An inexact primal-dual method with correction step for a saddle point problem in image debluring *

Changjie Fang,† Liliang Hu, Shenglan Chen

Key Lab of Intelligent Analysis and Decision on Complex Systems,

Chongqing University of Posts and Telecommunications, Chongqing 400065, China

School of Science, Chongqing University of Posts and Telecommunications,

Chongqing 400065, China

Abstract. In this paper,we present an inexact primal-dual method with correction step for a saddle point problem by introducing the notations of inexact extended proximal operators with symmetric positive definit ematrix D. Relaxing requirement on primal-dual step sizes, we prove the convergence of the proposed method. We also establish the $O(1/N)$ convergence rate of our method in the ergodic sense. Moreover, we apply our method to solve TV-L_1 image deblurring problems. Numerical simulation results illustrate the efficiency of our method.

Key Words. primal-dual method, inexact extended proximal operators, convergence rate, prediction and correction, image deblurring

1 Introduction

Let X and Y be two finite-dimensional real vector spaces equipped with an inner product $\langle\cdot,\cdot\rangle$ and a norm $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$. In this paper,we consider the following saddle point problem:

$$\min_{x\in X}\max_{y\in Y} L(x,y) = f(x) + \langle Ax, y\rangle - g(y) \tag{1.1}$$

where A is a bounded linear operator, $f : X \to (-\infty, +\infty]$ and $h : Y \to (-\infty, +\infty]$ are proper lower semicontinuous (l.s.c) convex functions.

Recall that (x^*, y^*) is called the saddle point of (1.1), if

$$L(x^*, y) \le L(x^*, y^*) \le L(x, y^*), \forall x \in X, y \in Y \tag{1.2}$$

Now we consider the primal problem

$$\min_{x\in X} f(x) + h(Ax) \tag{1.3}$$

*This work is partially supported by the National Natural Science Foundation of China (No. 11771350), Basic and Advanced Research Project of CQ CSTC (Nos. cstc2020jcyj-msxmX0738 and cstc2018jcyjAX0605)

†Corresponding author, E-mail address: fangcj@cqupt.edu.cn

together with its dual problem

$$\max_{y\in Y} f^*(-A^*y) + h^*(y) \tag{1.4}$$

where h^* denotes the Legendre-Fenchel conjugate of a convex l.s.c. function h, A^* denotes the adjoint of the bounded linear operator A.

If a primal-dual solution pair (x^*, y^*) of (1.3) and (1.4) exists, i.e.,

$$0 \in \partial f(x^*) + A^*y^*, 0 \in \partial h(Ax^*) - y^*,$$

then the problem (1.3) is equivalent to the following saddle-point formulation:

$$\min_{x\in X} \max_{y\in Y} f(x) + \langle Ax, y\rangle - h^*(y). \tag{1.5}$$

Hence, Problem (1.5) is a special case of Problem (1.1).

It is well known that many application problems can be formulated as the saddle point problem (1.1) such as image restoration, magnetic resonance imaging and computer vision; see, for example, [17, 19, 22, 27].

Two of the most popular approaches are first-order primal-dual methods [4, 9], in particular the Primal-Dual Hybrid Gradient (PDHG) method [27], and Alternating Direction Method of Multipliers (ADMM) method [2, 11]. For PDHG method, both a primal and a dual variable are updated in each iteration and thus some difficulties that arise when working only on the primal or dual variable can be avoided. In ADMM method, separating the minimization over the two primal variables into two steps is precisely what allows for decomposition when f or g, or both, are separable. In [13], it was showed that PDHG is not necessarily convergent even when the step sizes are fixed as tiny constants. In [9], PDHG was interpreted as projected averaged gradient method, and its convergence was studied by imposing additional restrictions ensuring that the step sizes λ and τ are small. In [4], a primal-dual method with inertial step $\theta(\theta \in [0, 1])$ was proposed (denote by PDI) with convergence rate $O(1/N)$ in terms of primal-dual gap, and for $\theta = 1$, the convergence of PDI was proved with the requirement on step sizes $\tau\lambda < 1/\|A^TA\|$. In [12], some prediction-correction contraction methods were presented in which the convergence was guaranteed with relaxed step sizes satisfying $\lambda\tau < 4/(1 + \theta)^2\|A^TA\|^2$ for $\theta \in (-1, 1]$. Further, a primal-dual method (named PDL) in the prediction-correction fashion was proposed in [25] and the pairwise primal-dual stepsizes λ_i and τ_i were relaxed to $\lambda_i\tau_i < 1(i = 1, 2)$.

