Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 16, No. 4, pp. 663–685, 2020

A REFINED PRIMAL-DUAL ALGORITHM FOR A SADDLE-POINT PROBLEM WITH APPLICATIONS TO IMAGING*

Min Li†, Xinxin Li and Zhongming Wu

Abstract: There are rich literatures on primal-dual algorithms for a saddle-point problem; and they have been demonstrated to be very efficient for some image restoration models with the total variation regularization. How to determine the step sizes is crucial for ensuring the efficiency of these primal-dual algorithms, and it has received intensive attention in the literature. This paper shows that the step sizes can be substantially refined if the output of a primal-dual algorithm at each iteration is corrected slightly. A modified primal-dual algorithm with refined step sizes is thus proposed. We prove rigorously the convergence of this new algorithm, and establish its worst-case convergence rate measured by the iteration complexity in ergodic and non-ergodic senses. The acceleration effectiveness of the refined step sizes is demonstrated by the TV image deblurring and inpainting problems.

Key words: *saddle-point problem, image restoration, primal-dual algorithm, refined step size, convergence rate*

Mathematics Subject Classification: *68U10, 90C25, 65K10*

---

1 Introduction

We consider to find a saddle point $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ of $\Phi(x, y)$, that is

$$\forall\, y \in \mathcal{Y} \qquad \Phi(x^*, y) \le \Phi(x^*, y^*) \le \Phi(x, y^*) \qquad \forall\, x \in \mathcal{X}, \tag{1.1}$$

where $\Phi(x, y) := \theta_1(x) - y^T Ax - \theta_2(y)$, $A \in \Re^{m\times n}$, $\mathcal{X} \subseteq \Re^n$, $\mathcal{Y} \subseteq \Re^m$ 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 saddle-point set of $\Phi(x, y)$ is assumed to be nonempty throughout our discussion. The problem (1.1) captures a variety of applications in different areas. For example, finding a saddle-point for the Lagrange function of the canonical convex minimization model with linear equality or inequality constraints is a special case of (1.1). In particular, a number of variational models with the total variation (TV) regularization in [27] arising in image restoration can also be reformulated as special cases of (1.1), see details in, e.g., [5,28,30,31].

*This work was supported by National Natural Science Foundation of China grants 11771078, 71661147004 and 11601183, Natural Science Foundation of Jiangsu Province grant BK20181258, Project 333 of Jiangsu Province grant BRA2018351, and Natural Science Foundation for Young Scientist of Jilin Province grant 20180520212JH.

†Corresponding author.

For many applications such as some mentioned TV models in image restoration, the functions θ_1 and θ_2 usually have some special properties and it deserves to explore them in algorithmic design. This has inspired a very active research topic of designing splitting algorithms where the functions θ_i are treated individually and thus the resulting subproblems are often easier, or even easy enough to have closed-form solutions. We focus on the primal-dual type of algorithms whose iterative schemes can be summarized as

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

where $\eta\in[-1,1]$ is a combination parameter; r and s are positive numbers[1]. As delineated in many papers such as $[3,5,30,31]$, $1/r$ and $1/s$ are the step sizes associated with the gradients (or subgradients) of θ_1 and θ_2, respectively. The case of (1.2) with $\eta=0$ was first investigated in [31]; and the so-called primal-dual hybrid gradient (PDHG) algorithm was proposed to solve some TV image restoration models which are all special cases of the model (1.1) where $\theta_1(x)\equiv 0$ and θ_2 is a quadratic term. For PDHG in [31], the parameters r and s were chosen in a specific form and the particular strategy was shown to be efficient for some specific TV denoising and deblurring models. But the convergence of PDHG with the recommended strategy of determining r and s has not yet been proved rigorously. Then, it was shown in [9] that the PDHG algorithm in [31] is related to the inexact Uzawa method in [1]; and under some restrictive conditions which essentially require r and s to be sufficiently large, the convergence of PDHG for TV denoising application was proved in [9]. Under other restrictions which also enforce the step sizes r and s to go to infinity, the convergence of a more general scheme of the PDHG algorithm was established in [3] in the context of subgradient method; and it was shown to be efficient for solving denoising and deblurring models where the data fidelity functions is defined as the generalized Kullback-Leibler divergence or the edge preserving removal of impulsive noise. In [5], the combination parameter η was extended to $[0,1]$; and it was shown that the primal-dual method has a worst-case $O(1/t)$ convergence rate where t is the iteration counter[2] and it can be accelerated by some acceleration techniques in the literature (e.g., [20,21]) so as to obtain an accelerated primal-dual algorithm with a worst-case convergence rate of $O(1/t^2)$. Also, it was shown in [5] that the primal-dual scheme (1.2) is closely related to many existing methods including the extrapolational gradient method [17,24], the Douglas-Rachford splitting method [8,18, 22], and the alternating direction method of multipliers [11]. In [15], the authors focused on a special case of (1.1) with $\theta_1(x)\equiv 0$ and $\theta_2(y)=\frac{\lambda}{2}\|By-b\|^2$, they proposed four types of algorithm both in prediction-correction framework, which share the prediction step as follows:

$$\left\{\begin{array}{rcl} \tilde{x}^k & := & \operatorname{Arg\,min}\big\{-(y^k)^TAx+\frac{r}{2}\|x-x^k\|^2 \,\big|\, x\in\mathcal{X}\big\}, \\ \bar{x}^k & := & \tilde{x}^k+\eta(\tilde{x}^k-x^k), \\ \tilde{y}^k & := & \operatorname{Arg\,max}\big\{-\frac{\lambda}{2}\|By-b\|^2-y^TA\bar{x}^k-\frac{s}{2}\|y-y^k\|^2 \,\big|\, y\in\mathcal{Y}\big\}, \end{array}\right. \tag{1.3}$$

[1] Adaptive strategies adjusting r and s are very interesting to investigate such as [12]; but for simplification we focus on the case where they are constants in our discussion and as we shall show, our emphasis is to investigate the uniform bound for r and s, instead of each individual choice.

[2] As the work $[20,21]$ and many others, a worst-case $O(1/t)$ convergence rate means the accuracy to a solution under certain criteria is of the order $O(1/t)$ after t iterations of an iterative scheme; or equivalently, it requires at most $O(1/\epsilon)$ iterations to achieve an approximate solution with an accuracy of ϵ.

but adopt four different correction strategies to obtain the new iteration, and the range of the combination parameter η was further enlarged to $[-1, 1]$. We only list the equivalent form as in (1.3), more details can be found in [15, Algorithms 1-4]. In particular, the scheme (1.2) was studied in the proximal point algorithm (see [19, 26]) context and some algorithms in prediction-correction framework were proposed where the output of (1.2) (i.e., the predictor) was suggested to be corrected by some correction steps. The convergence of these primal-dual type methods in [13–15] was established from contraction perspective (see [2]). Later, the analysis in [15] was used in [23] to present a pre-conditioning version of the primal-dual algorithm in [5]. Recently, the ergodic convergence rates of more general versions of the first-order primal-dual gradient algorithm were established in [6, 25].

For primal-dual type algorithms summarized in (1.2), how to choose r and s is crucial for ensuring their numerical efficiency. As we have mentioned, the convergence of the original PDHG in [31] or its variants in [3, 9] was proved under the assumption that r and s are sufficiently large. This assumption equally means that the step sizes $1/r$ and $1/s$ must tend to zero asymptotically in iterations. A step size tending to zero inevitably leads to slow convergence in algorithmic performance, and as a common sense, it should be absolutely avoided for any algorithm whenever possible. The strong desire of avoiding too large values of r and s can also be understood in the following way. In fact, for the unified scheme (1.2), we can alternatively understand the parameters r and s as proximal parameters while $\frac{r}{2}\|x-x^k\|^2$ and $\frac{s}{2}\|y-y^k\|^2$ are proximal regularization terms (see [19, 26]). This means the proximal term $\frac{r}{2}\|x-x^k\|^2$ or $\frac{s}{2}\|y-y^k\|^2$ with a larger value of r or s plays a more dominant role in its objective function; hence a very large value of r (resp. s) enforces the new iterate x^{k+1} (resp. y^{k+1}) to be very close to the previous iterate x^k (resp. y^k). That is, slow convergence occurs. Therefore, too large values should be avoided for r and s empirically. Overall, r and s should be large enough such that the convergence is guaranteed but meanwhile they should not be too large in order to avoid too small steps sizes; or equivalently we prefer some smaller values for r and s whenever they can ensure the convergence. These different requirements in theoretical and practical senses thus require us to choose r and s very judiciously to implement primal-dual type algorithms. Note given the symmetric role of r and s in (1.2) and empirical advantages observed in [31], it is usual to discuss rs, rather than r and s individually, when the requirement of these two parameters are considered in the literature. In [5, 15, 30], the requirement for rs is

