbf{P}_{\mathcal{Y}}(y^k+\frac{1}{s}\nabla\tilde{x}^k),\end{cases}\\ \begin{cases}x^{k+1} &= x^k-\gamma\alpha_k^*\frac{1}{r}(x^k-\tilde{x}^k),\\ y^{k+1} &= y^k-\gamma\alpha_k^*\{\frac{1}{rs}\nabla(x^k-\tilde{x}^k)+\frac{1}{s}(y^k-\tilde{y}^k)\},\end{cases}\end{cases}$$

in which

$$\begin{aligned}\alpha_k^* &= \frac{\|u^k-\tilde{u}^k\|^2_{Q^T+Q}}{2\|u^k-\tilde{u}^k\|_2^2}\\ &= \frac{q^k}{\|x^k-\tilde{x}^k\|_2^2+\|y^k-\tilde{y}^k\|_2^2},\end{aligned}$$

and

$$q^k = r\|x^k-\tilde{x}^k\|_2^2+s\|y^k-\tilde{y}^k\|_2^2-(y^k-\tilde{y}^k)^T\nabla(x^k-\tilde{x}^k).$$

We also give the specific iterative form of PDHG-HeY proposed in [15], which enjoys a form:

$$\begin{cases}\begin{cases}\tilde{x}^k &= (\mu B^TB+rI)^{-1}(\mu B^TBz+rx^k-\nabla^Ty^k),\\ \tilde{y}^k &= \mathbf{P}_{\mathcal{Y}}(y^k+\frac{1}{s}\nabla\tilde{x}^k),\end{cases}\\ \begin{cases}x^{k+1} &= x^k-\gamma\beta_k^*\big\{r(x^k-\tilde{x}^k)-\nabla^T(y^k-\tilde{y}^k)\big\},\\ y^{k+1} &= y^k-\gamma\beta_k^*\big\{s(y^k-\tilde{y}^k)\big\},\end{cases}\end{cases}$$

where

$$\beta_k^* = \frac{q^k}{\|r(x^k-\tilde{x}^k)-\nabla^T(y^k-\tilde{y}^k)\|_2^2+\|s(y^k-\tilde{y}^k)\|_2^2}.$$

Fig. 1 Original image, missing image, and image recovered from RPDHG ($r = 5$ and $s = 1.2$) with 199 iterations.

Fig. 2 Original image, missing image, and image recovered from RPDHG ($r = 5$ and $s = 1.2$) with 289 iterations.

Fig. 3 Original image, missing image, and image recovered from RPDHG ($r = 5$ and $s = 1.2$) with 171 iterations.

Next, taking TV image inpainting as an example, we adpot some stop rules including iteration relative error (Itr-RE) and relative error (RE), which are defined as

$$\text{Itr-RE(k)} = \frac{\|\tilde{u}^k - u^k\|_2^2}{\|u^k\|_2^2} \quad \text{and} \quad \text{RE(k)} = \frac{\|x^k - x^*\|_2^2}{\|x^*\|_2^2}$$

respectively, where u^k and $\tilde{u}^k$ are k-th iteration point and prediction point, x^* is the good image. We also compare different methods by signal-to-noise ratio value (SNR). The reason why we take the stop rule by Itr-RE is (15), and due to the simplification of the model, the relative error (RE) would have a lower bound.

In our experiments, we take $\mu = 500$. Noting that $\rho(\nabla^T\nabla) = 8$, referring to [2,9,25,27], we adopt $s = 8.1/r$ for PDHG. By [15] and a lot of numerical attempts, we take $s = 7/r$ and $\gamma = 1.2$ in PDHG-HeY to