$$rs > \|AA^T\| \tag{1.4}$$

to ensure the convergence for primal-dual algorithms. In particular, it was analyzed in [15] that the requirement (1.4) can be relaxed to

$$rs > \frac{(1+\eta)^2}{4}\|AA^T\| \tag{1.5}$$

provided that the output of (1.2) was corrected by certain correction step. Based on our previous analysis, we are interested in finding a smaller upper bound for the quantity rs whenever possible. Note the lower bound for rs is reduced by $\frac{(1+\eta)^2}{4}$ times in (1.4); thus there are more choices for r and s and better step sizes for primal-dual algorithms might be sought. This correction step in [15] (see Algorithm 1 therein), however, requires to compute a sophisticated step size at each iteration which needs considerable computation especially for TV variational models in image restoration.

The contributions of this paper can be summarized as follows.

(1). We propose a new correction step to correct the output (1.2) in which the restriction for rs only needs to obey the requirement (1.5); and the step size in the correction step is just a constant which can be easily determined, as we will show later. Thus, a modified primal-dual algorithm with refined step sizes is proposed for (1.1); and compared with the Algorithm 1 in [15], the new algorithm has the same requirement on rs while with a much cheaper correction step.

(2). We prove the global convergence. And then, we establish the worst-case convergence rate in both ergodic and non-ergodic senses for the new primal-dual algorithm under the additional assumptions (e.g., $\mathcal{X} = \Re^n$ and $\mathcal{Y} = \Re^m$). Note the convergence rate in ergodic sense has been analyzed in [5,6] but it is derived from an entirely new technique in this paper; and the convergence rate in non-ergodic sense is also analyzed.

(3). We conduct some numerical experiments on TV image deblurring and inpainting problems to illustrate the advantages of this new algorithm. According to the numerical results, we find that the performance of the new algorithm is sensitive with the step size in the correction step, and the larger step size leads to better performance. Moreover, the numerical results indicate that the sequence generated by the proposed algorithm can converge to the ground-truth solution quickly. In addition, the proposed algorithm outperforms some existing methods, including those proposed in [15], in terms of iteration numbers and CPU computing time for solving problem (1.1).

The rest of this paper is organized as follows. In Section 2, we summarize some preliminaries that are useful for further analysis. In Section 3, a primal-dual algorithm with refined step sizes is proposed; in Section 4, we prove the global convergence and establish a worst-case convergence rate measured by the iteration complexity in ergodic and non-ergodic senses. The results of numerical experiments on TV image deblurring and inpainting problems are shown in Section 5. Finally, we draw some conclusions in Section 6.

2 Preliminaries

In this section, we recall some preliminaries which will be used in our analysis.

Set $u := (x, y)$ and $\Omega := \mathcal{X} \times \mathcal{Y}$. We denote by Ω^* the saddle-point set of $\Phi(x, y)$. Then it is nonempty under our nonempty assumption on the saddle-point set of $\Phi(x, y)$.

In the following theorem, we establish a characterization for the saddle-point set Ω^*. This characterization is useful for establishing the worst-case convergence rate in non-ergodic sense for the new algorithm.

Theorem 2.1. *The saddle-point set Ω^* is convex and it can be characterized as*

$$\Omega^* := \bigcap_{(x,y)\in\Omega} \big\{(\tilde{x}, \tilde{y}) \in \Omega : \Phi(x, \tilde{y}) - \Phi(\tilde{x}, y) \geq 0\big\}. \tag{2.1}$$

Proof. The proof is omitted, as it is an incremental extension of Theorem 2.3.5 in [10], or Theorem 2.1 in [16]. $\square$

The following lemma, presented in [7, Lemma 1.1], is useful for establishing a worst-case convergence rate on the consecutive iterates distance.

Lemma 2.2. *If a sequence $\{a_t\} \subseteq \Re$ obeys: (1) $a_t \geq 0$; (2) $\sum_{t=1}^{\infty} a_t < +\infty$; (3) $\{a_t\}$ is monotonically non-increasing, then we have $a_t = o(1/t)$.*

$\boxed{3}$ A Primal-Dual Algorithm with Refined Step Sizes

In this section, we present the primal-dual algorithm with refined step sizes and give some remarks.

First, as we have mentioned, we are more interested in the requirement (1.5) than (1.4) because it allows more choices for r and s and thus it is more possible to seek better step sizes for primal-dual algorithms. Let us rewrite it here as

Condition (1.5) of Step Sizes:

$$rs > \frac{(1+\eta)^2}{4}\|AA^T\|.$$

Note that the above inequality indicates that r and s can be any positive numbers when $\eta = -1$, because it reduces to $rs > 0$ when $\eta = -1$. One obvious conclusion based on (1.5) is that for any r and s satisfying (1.5), there exists a positive scale $\tau > (1+\eta)^2/4$ such that

$$rs > \tau\|AA^T\|. \tag{3.1}$$

For example, if $rs > (1+\eta)^2\|AA^T\|/4$, we can take value less than $rs/\|AA^T\|$ for τ such that $\tau > (1+\eta)^2/4$ is as great as possible.

Moreover, with the determined τ by (3.1), we can determine a positive parameter σ by

$$\sigma := \frac{2\sqrt{\tau} + (1+\eta)\mathrm{sgn}(\tau-1)}{\sqrt{\tau} + \frac{1}{\sqrt{\tau}} + (1+\eta)\mathrm{sgn}(\tau-1)} > 0, \tag{3.2}$$

where

$$\mathrm{sgn}(\tau-1) := \begin{cases} 1, & \text{if} \quad \tau > 1; \\ 0, & \text{if} \quad \tau = 1; \\ -1, & \text{if} \quad \tau < 1. \end{cases}$$

In fact, it is easy to verify that

$$\sigma : \begin{cases} \in (1,2), & \text{if} \quad \tau > 1; \\ = 1, & \text{if} \quad \tau = 1; \\ \in (0,1), & \text{if} \quad \tau < 1; \end{cases}$$

and thus σ is larger when τ is larger. Note that in Algorithms 2 and 3 in [15], the step sizes always equal to 1. By using the definition (3.2), even under the same condition $rs > \|AA^T\|$, we get that the upper bound of step sizes is larger than 1.

Algorithm 1: A primal-dual algorithm with refined step sizes for (1.1).

Step 0. Choose $\eta \in [-1, 1]$; r and s satisfying the condition (1.5). Let τ be the constant satisfying (3.1) and σ be determined by (3.2). With $u^k := (x^k, y^k) \in \mathcal{X} \times \mathcal{Y}$, the new iterate $u^{k+1} := (x^{k+1}, y^{k+1})$ is generated by the following steps.
Step 1 [Prediction Step]. Implement a primal-dual step to generate a prediction $\tilde{u}^k := (\tilde{x}^k, \tilde{y}^k)$:

$$\left\{ \begin{array}{lcl} \tilde{x}^k & := & \operatorname{Arg\,min}\left\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \,\middle|\, x \in \mathcal{X}\right\}, \\ \bar{x}^k & := & \tilde{x}^k + \eta(\tilde{x}^k - x^k), \\ \tilde{y}^k & := & \operatorname{Arg\,max}\left\{\Phi(\bar{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \,\middle|\, y \in \mathcal{Y}\right\}. \end{array} \right. \tag{3.3}$$

Step 2 [Correction Step]. Correct the predictor and generate a new iterate $u^{k+1} := (x^{k+1}, y^{k+1})$:

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} := \begin{pmatrix} x^k \\ y^k \end{pmatrix} - \alpha \begin{pmatrix} I_n & \frac{1}{r}A^T \\ \frac{\eta}{s}A & I_m \end{pmatrix} \begin{pmatrix} x^k - \tilde{x}^k \\ y^k - \tilde{y}^k \end{pmatrix}, \tag{3.4}$$

where $\alpha \in (0, \sigma]$.