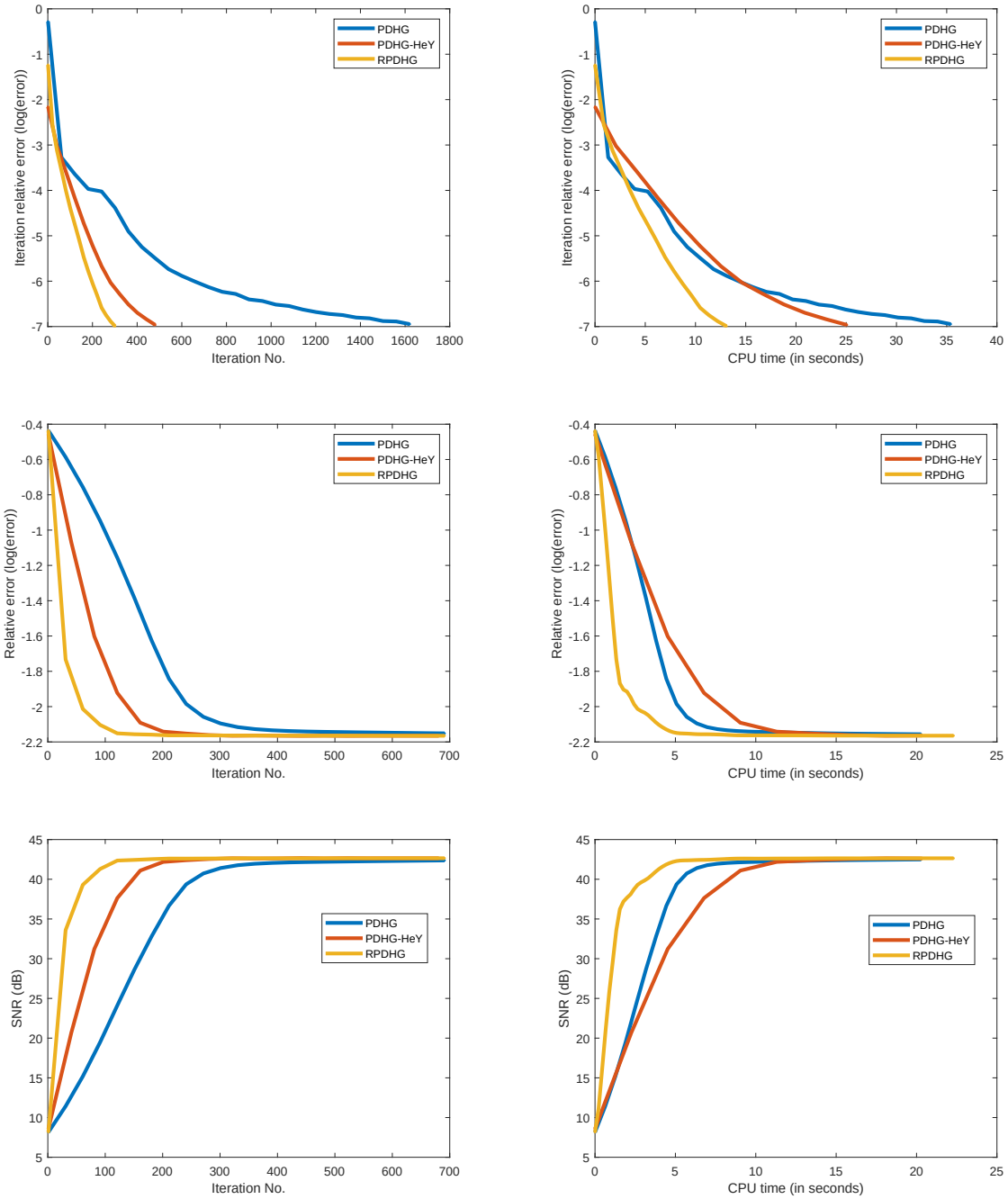

Fig. 4 The implementation of different evaluation indicators with iterations and CPU time for Figure 1.

obtain an efficient numerical performance. Finally, we take $s = 6/r$ in RPDHG and use $\gamma = 1$ for simplification.

5.2 Comparison

This subsection compares the RPDHG (20)-(21) to the original PDHG and PDHG-HeY (Algorithm 1) in [15], we show their implementation for different evaluation indicators with iterations and CPU time.

To begin with, we need to determine an optimal r for the above methods. We list the first iteration to the stop rule Tol = Itr-RE$(k) < 10^{-6}$ for different r in Tables 1, 2 and 3. As we can see, $r = 80$ and $r = 5$ have a better numerical performance for PDHG and RPDHG respectively; and $r = 2$ is more efficient for PDHG-HeY.

Table 1 The iteration number to stop rule Tol = Itr-RE(k) $< 10^{-6}$ for the original PDHG.