Remark 3.1. As we have mentioned, the correction step (3.4) is computationally cheap because its step size is just a constant. If we define

$$M := \begin{pmatrix} I_n & \frac{1}{r}A^T \\ \frac{\eta}{s}A & I_m \end{pmatrix}, \tag{3.5}$$

then using the notation $u^k := (x^k, y^k)$ and $\tilde{u}^k := (\tilde{x}^k, \tilde{y}^k)$, the correction step (3.4) can be rewritten as

$$u^{k+1} := u^k - \alpha M(u^k - \tilde{u}^k). \tag{3.6}$$

Remark 3.2. The proposed algorithm is a generalization of Algorithms 2 and 3 in [15], and competitive with Algorithm 1 in [15]. There are three reasons as follows. Firstly, the condition (1.5) is more relaxed than $rs > \|AA^T\|$ if $\eta \in [-1, 1)$. Secondly, if we use the same condition $rs > \|AA^T\|$, the upper bound of the step size α is larger than 1 from (3.3). Thirdly, the proposed algorithm used the same condition (1.5) as Algorithm 1 in [15], while the computational cost of step size of the proposed algorithm is smaller than Algorithm 1 in [15].

4 Convergence Analysis

In this section, we prove the global convergence and establish the worst-case convergence rate for the proposed primal-dual algorithm with refined step sizes in ergodic and non-ergodic senses. We first need to show some contraction properties for its sequence.

4.1 Contraction properties

First of all, let us take a look at how accurate the predictor $\tilde{u}^k$ generated by (3.3) is to a saddle point of $\Phi(x, y)$.

Lemma 4.1. *For given $u^k := (x^k, y^k)$, let $\tilde{u}^k := (\tilde{x}^k, \tilde{y}^k) \in \Omega$ be generated by* (3.3). *Then we have*

$$\Phi(x, \tilde{y}^k) - \Phi(\tilde{x}^k, y) \geq (u - \tilde{u}^k)^T Q(u^k - \tilde{u}^k) \quad \forall\, u := (x, y) \in \Omega, \tag{4.1}$$

where

$$Q := \begin{pmatrix} rI_n & A^T \\ \eta A & sI_m \end{pmatrix}. \tag{4.2}$$

Proof. Using (1.1) and the optimality condition of (3.3), we have

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T[-A^T y^k + r(\tilde{x}^k - x^k)] \geq 0 \qquad \forall x \in \mathcal{X},$$

and

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{A[\tilde{x}^k + \eta(\tilde{x}^k - x^k)] + s(\tilde{y}^k - y^k)\} \geq 0 \quad \forall y \in \mathcal{Y}.$$

Combining the above two inequalities, it yields

$$\begin{gathered}\big(\theta_1(x) - (\tilde{y}^k)^T A x - \theta_2(\tilde{y}^k)\big) - \big(\theta_1(\tilde{x}^k) - y^T A\tilde{x}^k - \theta_2(y)\big) \\ + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \end{pmatrix}^T \begin{pmatrix} rI_n & A^T \\ \eta A & sI_m \end{pmatrix} \begin{pmatrix} \tilde{x}^k - x^k \\ \tilde{y}^k - y^k \end{pmatrix} \geq 0,\end{gathered}$$

for any $u := (x, y) \in \Omega$. Using the notation in (4.2), the lemma is proved. □

Recall the matrix M defined in (3.5). Then, for Q defined in (4.2), we have that

$$Q = HM, \tag{4.3}$$

where

$$H := \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix}. \tag{4.4}$$

Our motivation of correcting the predictor in (3.3) by the correction step (3.4) is thus also clear from (4.1). In detail, the term $Q(\tilde{u}^k - u^k)$ appears like a proximal regularization term but the matrix Q in (4.2) is not symmetric. In both VI and proximal point algorithm (PPA) senses, it is not easy to handle the lack of symmetry. We thus have the desire to decompose the matrix Q into the multiplication of a symmetric matrix H and an asymmetric one M. Note the asymmetry of Q is totally remained in M. Thus, if we use $u^k - \alpha M(u^k - \tilde{u}^k)$ as the new iterate, then the diagonal matrix (which is of course symmetric) H plays the role of a regularization coefficient and the inequality (4.1) becomes a convenient tool for analysis, as we shall demonstrate later.

Lemma 4.2. *Let r and s satisfy the condition* (1.5)*; the matrices Q, M and H be given by* (4.2), (3.5) *and* (4.4)*, respectively; the constant τ satisfy the condition* (3.1) *and the constant σ be given in* (3.2). *Then we have*

$$G := Q^T + Q - \sigma M^T H M \succ 0. \tag{4.5}$$

Proof. Since $Q = HM$ and $H^T = H$, we have

$$M^T H M = Q^T M \;=\; \begin{pmatrix} rI_n & \eta A^T \\ A & sI_m \end{pmatrix} \begin{pmatrix} I_n & \frac{1}{r}A^T \\ \frac{\eta}{s}A & I_m \end{pmatrix}$$

$$= \begin{pmatrix} rI_n + \frac{\eta^2}{s}A^TA & (1+\eta)A^T \\ (1+\eta)A & sI_m + \frac{1}{r}AA^T \end{pmatrix}. \tag{4.6}$$

Using (4.2), we get

$$Q^T + Q = \begin{pmatrix} 2rI_n & (1+\eta)A^T \\ (1+\eta)A & 2sI_m \end{pmatrix}. \tag{4.7}$$

It follows from (4.6) and (4.7) that

$$\begin{aligned} G &= Q^T + Q - \sigma M^T HM \\ &= \begin{pmatrix} (2-\sigma)rI_n - \frac{\sigma\eta^2}{s}A^TA & (1+\eta)(1-\sigma)A^T \\ (1+\eta)(1-\sigma)A & (2-\sigma)sI_m - \frac{\sigma}{r}AA^T \end{pmatrix} \\ &= \begin{pmatrix} \frac{\sigma}{\tau}rI_n - \frac{\sigma\eta^2}{s}A^TA & 0 \\ 0 & \frac{\sigma}{\tau}sI_m - \frac{\sigma}{r}AA^T \end{pmatrix} \\ &\quad + \begin{pmatrix} (2-\sigma-\frac{\sigma}{\tau})rI_n & (1+\eta)(1-\sigma)A^T \\ (1+\eta)(1-\sigma)A & (2-\sigma-\frac{\sigma}{\tau})sI_m \end{pmatrix}. \end{aligned} \tag{4.8}$$

Recall that we have $\tau > (1+\eta)^2/4$. In the following, we will prove $G \succ 0$ under the condition (3.1). To prove this assertion, there are three cases for τ to discuss, that is $\tau > 1$, $(1+\eta)^2/4 < \tau < 1$ and $\tau = 1$.

1) If $\tau > 1$, from (3.2) and $-1 \le \eta \le 1$ we have

$$\sigma = \frac{2\sqrt{\tau} + 1 + \eta}{\sqrt{\tau} + \frac{1}{\sqrt{\tau}} + 1 + \eta} > 0. \tag{4.9}$$

It follows from (4.9) that

$$(1+\eta)(1-\sigma) = -\sqrt{\tau}(2 - \sigma - \frac{\sigma}{\tau}). \tag{4.10}$$

Since $-1 \le \eta \le 1$ and $\tau > 1$, from (4.9) and (4.10), we have $\sigma > 1$ and thus

$$2 - \sigma - \frac{\sigma}{\tau} = \frac{(1+\eta)(\sigma-1)}{\sqrt{\tau}} \ge 0. \tag{4.11}$$

Then from (4.8) and (4.10), we get

$$\begin{aligned} G &= Q^T + Q - \sigma M^T HM \\ &= \sigma \begin{pmatrix} \frac{r}{\tau}I_n - \frac{\eta^2}{s}A^TA & 0 \\ 0 & \frac{s}{\tau}I_m - \frac{1}{r}AA^T \end{pmatrix} + (2 - \sigma - \frac{\sigma}{\tau}) \begin{pmatrix} rI_n & -\sqrt{\tau}A^T \\ -\sqrt{\tau}A & sI_m \end{pmatrix}. \end{aligned}$$

By using the fact that $rs > \tau\|AA^T\|$, $-1 \le \eta \le 1$, $\sigma > 0$ and (4.11), we have

$$\sigma \begin{pmatrix} \frac{r}{\tau}I_n - \frac{\eta^2}{s}A^TA & 0 \\ 0 & \frac{s}{\tau}I_m - \frac{1}{r}AA^T \end{pmatrix} \succ 0 \quad \text{and} \quad (2-\sigma-\frac{\sigma}{\tau}) \begin{pmatrix} rI_n & -\sqrt{\tau}A^T \\ -\sqrt{\tau}A & sI_m \end{pmatrix} \succeq 0.$$

And thus, the matrix G is positive definite.

2) Similarly, we could prove that the matrix $G = Q^T + Q - \sigma M^T H M$ is positive definite if $(1+\eta)^2/4 < \tau < 1$ and

$$\sigma = \frac{2\sqrt{\tau} - 1 - \eta}{\sqrt{\tau} + \frac{1}{\sqrt{\tau}} - 1 - \eta}.$$

3) If $\tau = 1$, we have

$$\sigma = \frac{2\sqrt{\tau}}{\sqrt{\tau} + \frac{1}{\sqrt{\tau}}} = 1.$$

And thus we get $2 - \sigma - \frac{\sigma}{\tau} = 0$ and $1 - \sigma = 0$. Note that $-1 \le \eta \le 1$, $\tau = 1$ and $rs > \|AA^T\|$. From (4.8), we get

$$G = Q^T + Q - M^T H M = \begin{pmatrix} rI_n - \frac{\eta^2}{s} A^T A & 0 \\ 0 & sI_m - \frac{1}{r} AA^T \end{pmatrix} \succ 0.$$

The proof is complete. □

Now, we can prove an important inequality in the following theorem which will give us an estimate for the accuracy of $\tilde{u}^k$ in terms of some simple quadratic terms.

Theorem 4.3. *For given $u^k = (x^k, y^k)$, let $\tilde{u}^k := (\tilde{x}^k, \tilde{y}^k) \in \Omega$ be generated by* (3.3) *and u^{k+1} be updated by* (3.6)*. Then for any $u \in \Omega$, we have*

$$\begin{aligned} &\Phi(x, \tilde{y}^k) - \Phi(\tilde{x}^k, y) \\ &\ge \frac{1}{2\alpha}\big(\|u - u^{k+1}\|_H^2 - \|u - u^k\|_H^2\big) + \frac{1}{2}\|u^k - \tilde{u}^k\|_G^2, \end{aligned} \tag{4.12}$$

where the matrices H and G are defined by (4.4) *and* (4.5)*, respectively.*

Proof. Using (4.1) and $\alpha > 0$, we have

$$\alpha\big(\Phi(x, \tilde{y}^k) - \Phi(\tilde{x}^k, y)\big) \ge \alpha(u - \tilde{u}^k)^T Q(u^k - \tilde{u}^k) \ \ \forall u \in \Omega. \tag{4.13}$$

It follows from $Q = HM$ (see (4.3)) and the update form (3.6) that

$$\begin{aligned} \alpha(u - \tilde{u}^k)^T Q(u^k - \tilde{u}^k) &= \alpha(u - \tilde{u}^k)^T HM(u^k - \tilde{u}^k) \\ &= (u - \tilde{u}^k)^T H(u^k - u^{k+1}). \end{aligned} \tag{4.14}$$

Applying the identity

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

to the right-hand side of (4.14) with

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

we thus obtain

$$\begin{aligned} &(u - \tilde{u}^k)^T H(u^k - u^{k+1}) \\ &= \frac{1}{2}\big(\|u - u^{k+1}\|_H^2 - \|u - u^k\|_H^2\big) + \frac{1}{2}(\|u^k - \tilde{u}^k\|_H^2 - \|u^{k+1} - \tilde{u}^k\|_H^2). \end{aligned} \tag{4.15}$$

For the last term of the equation (4.15), we have

$$\begin{aligned}
&\|u^k-\tilde{u}^k\|_H^2-\|u^{k+1}-\tilde{u}^k\|_H^2\\
&\quad= \|u^k-\tilde{u}^k\|_H^2-\|(u^k-\tilde{u}^k)-(u^k-u^{k+1})\|_H^2\\
&\quad\overset{(3.6)}{=} \|u^k-\tilde{u}^k\|_H^2-\|(u^k-\tilde{u}^k)-\alpha M(u^k-\tilde{u}^k)\|_H^2\\
&\quad= 2\alpha(u^k-\tilde{u}^k)^T HM(u^k-\tilde{u}^k)-\alpha^2\|u^k-\tilde{u}^k\|_{M^THM}^2\\
&\quad\overset{(4.3)}{=} \alpha(u^k-\tilde{u}^k)^T(Q^T+Q)(u^k-\tilde{u}^k)-\alpha^2\|u^k-\tilde{u}^k\|_{M^THM}^2.
\end{aligned}$$

Combining (4.13), (4.14), (4.15) and the above equations, we get

$$\begin{aligned}
&\Phi(x,\tilde{y}^k)-\Phi(\tilde{x}^k,y)\\
&\quad\geq \frac{1}{2\alpha}\big(\|u-u^{k+1}\|_H^2-\|u-u^k\|_H^2\big)+\frac{1}{2}\big(\|u^k-\tilde{u}^k\|_{(Q^T+Q)}^2\\
&\qquad -\alpha\|u^k-\tilde{u}^k\|_{M^THM}^2\big).
\end{aligned} \tag{4.16}$$

Using (4.5) and $0<\alpha\leq\sigma$, we get

$$\begin{aligned}
&\|u^k-\tilde{u}^k\|_{(Q^T+Q)}^2-\alpha\|u^k-\tilde{u}^k\|_{M^THM}^2\\
&\quad\geq \|u^k-\tilde{u}^k\|_{(Q^T+Q)}^2-\sigma\|u^k-\tilde{u}^k\|_{M^THM}^2=\|u^k-\tilde{u}^k\|_G^2.
\end{aligned}$$

Substituting the above inequality into (4.16), we get the assertion (4.12). □

Now we are ready to show that the sequence $\{u^k\}$ generated by the proposed algorithm is contractive with respect to the saddle-point set Ω^*.

Theorem 4.4. *For given $u^k=(x^k,y^k)$, let $\tilde{u}^k:=(\tilde{x}^k,\tilde{y}^k)\in\Omega$ be generated by* (3.3) *and u^{k+1} be updated by* (3.6). *Then we have*

$$\|u^{k+1}-u^*\|_H^2\leq\|u^k-u^*\|_H^2-\alpha\|u^k-\tilde{u}^k\|_G^2\quad\forall u^*\in\Omega^*, \tag{4.17}$$

where the matrices H and G are defined by (4.4) *and* (4.5)*, respectively.*

Proof. Setting $u=u^*$ in (4.12), we get

$$\begin{aligned}
&\|u^k-u^*\|_H^2-\|u^{k+1}-u^*\|_H^2\\
&\quad\geq \alpha\|u^k-\tilde{u}^k\|_G^2+2\alpha\big(\Phi(\tilde{x}^k,y^*)-\Phi(x^*,\tilde{y}^k)\big).
\end{aligned} \tag{4.18}$$

Since $u^*:=(x^*,y^*)\in\Omega^*$, from (2.1) we have

$$\Phi(\tilde{x}^k,y^*)-\Phi(x^*,\tilde{y}^k)\geq 0.$$

The assertion (4.17) follows from (4.18) and the above inequality directly. □

4.2 Global convergence

Based on the contraction property established in the last subsection, we can easily prove the global convergence of the proposed algorithm. The proof follows the standard analytic framework of contraction methods, see e.g. [2]; but we include the detail for completeness.