	Figure 1		Figure 2		Figure 3	
r	Iterations	log(Itr-RE)	Iterations	log(Itr-RE)	Iterations	log(Itr-RE)
$r = 5$	1599	-6.001	1875	-6.0042	1165	-6.0020
$r = 10$	1310	-6.0102	1503	-6.0045	888	-6.0038
$r = 50$	690	-6.0089	929	-6.0072	490	-6.0012
$r = 60$	664	-6.0092	856	-6.0052	474	-6.0196
$r = 70$	658	-6.0040	820	-6.0012	479	-6.0000
$r = 75$	648	-6.0073	828	-6.0080	475	-6.0089
$r = 80$	646	-6.0043	821	-6.0060	484	-6.0000
$r = 85$	653	-6.0085	833	-6.0006	490	-6.0038
$r = 90$	648	-6.0038	826	-6.0012	505	-6.0024
$r = 100$	652	-6.0096	851	-6.0014	540	-6.0039
$r = 120$	706	-6.0014	890	-6.0117	561	-6.0179

Table 2 The iteration number to stop rule Tol = Itr-RE(k) $< 10^{-6}$ for the PDHG-HeY with $\gamma = 1.2$.

	Figure 1		Figure 2		Figure 3	
r	Iterations	log(Itr-RE)	Iterations	log(Itr-RE)	Iterations	log(Itr-RE)
$r = 1$	340	-6.0002	600	-6.0022	355	-6.0018
$r = 1.5$	288	-6.0013	468	-6.0178	279	-6.0009
$r = 2$	279	-6.0031	417	-6.0094	263	-6.0001
$r = 2.5$	295	-6.0033	436	-6.0011	272	-6.0049
$r = 3$	301	-6.0026	436	-6.0012	273	-6.0071
$r = 4$	414	-6.0008	522	-6.0026	401	-6.0011
$r = 5$	597	-6.0026	678	-6.0016	572	-6.0026
$r = 6$	803	-6.0022	861	-6.0009	765	-6.0010

Table 3 The iteration number to stop rule Tol = Itr-RE(k) $< 10^{-6}$ for the RPDHG with $\gamma = 1$.

	Figure 1		Figure 2		Figure 3	
r	Iterations	log(Itr-RE)	Iterations	log(Itr-RE)	Iterations	log(Itr-RE)
$r = 2$	536	-6.0009	616	-6.0027	513	-6.0012
$r = 3$	273	-6.0013	414	-6.0009	251	-6.0051
$r = 4$	212	-6.0086	299	-6.0065	186	-6.0051
$r = 5$	199	-6.0169	289	-6.0102	171	-6.0068
$r = 6$	212	-6.0069	328	-6.0055	192	-6.0017
$r = 8$	346	-6.0013	521	-6.0051	345	-6.0052
$r = 10$	671	-6.0052	941	-6.0019	665	-6.0023

Therefore, in the following, we only compare the performance of PDHG with $r = 80$ $s = 0.10125$, PDHG-HeY with $r = 2$ $s = 3.5$ $\gamma = 1.2$ and RPDHG with $r = 5$ $s = 1.2$ $\gamma = 1$.

For the first experiment, Figure 1 shows the visual result by using the proposed method RPDHG with 199 iterations. As can be seen, it has a good visual effect. As the Figure 4 vividly shows: the modified PDHG methods PDHG-HeY and RPDHG are more efficient with iterations; the RPDHG needs a less iterations to stop rule and it has a steeper sub-linear convergence rate numerically. It also need less iterations to a stable relative error and SNR value. As the same time, noting the proposed method needs some extra extrapolation, we also show the implementation for the change of Itr-RE, relative error and SNR value with CPU time. As illustrated in Figure 4, it is clear that the RPDHG costs a less CPU time to the original PDHG and PDHG-HeY, which indicates the RPDHG has an accelerated effect for TV image inpainting.