Theorem 4.5. *The sequence $\{u^k\}$ generated by the proposed primal-dual algorithm converges to some u^∞ in the set Ω^*.*

Proof. It follows from (4.17) that the sequence $\{u^k\}$ is bounded and

$$\lim_{k\to\infty} \|u^k - \tilde{u}^k\|_G = 0. \tag{4.19}$$

Combining this with $G \succ 0$, we deduce that $\{\tilde{u}^k\}$ is also bounded. Let u^∞ be a cluster point of $\{\tilde{u}^k\}$ and $\{\tilde{u}^{k_j}\}$ be a subsequence which converges to u^∞. Let $\{u^{k_j}\}$ and $\{\tilde{u}^{k_j}\}$ be the induced sequences by $\{u^k\}$ and $\{\tilde{u}^k\}$, respectively. It follows from (4.1) that $\tilde{u}^{k_j} \in \Omega$,

$$\Phi(x, \tilde{y}^{k_j}) - \Phi(\tilde{x}^{k_j}, y) \geq (u - \tilde{u}^{k_j})^T Q(u^{k_j} - \tilde{u}^{k_j}) \ \ \forall\, u \in \Omega.$$

Letting $j \to \infty$, it follows from (4.19) and the continuity of $\Phi(x, y)$ that $u^\infty := (x^\infty, y^\infty) \in \Omega$,

$$\Phi(x, y^\infty) - \Phi(x^\infty, y) \geq 0 \qquad \forall\, u := (x, y) \in \Omega.$$

The above inequality indicates that u^∞ is a saddle point of (1.1). By using (4.19) and $\lim_{j\to\infty} \tilde{u}^{k_j} = u^\infty$, the subsequence $\{u^{k_j}\}$ also converges to u^∞. Due to (4.17), we have

$$\|u^{k+1} - u^\infty\|_H \leq \|u^k - u^\infty\|_H,$$

and thus $\{u^k\}$ converges to u^∞. The proof is complete. □

4.3 Convergence rate in ergodic sense

Now we show the worst-case $O(1/t)$ convergence rate in ergodic sense for the proposed algorithm. As we have mentioned, the first such result was established in [5] by using the technique in [20, 21]. But here we are considering a different primal-dual algorithm (the choices for r and s are more relaxed) combined with a correction step; and the technique for analysis is mainly inspired by that in [15] and thus different from that in [5]. Note this convergence rate result is lacked in [15]. The base for the analysis in this subsection is Theorem 4.3.

Theorem 4.6. *For given $u^k := (x^k, y^k)$, let $\tilde{u}^k$ and u^{k+1} be generated by the proposed algorithm. Then for any integer $t > 0$, we have*

$$\Phi(x, \tilde{y}_t) - \Phi(\tilde{x}_t, y) \geq -\frac{1}{2\alpha(t+1)}\|u - u^0\|_H^2 \quad \forall u \in \Omega, \tag{4.20}$$

where

$$(\tilde{x}_t, \tilde{y}_t) := \tilde{u}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{u}^k. \tag{4.21}$$

Proof. First, from (4.12), for the sequence generated by the proposed algorithm, we have $\tilde{u}^k \in \Omega$,

$$\Phi(x, \tilde{y}^k) - \Phi(\tilde{x}^k, y) + \frac{1}{2\alpha}\|u - u^k\|_H^2 \geq \frac{1}{2\alpha}\|u - u^{k+1}\|_H^2 \ \forall u \in \Omega. \tag{4.22}$$

Note that $\tilde{u}^k \in \Omega$ for all $k \geq 0$. Together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (4.21) implies $\tilde{u}_t \in \Omega$. Summing the inequality (4.22) over $k = 0, 1, \ldots, t$, we obtain

$$\begin{aligned}&(t+1)\big(\theta_1(x) + \theta_2(y)\big) - \sum_{k=0}^{t}\big(\theta_1(\tilde{x}^k) + \theta_2(\tilde{y}^k)\big) - \Big(\sum_{k=0}^{t}\tilde{y}^k\Big)^T Ax + y^T A\Big(\sum_{k=0}^{t}\tilde{x}^k\Big)\\ &\geq -\frac{1}{2\alpha}\|u - u^0\|_H^2 \quad \forall u \in \Omega.\end{aligned}$$

Using the notation of $\tilde{u}_t$, it can be written as

$$\begin{aligned}&\big(\theta_1(x)+\theta_2(y)\big)-\frac{1}{t+1}\sum_{k=0}^{t}\big(\theta_1(\tilde{x}^k)+\theta_2(\tilde{y}^k)\big)-(\tilde{y}_t)^TAx+y^TA\tilde{x}_t\\ &\geq -\frac{1}{2\alpha(t+1)}\|u-u^0\|_H^2 \quad \forall u\in\Omega. \end{aligned}\tag{4.23}$$

Since $\theta_1(x)$ and $\theta_2(y)$ are convex and

$$\tilde{u}_t=\frac{1}{t+1}\sum_{k=0}^{t}\tilde{u}^k,$$

we have that

$$\theta_1(\tilde{x}_t)+\theta_2(\tilde{y}_t)\leq\frac{1}{t+1}\sum_{k=0}^{t}(\theta_1(\tilde{x}^k)+\theta_2(\tilde{y}^k)).$$

Substituting it in (4.23), the assertion of this theorem follows directly. □

It follows from (4.17) that the sequence $\{u^k\}$ generated by the propose algorithm is bounded. According to (4.19) and $G\succ 0$, the sequence $\{\tilde{u}^k\}$ is also bounded. Since the saddle-point set Ω^* is assumed to be nonempty, we suppose that there exists a saddle point $u^*:=(x^*,y^*)\in\Omega^*$. Then it follows from (4.20) that

$$\Phi(\tilde{x}_t,y^*)-\Phi(x^*,\tilde{y}_t)\leq\frac{1}{2\alpha(t+1)}\|u^*-u^0\|_H^2.\tag{4.24}$$

From the definition of $\Phi(\cdot,\cdot)$ in (1.1), the left-hand side of (4.24) can be rewritten as

$$\begin{aligned}&\Phi(\tilde{x}_t,y^*)-\Phi(x^*,\tilde{y}_t)\\ &=\theta_1(\tilde{x}_t)-(y^*)^TA\tilde{x}_t-\theta_2(y^*)-\theta_1(x^*)+(\tilde{y}_t)^TAx^*+\theta_2(\tilde{y}_t)\\ &=\Big(\theta_1(\tilde{x}_t)-\theta_1(x^*)-(A^Ty^*)^T(\tilde{x}_t-x^*)\Big)+\Big(\theta_2(\tilde{y}_t)-\theta_2(y^*)-(-Ax^*)^T(\tilde{y}_t-y^*)\Big).\end{aligned}\tag{4.25}$$

Assume that $\mathcal{X}:=\Re^n$ and $\mathcal{Y}:=\Re^m$. According to the first-order optimality condition of the saddle-point problem (1.1), we know that $A^Ty^*\in\partial\theta_1(x^*)$ and $-Ax^*\in\partial\theta_2(y^*)$. Substituting (4.25) into (4.24), we get

$$D_{\theta_1}(\tilde{x}_t,x^*)+D_{\theta_2}(\tilde{y}_t,y^*)\leq\frac{1}{2\alpha(t+1)}\|u^*-u^0\|_H^2,$$

where the generalized Bregman distance $D_{\theta_i}(\cdot,\cdot)$ $(i=1,2)$ [25,29] is defined as

$$D_{\theta_i}(z,\tilde{z}):=\theta_i(z)-\theta_i(\tilde{z})-p^T(z-\tilde{z})\quad\forall z,\tilde{z}\in\mathrm{dom}\theta_i,\tag{4.26}$$

with $p\in\partial\theta_i(\tilde{z})$. This implies that $\tilde{u}_t:=(\tilde{x}_t,\tilde{y}_t)$ is an approximate saddle point of $\Phi(x,y)$ with an accuracy of $O(1/t)$ in the generalized Bregman distance. That is, a worst-case $O(1/t)$ convergence rate of the proposed algorithm in ergodic sense is established.

4.4 Convergence rate in non-ergodic sense

Now, we prove the worst-case $o(1/\sqrt{t})$ convergence rate in non-ergodic sense for the proposed algorithm. We first find a criterion to measure the accuracy of an iterate, as the criterion (2.1) can be used for establishing the non-ergodic convergence rate. In fact, recall (4.1) and $Q = HM$. If $\|M(u^k - \tilde{u}^k)\|_H = 0$, we get $\tilde{u}^k \in \Omega$,

$$\Phi(x, \tilde{y}^k) - \Phi(\tilde{x}^k, y) \geq (u - \tilde{u}^k)^T HM(u^k - \tilde{u}^k) = 0 \qquad \forall\, u \in \Omega.$$

We thus can claim that $\tilde{u}^k$ is a saddle point of (1.1) if $\|M(u^k - \tilde{u}^k)\|_H = 0$; which means we can measure the accuracy of $\tilde{u}^k$ to a saddle point of (1.1) by the quantity $\|M(u^k - \tilde{u}^k)\|_H$.

Let us take a deeper look at the sequence generated by the proposed algorithm and establish more properties.

Lemma 4.7. *For given $u^k := (x^k, y^k)$, let $\tilde{u}^k := (\tilde{x}^k, \tilde{y}^k) \in \Omega$ be generated by* (3.3) *and u^{k+1} be updated by* (3.6)*. Then for any integer $k > 0$, we have*

$$\begin{aligned} &\alpha(u^k - \tilde{u}^k)^T M^T HM[(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})] \\ &\geq \frac{1}{2}\|(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\|^2_{(Q^T+Q)}, \end{aligned} \tag{4.27}$$

where the matrices Q, M and H are defined by (4.2), (3.5) *and* (4.4)*, respectively.*

Proof. First, setting $u = \tilde{u}^{k+1}$ in (4.1), we have

$$\Phi(\tilde{x}^{k+1}, \tilde{y}^k) - \Phi(\tilde{x}^k, \tilde{y}^{k+1}) \geq (\tilde{u}^{k+1} - \tilde{u}^k)^T Q(u^k - \tilde{u}^k). \tag{4.28}$$

Note that (4.1) is also true for $k := k + 1$ and thus we have

$$\Phi(x, \tilde{y}^{k+1}) - \Phi(\tilde{x}^{k+1}, y) \geq (u - \tilde{u}^{k+1})^T Q(u^{k+1} - \tilde{u}^{k+1}) \quad \forall\, u \in \Omega.$$

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

$$\Phi(\tilde{x}^k, \tilde{y}^{k+1}) - \Phi(\tilde{x}^{k+1}, \tilde{y}^k) \geq (\tilde{u}^k - \tilde{u}^{k+1})Q(u^{k+1} - \tilde{u}^{k+1}). \tag{4.29}$$

Adding (4.28) and (4.29) and using the monotonicity of F, we get

$$(\tilde{u}^k - \tilde{u}^{k+1})^T Q[(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})] \geq 0. \tag{4.30}$$

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 the both sides of (4.30), and using $u^T Qu = \frac{1}{2}u^T(Q^T + Q)u$, we get

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

Substituting the term $u^k - u^{k+1} = \alpha M(u^k - \tilde{u}^k)$ into the left-hand side of the last inequality, and using $Q = HM$, we obtain (4.27). The proof is complete. □

Theorem 4.8. *For given $u^k := (x^k, y^k)$, let $\tilde{u}^k := (\tilde{x}^k, \tilde{y}^k) \in \Omega$ be generated by* (3.3) *and u^{k+1} be updated by* (3.6)*. Then for any integer $k > 0$, we have*

$$\|M(u^k - \tilde{u}^k)\|^2_H - \|M(u^{k+1} - \tilde{u}^{k+1})\|^2_H \geq 0, \tag{4.31}$$

where the matrices M and H are defined by (3.5) *and* (4.4)*, respectively.*

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

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

we obtain

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

Inserting (4.27) into the first term of the right-hand side of the last equality, we obtain

$$\begin{array}{rl} & \|M(u^k - \tilde{u}^k)\|_H^2 - \|M(u^{k+1} - \tilde{u}^{k+1})\|_H^2 \\ \geq & \dfrac{1}{\alpha}\|(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\|_{(Q^T+Q)}^2 - \|(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\|_{M^T H M}^2 \\ = & \dfrac{1}{\alpha}\big(\|(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\|_{(Q^T+Q)}^2 \\ & -\alpha\|(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\|_{M^T H M}^2\big). \end{array} \tag{4.32}$$

Since $0 < \alpha \leq \sigma$, we have

$$Q^T + Q - \alpha M^T H M \succeq Q^T + Q - \sigma M^T H M \succ 0.$$

It follows from the above formula that the right hand side of (4.32) is nonnegative. □

Note that since $G \succ 0$ and $M^T H M \succeq 0$, it follows from (4.17) that there is a constant $c_1 > 0$ such that

$$\|u^{k+1} - u^*\|_H^2 \leq \|u^k - u^*\|_H^2 - c_1\|u^k - \tilde{u}^k\|_{M^T H M}^2 \quad \forall u^* \in \Omega^*. \tag{4.33}$$

To establish the worst-case convergence rate in non-ergodic sense for the proposed algorithm, we give an important lemma in the following.

Lemma 4.9. *Let $\{u^k\}$ and $\{\tilde{u}^k\}$ be the sequences generated by the proposed algorithm. Then for any integer $t > 0$, we have*

$$\|M(u^t - \tilde{u}^t)\|_H = o(1/\sqrt{t}). \tag{4.34}$$

Proof. First, it follows from (4.33) that

$$\sum_{k=0}^{\infty} c_1\|M(u^k - \tilde{u}^k)\|_H^2 \leq \|u^0 - u^*\|_H^2 \quad \forall\, u^* \in \Omega^*. \tag{4.35}$$

According to (4.31), the sequence $\{\|M(u^k - \tilde{u}^k)\|_H\}$ is nonincreasing. The assertion (4.34) follows from (4.35) and Lemma 2.2 immediately. □

Notice that Ω^* is convex and closed (see Theorem 2.1). From Theorem 2.1 and Lemma 4.1, we get that $\tilde{u}^t$ is a saddle point of (1.1) if $\|M(u^t - \tilde{u}^t)\|_H = 0$. A worst-case $o(1/\sqrt{t})$ convergence rate in non-ergodic sense is thus established for the proposed algorithm.

Theorem 4.10. *Let $\{u^t\}$ and $\{\tilde{u}^t\}$ be the sequences generated by the proposed algorithm. Then for any integer $t > 0$ and $u^* := (x^*, y^*) \in \Omega^*$, we have*

$$\Phi(x^*, \tilde{y}^t) - \Phi(\tilde{x}^t, y^*) \geq -o(1/\sqrt{t}).$$

Proof. It follows from Lemma 4.1 that

$$\Phi(x,\tilde{y}^t)-\Phi(\tilde{x}^t,y)\geq(\tilde{u}^t-u)^TQ(\tilde{u}^t-u^t). \tag{4.36}$$

Note that $\{\tilde{u}^t\}$ is bounded. By using (4.36), (4.3) and (4.34), for any $u^*\in\Omega^*$, we have

$$\begin{aligned}\Phi(x^*,\tilde{y}^t)-\Phi(\tilde{x}^t,y^*) &\geq (u^*-\tilde{u}^t)^THM(u^t-\tilde{u}^t)\\ &\geq -\|u^*-\tilde{u}^t\|_H\|M(u^t-\tilde{u}^t)\|_H\\ &= -o(1/\sqrt{t}).\end{aligned}$$

The proof is complete. □

Remark 4.11. Assume that $\mathcal{X}:=\Re^n$ and $\mathcal{Y}:=\Re^m$. From (4.25) and Theorem 4.10, we get

$$D_{\theta_1}(\tilde{x}^t,x^*)+D_{\theta_2}(\tilde{y}^t,y^*)\leq o(1/\sqrt{t}),$$

where the generalized Bregman distance $D_{\theta_i}(\cdot,\cdot)$ $(i=1,2)$ is defined in (4.26).

5 Numerical Experiments

To investigate the numerical performance of the proposed method, in this section, we apply Algorithm 1 (denoted by "R-PDA") to solve the TV image deblurring and inpainting problems, and compare with some existing works. All codes were written by Matlab 2014b and all the numerical experiments were conducted on a personal computer with 4GB memory.