For the second and third experiments, Figure 2 and 3 show the visual result by using the novel method. Moreover, similar with the first experiment, we show the implementation of different evaluation indicators with iterations and CPU time in Figure 5 and Figure 6. As they vividly show, the RPDHG (20)-(21)

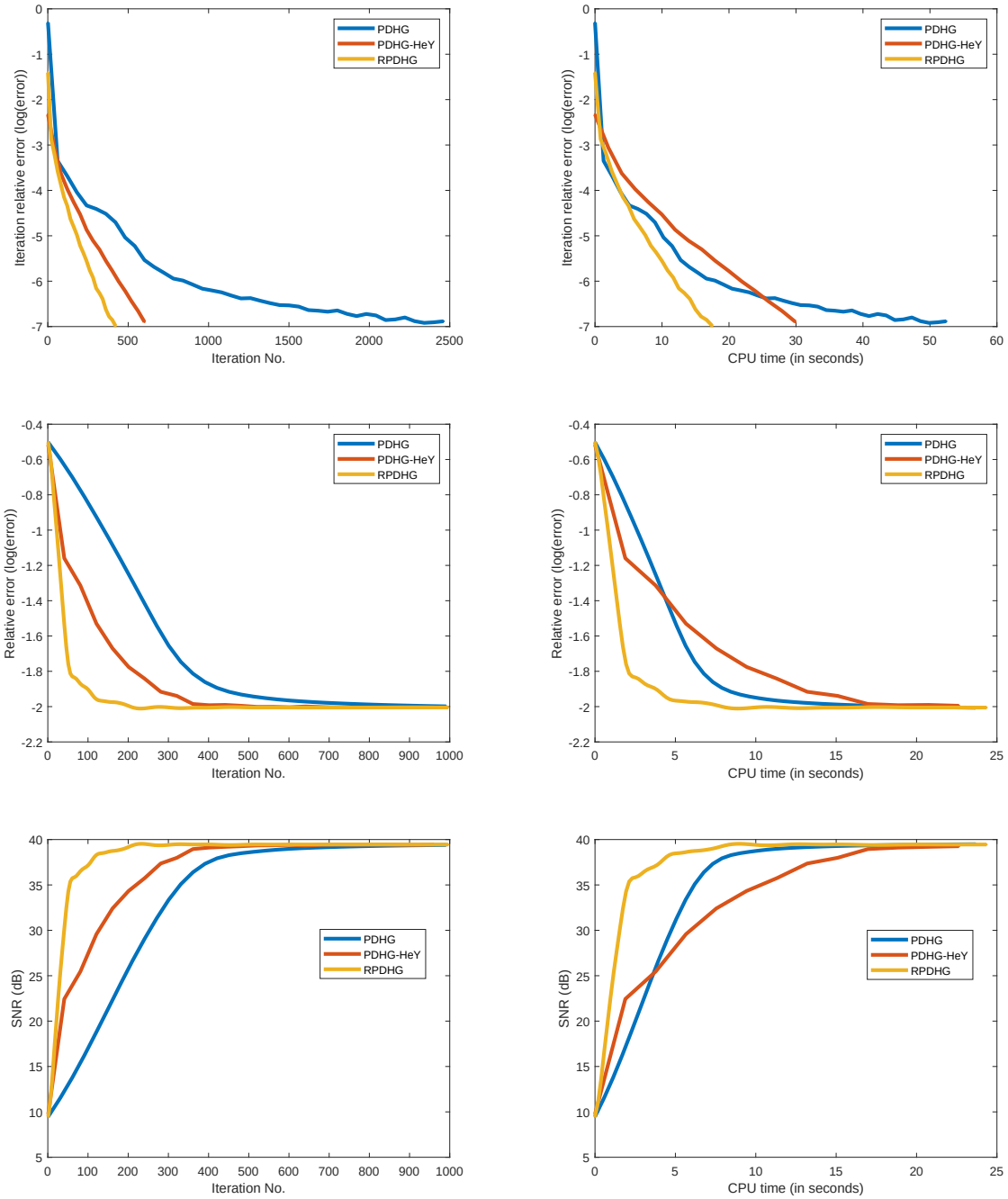

Fig. 5 The performance of different evaluation indicators with iterations and CPU time for Figure 2.

has a faster convergence rate both in iterations and CPU time. Therefore, the necessity of considering the RPDHG for other applications is meaningful.