5.1 Image deblurring problem

In this subsection, we apply the R-PDA to solve the TV image deblurring problem and compare it with the algorithms proposed in [15]. As discussed in [15,31], the TV deblurring model can be reformulated as to find $(x^*,y^*)\in\mathcal{X}\times\Re^n$ such that

$$\forall y\in\Re^n \qquad \Phi_1(x^*,y)\leq\Phi_1(x^*,y^*)\leq\Phi_1(x,y^*) \qquad \forall x\in\mathcal{X}, \tag{5.1}$$

where $\Phi_1(x,y):=-y^T\nabla x-\frac{\lambda}{2}\|By-b\|^2$, $\mathcal{X}=\{x\in\Re^{n\times 2}\mid \|x\|_\infty\leq 1\}$, b is the given observed image, ∇ is the discrete gradient operator [31], $\lambda>0$ is a balanced constant and B is a matrix that represents a space-invariant blurring operator. It is obvious that problem (5.1) is a special case of (1.1) and we can apply R-PDA to solve it. When applying R-PDA scheme to solve problem (5.1), the x-subproblem in (3.3) reduces to

$$\tilde{x}^k=P_{\mathcal{X}}(x^k+\frac{1}{r}\nabla^Ty^k),$$

where $P_{\mathcal{X}}$ denotes the projection onto $\mathcal{X}$ which can be easily computed. The y-subproblem in (3.3) corresponds to solve the following system of equations:

$$\lambda B^T(By-b)+\nabla\bar{x}^k+s(y-y^k)=0,$$

whose solution can be obtained by the Fast Fourier Transform (FFT) or Discrete Cosine Transform (DCT) [31].

We test the images Cameraman (256×256), Peppers (256×256), Hatgirl (256×256) and Barbara (512×512) presented in the first row in Figure 1. These original images are then degraded by severe motion blur, and the motion blur operator is generated by the script

Figure 1: Original images (the first row) and degraded images (the second row). From left to right: Cameraman, Peppers, Hatgirl, Barbara.

`fspecial` in the MATLAB Image Processing Toolbox with `theta` $= 135$ and `len` $= 91$. The degraded images are presented in the second row in Figure 1.

In order to recover these corrupted images and due to the sole purpose of investigating the efficiency of each tested algorithm, as suggested in [15], we take $\lambda = 250$ in (5.1) and all the algorithms tested in this subsection use the stopping criterion as follows:

$$\max\left\{\frac{\|x^{k+1}-x^k\|}{\|x^{k+1}\|}, \frac{\|y^{k+1}-y^k\|}{\|y^{k+1}\|}\right\} < \mathrm{Tol}, \tag{5.2}$$

where $\{x^k\}$ and $\{y^k\}$ are generated by the tested algorithm, and we take the error Tol $:= 10^{-4}$. Besides, we initial each algorithm with the degraded images. Besides, the quality of recovered images is measured by the value of the signal-to-noise ratio (SNR) given by

$$\mathrm{SNR} := 20\log_{10}\frac{\|y^*\|}{\|\tilde{y}-y^*\|}, \tag{5.3}$$

where y^* is the original image and $\tilde{y}$ is the image restored by a certain tested algorithm.

Table 1: Iteration numbers of R-PDA with different α and η for Hatgirl

$\alpha\backslash\eta$	-1	-0.9	-0.7	-0.5	-0.3	-0.1	0	0.1	0.3	0.5	0.7	0.9	1
0.3σ	152	152	151	151	151	150	150	150	150	149	149	150	150
0.6σ	98	98	98	98	97	97	97	96	96	96	95	95	96
0.9σ	79	78	78	78	78	77	77	77	77	76	76	76	77
σ	74	74	73	73	73	73	72	72	72	71	71	72	73

Now we elaborate on the parameters setting involved in R-PDA. From the conditions (1.5) and (3.1), we know that τ should satisfy

$$\frac{(1+\eta)^2}{4} < \tau < \frac{rs}{\|\nabla^T\nabla\|},$$

and it follows from [4] that $\|\nabla^T \nabla\| \leq 8$, thus we take $\tau := rs/8 - 0.01$ in the experiment. First, we test the sensitivity of the parameters r, s and η. We fix $\eta = 0.5$ and test the performance of R-PDA with different r, s. We find that the situation $r = 100$, $s = 1/12$ performs better than others. Here we omit to list the results due to the limited space. Next, we fix $r = 100$, $s = 1/12$ and observe the performance of R-PDA with different values of α and $\eta \in [-1, 1]$. We test 4 and 13 instances for α and η, the concrete choices can be found in Table 1, where σ is defined in (3.2). We report the preliminary results in Table 1 just including the iteration numbers for deblurring the image (Hatgirl) with the same SNR values. From the results we can find that the numerical performance is sensitive with α and the choice with $\alpha := \sigma$ and $\eta = 0.7$ usually performs better than others. Therefore, we take $\alpha := \sigma$ and $(r, s, \eta) = (100, 1/12, 0.7)$ for R-PDA in the experiments in this subsection.

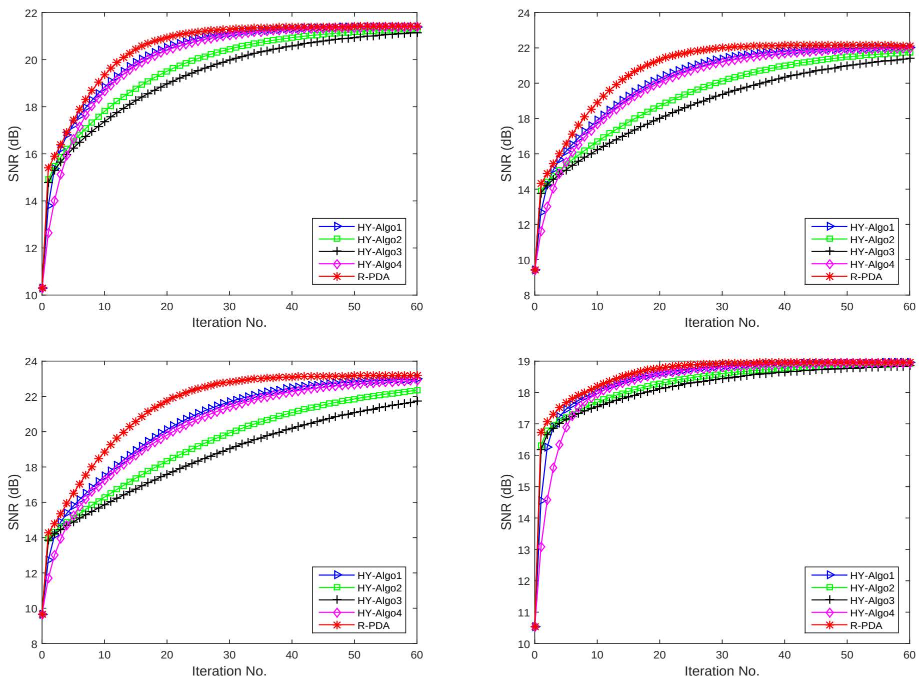

Figure 2: Evolutions of SNRs with respect to iterations for Cameraman (left in the first row), Peppers (right in the first row), Hatgirl (left in the second row) and Barbara(right in the second row).

To illustrate the efficiency of the proposed method, we further compare the numerical performance of R-PDA with Algorithms 1-4 proposed in [15], i.e., the iteration (1.3) with four different correction ways, we denote them by HY-Algo1, HY-Algo2, HY-Algo3 and HY-Algo4, respectively. In our experiments, we take the optimal parameters as that in [15, Table 2] for HY-Algo1, HY-Algo2, HY-Algo3 and HY-Algo4. For the results, we plot the evolutions of SNR values of different methods with respect to iterations in Figure 2 for both tested images, which show that the R-PDA converges faster and can achieve better quality with higher SNR values than that proposed in [15]. In addition, in order to indicate that the sequence generated by each tested algorithm can converge to the ground-truth solution, we also show the evolutions of the value of $\|y^k - \bar{y}\|/\|\bar{y}\|$ with respect to iteration number for

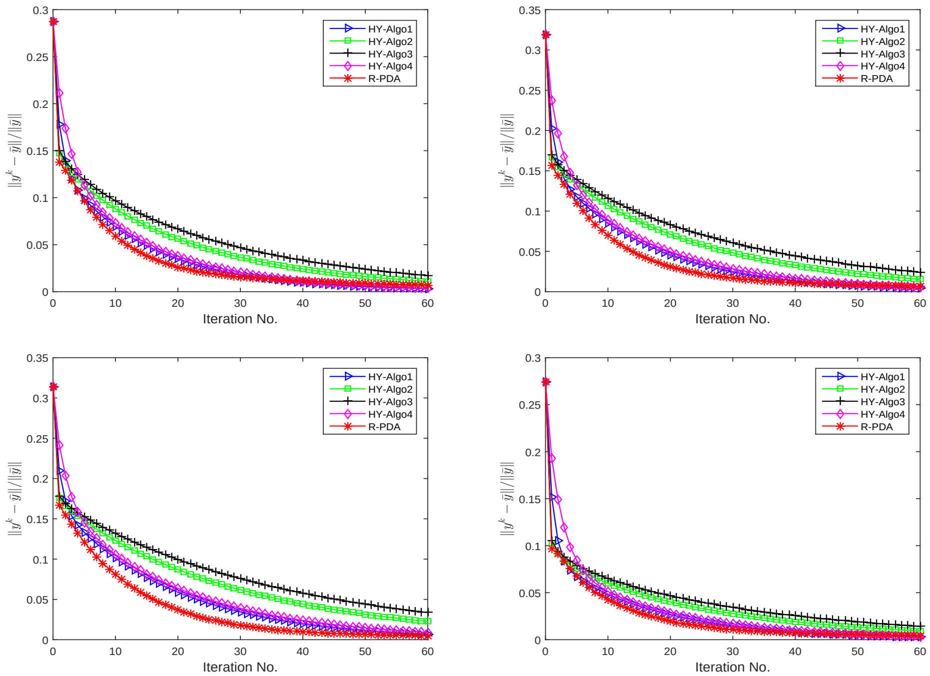

Figure 3: Evolutions of the value of $\|y^k - \bar{y}\|/\|\bar{y}\|$ with respect to iterations for Cameraman (left in the first row), Peppers (right in the first row), Hatgirl (left in the second row) and Barbara(right in the second row).

each algorithm in Figure 3, where $\bar{y}$ denotes the benchmark solution which is produced by the the standard primal-dual method in [5] after 1000 iterations.

In order to further visualize the numerical comparison, we list the images restored by HY-Algo1, HY-Algo2, HY-Algo3, HY-Algo4 and R-PDA in Figure 4 for Peppers and Barbara. Here we omit to show the recovered images Cameraman and Hatgirl due to the limited space.

5.2 Image inpainting problem

In this subsection, we apply the R-PDA to solve the TV image inpainting problem, and compare it with the first-order primal-dual algorithm (denoted by CP) in [5, Algorithm 1], the PDHG ((1.2) with $\eta = 0$) in [31] and HY-Algo1 in [15, Algorithm 1].

The image inpainting problem with TV regularization shares the form of (5.1) with $B \in \Re^{N\times N}$ represented a mask operator, which characterizes the missing information of the original image. We test the images House (256×256) and Peppers (512×512). We present the original and degraded images in Figure 5. For the degraded images, the operator B is a character mask for House and a texture Peppers, where about 15% and 60% of pixels are missed, respectively. Besides, we add the zero-mean Gaussian noise with standard deviation 0.02 for both images. For both tested methods, we take $\lambda = 50$ in the model (5.1). For the parameters of the compared methods, as suggested in [15], we take $(r, s) = (50, r/8)$ for CP, $(r, s) = (1/8, 100)$ for PDHG, and $(r, s, \eta, \gamma) = (50/3, 1/3, -0.4, 1.3)$ for HY-Algo1.

Figure 4: From left to right: Images Peppers (the first row) and Barbara (the second row) restored by HY-Algo1, HY-Algo2, HY-Algo3, HY-Algo4 and R-PDA.

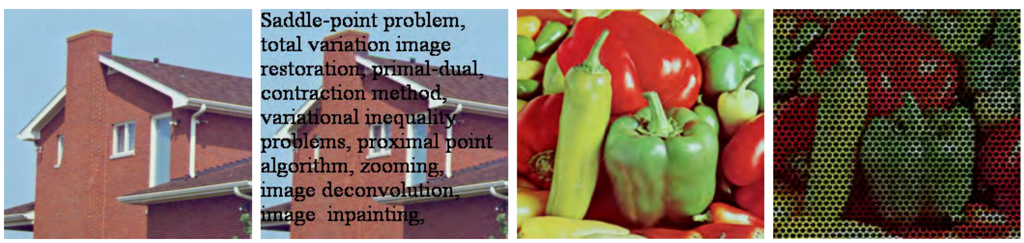

Figure 5: From left to right: Original House, degraded House, Original Peppers and degraded Peppers.

Note that here $1/r$ and $1/s$ denote the step sizes τ and σ used in [5,15,31] for the compared methods. For the R-PDA, we find that when we take $(r, s, \eta) = (1, 20/3, -0.7)$ and the other parameters choice similar to that in last subsection, the R-PDA performs better. Besides, we use the same stopping criteria as that in (5.2) with $\mathrm{Tol} := 10^{-3}$ and the SNR value given by (5.3) to measure the quality of the restored images. All the tested methods initial their iterations with the degraded images.

We report the numerical results of these tested methods in Table 2. We find that R-PDA outperforms the other three methods in terms of the number of iterations and CPU computing time. Besides, R-PDA can achieve better quality than CP and PDHG, and also is competitive with HY-Algo1. In addition, we show the restored images by the tested algorithms in Figure 6 to further illustrate the efficiency of the proposed method.

6 Conclusions

In this paper, we proposed a modified primal-dual algorithm with refined step sizes, which is computationally cheap because the step size used in correction step is just a constant. Under the standard assumptions, the global convergence and the rate of convergence in ergodic and

Table 2: Numerical results for image inpainting problem

Algs. \ Images	House			Peppers		
	Iter.	CPU(s)	SNR	Iter.	CPU(s)	SNR
CP method	182	10.84	28.85	262	64.02	19.83
HY-Algo1	62	6.54	28.86	136	60.06	19.89
HY-Algo4	92	6.15	28.86	169	43.45	19.89
R-PDA	59	5.78	28.85	131	44.69	19.88

Figure 6: From left to right: Images House (the first row) and Peppers (the second row) restored by CP, PDHG, HY-Algo1 and R-PDA.

non-ergodic senses of the proposed method have proved. Some numerical results on image deblurring and inpainting problems were also reported to illustrate the efficiency of the proposed methods.

Acknowledgement

The authors are deeply grateful to thank the associate editor and two anonymous referees for their suggestions which largely improve the quality of our original manuscript.

References

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

[2] E. Blum and W. Oettli, *Mathematische Optimierung, Econometrics and Operations Research XX*, Springer Verlag, 1975.

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

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

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

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

[7] W. Deng, M.-J. Lai, Z.M. Peng and W.T. Yin, Parallel multi-block ADMM with $o(1/k)$ convergence, *J. Sci. Comput.* 71 (2017) 712–736.

[8] J. Douglas and H.H. Rachford, On the numerical solution of the heat conduction problem in 2 and 3 space variables, *Trans. American Math. Soc.* 82 (1956) 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) 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) 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, arXiv:1305.0546v2 [math.NA], 2015.

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

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

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

[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) 700–709.

[17] G.M. Korpelevič, An extragradient method for finding saddle points and for other problems, *Èkonom. i Mat. Metody* 12(4) (1976) 747–756.

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

[19] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, *Rev. Francaise d'Inform. Recherche Oper.* 4 (1970) 154–159.

[20] Y.E. Nesterov, A method for solving the convex programming problem with convergence rate $O(1/k^2)$, *Dokl. Akad. Nauk SSSR* 269 (1983) 543–547.

[21] Y.E. Nesterov, Gradient methods for minimizing composite functions, *Math. Program.* 140(1) (2013) 125–161.

[22] D. O'Connor and L. Vandenberghe, On the equivalence of the primal-dual hybrid gradient method and Douglas-Rachford splitting, *Math. Program.*, 179 (2020) 85–108.

[23] 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.

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

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

[26] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.* 14 (1976) 877–898.

[27] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Phys. D* 60 (1992) 227–238.

[28] 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.

[29] W. Yin, S. Osher, D. Goldfarb and J. Darbon, Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing, *SIAM J. Imaging Sci.* 1 (2008) 143–168.

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

[31] 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.

Manuscript received 25 January 219
revised 9 August 2019, 8 November 2019
accepted for publication 23 January 2020

Min Li
School of Management and Engineering
Nanjing University, Nanjing 210093, China
E-mail address: limin@nju.edu.cn

Xinxin Li
School of Mathematics, Jilin University
Changchun 130012, China
E-mail address: xinxinli@jlu.edu.cn

Zhongming Wu
School of Management Science and Engineering
Nanjing University of Information Science and Technology
Nanjing 210044, China
E-mail address: wuzm@nuist.edu.cn