6 Conclusions

Inspired by Newton's direction and alternating direction method with Gaussian back substitution, we present a reversible PDHG for saddle point problems. We also give its convergence analysis in a worst-case sense. On the one hand, compared with the original PDHG, the reversible PDHG is convergent only provided $rs > \rho(A^TA)/4$, which allows a bigger step size to potentially reduce required convergence iterations. On the other hand, the novel method provides an efficient ascent direction, and it can be regraded a variation of the classic line search method. Extensive numerical ex-

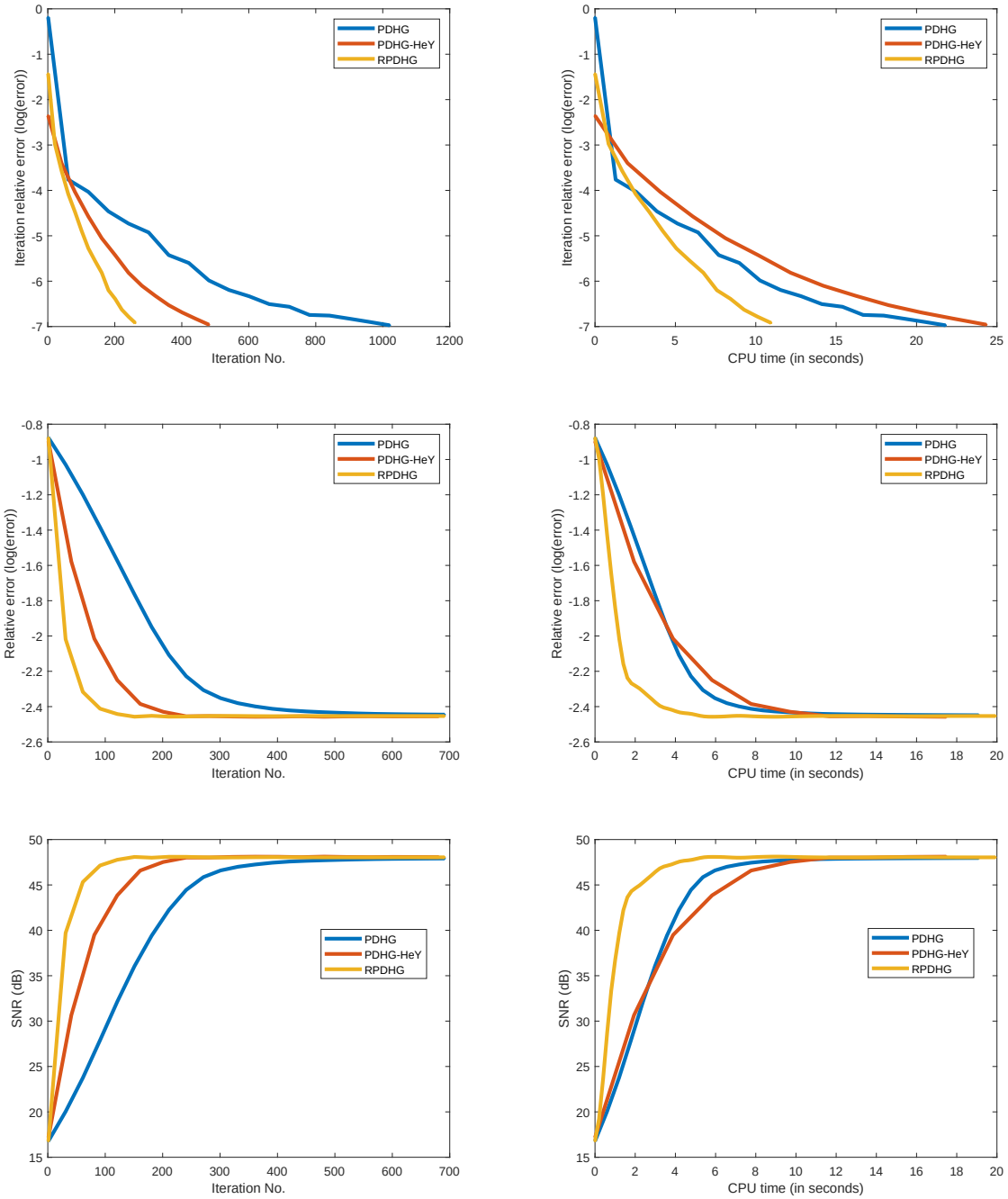

Fig. 6 The performance of different evaluation indicators with iterations and CPU time for Figure 3.

periments show that the proposed method has a faster convergence speed than the original PDHG for TV image inpainting. It would be interesting to know whether the proposed method (20)-(21) has a higher order convergence rate. We leave this as a topic for future work.

Acknowledgements

The author is greatly indebted to Professor Bingsheng He for many useful discussions and suggestions.

References

1. Arrow, K.J., Hurwicz, L., Uzawa, H.: Studies in linear and nonlinear programming. In: Chenery, H.B., Johnson,

S.M., Karlin, S., Marschak, T., Solow, R.M. (eds.) Stanford Mathematical Studies in the Social Science, vol. II. Stanford University Press, Stanford (1958)
2. Bonettini, S., Ruggiero, V.: On the convergence of primal-dual hybrid gradient algorithms for total variation image restoration. J. Math. Imaging Vis. 44, 236-253 (2012)
3. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University, (2004)
4. Chambolle, A., Pock, T.: A first-order primalCdual algorithms for convex problem with applications to imaging. J. Math. Imaging Vis. 40, 120-145 (2011)
5. Combettes, P.L., D$\tilde{u}$ng, D., V$\tilde{u}$, B.C.: Dualization of signal recovery problems. Set-Valued Anal. 18, 373-404 (2010)
6. Chan, T.F., Shen, J.: Mathematical models for local nontexture inpaintings. SIAM J. Appl Math. 62, 1019-1043 (2002)
7. Chen, C.H., Fu, X.L., He, B.S., Yuan, X.M.: On the iterative comlpexity of some projection methods for monotone linear variational inequalities. J. Optim. Theory Appl. 172, 914-928 (2017)
8. Cai, X.J., Gu, G.Y., He, B.S.: On the O(1/t) convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators. Comput. Optim. Appl., 57, 339-363 (2014)
9. Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal-dual algorithms for TV minimization. SIAM J. Imaging Sci. 3, 1015-1046 (2010)
10. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer Series in Operations Research, vol. I. Springer, New York (2003)
11. Goldfarb, D., Yin W.: Second-order cone programming methods for total variation-based image restoration. SIAM J. Sci. Comput., 27, 622-645 (2005)
12. Gu, G.Y., He, B.S., Yuan, X.M.: Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach. Comput. Optim. Appl. 59, 135-161 (2014)
13. Hansen, P., Nagy, J., OLeary, D.: Deblurring Images: Matrices, Spectra, and Filtering. Philadelphia: SIAM (2006)
14. He, B.S., You, Y.F., Yuan, X.M.: On the convergence of primal-dual hybrid gradient algorithm. SIAM J. Imaging Sci. 7, 2526-2537 (2014)
15. He, B.S., Yuan, X.M.: Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIAM J. Imaging Sci. 5, 119-149 (2012)
16. He, B.S., Ma, F., Yuan, X.M.: An agorithmic framework of generalized primal-dual hybrid gradient methods for saddle point problems. J. Math. Imaging Vis. 58, 279-293 (2017)
17. He, B.S., Ma, F., Yuan, X.M.: Optimally linearizing the alternating direction method of multipliers for convex programming. Comp. Opt. and Appl. 75(2), 361-388 (2020)
18. He, B.S., Tao, M., Yuan, X.M.: Alternating Direction Method with Gaussian Back Substitution for Separable Convex Programming. SIAM J. Optim. 22, 313-340(2012)
19. He, B.S., Ma, F., Yuan, X.M.: Optimal proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems. IMA J. Numer Anal. 40, 1188-1216 (2020)
20. He, B.S., Yuan, X.M., Zhang, W.X.: A customized proximal point algorithm for convex minimization with linear constraints, Comput. Optim. Appl., 56, 559-572 (2013)
21. He, B.S.: Constructing the first order splitting methods for convex optimization — A uniform framework using VI and PPA approaches. See lectures in `http://maths.nju.edu.cn/~hebma/Talk/Unified_Framework.pdf`
22. Nocedal, J., Wright, S.J.,: Numerical optimization. 2nd Edition, Springer, New York (2006)
23. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D 60, 227-238 (1992)
24. Rockafellar, R.T.,: Convex analysis. Princetin University Press, Princeton (1970)
25. Zhang, X., Burger, M., Osher, S.: A unified primal-dual algorithm framework based on Bregman iteration. J. Sci. Comput. 46(1), 20-46 (2010)
26. Xu, S.J., Yuan, J., He, B.S.: Proximal augmented Lagrangian method for convex optimization with linear inequality constraints. Optimization online, `http://www.optimization-online.org/DB_FILE/2019/05/7204.pdf`
27. Zhu, M., Chan, T.F.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM Report 08-34, UCLA, Los Angeles, CA (2008)