As first-order methods, however, they are sensitive to problem conditions, and hence might be performed up to a certain precision, for example, due to the application of a proximal operator lacking a closed-form solution. This problem may arise from examples studied in, for example, [3, 8, 10, 15]. An absolute error criterion was adopted in [7], where the subproblem errors are controlled by a summable sequence of error tolerances. To simplify the choice of the sequences,a relative error criterion was introduced in [16], where the coresponding parameters are required to be square summable. [18] introduced four diferent types of inexact proxima, where all the controlled errors were required to be summable. In [14], the inexact preconditioned PDHG method was studied by the selection of appropriate preconditioners and the introduction of bounded relative error of the subproblem, where convergence was established in case the error was neither summable nor square summable.

Motivated by the research works [18, 25], in this paper, we introduce three different types of inexact extended proxima closely related to the extended proximal operator with the matrix D([6]). Applying these notions, we proposed an inexact primal-dual method with correction step for solving the saddle point problem. Under some mild conditions, the convergence of the proposed method is proved, in which we relax requirement on pairwise primal-dual stepsize , for example, compared with that in [25]. In [25], primal-dual stepsizes λ_i and $\tau_i (i = 1, 2)$ are required to satisfy $\lambda_i \tau_i < 1$. In our method, the sequences $\{\lambda_k\}$ and $\{\tau_k\}$ are nondecreasing and bounded satisfying $R - \tau_k \lambda_k S^{-1} \succ 0$, where R and S are symmetric positive definite matrices. We also establish the $O(1/N)$ convergence rate in the ergodic sense. At the same time, we establish the convergence rates in case error tolerances $\{\delta_n\}$ and $\{\varepsilon_n\}$ are required to decrease like $O(1/n^{\alpha+\frac{1}{2}})$ for some $\alpha > 0$; see Theorem 3.2. In the numerical experiments part, we investigate the applications of our method in $TV - L_1$ image deblurring. Firstly, we show that the type-2 approximation of the extended proximal point can be computed by approximately minimizing duality gap; see (4.10) in Section 4. Further, the duality gap is used as the stopping criterion of inner loop, i.e., the second subproblem; see (4.11). In addition, we discuss the sensitivity of parameters in Algorithm 1. Finally, we show the efficiency of our method in image deblurring compared with some existing methods, for example, [4, 18, 25].

The rest of this paper is organized as follows. In Section 2, we introduce the concepts of inexact extended proximal operators and present some auxiliary lemmas. In Section 3, we describe our method and prove the convergence of our method. At the same time, we also analyze the cnvergence rate. Numerical experiment results are reported in Section 4. Some conclusions are presented in Section 5.

2 Preliminaries

In this section, we shall introduce some definitions. Suppose that h be a convex function in $\mathbb{R}^n$, $D \in \mathbb{R}^{n\times n}$ a symmetric positive definite matrix and $\tau > 0$. For any $D \succ 0$ and given $y \in \mathbb{R}^n$, denote

$$G_y(x) := h(x) + \frac{1}{2\tau}\|x - y\|_D^2, \ \forall\ x \in \mathbb{R}^n, \tag{2.1}$$

and define the extended proximal operator of h as

$$\text{Prox}_{\tau h}^D(y) := \bar{z} = \arg\min_{x\in X} G_y(x), \tag{2.2}$$

where $\|x\|_D^2 = \langle x, Dx\rangle$ and D^{-1} denotes the inverse of D. Because D is symmetric positive definite, $\text{Prox}_{\tau h}^D(y)$ is unique(see Lemma 2.4).

Definition 2.1. *Let $\varepsilon \geq 0$. $z \in X$ is said to be a type-0 approximation of the extended proximal point $Prox_{\tau h}^D(y)$ with precision ε if*

$$z \approx_0^\varepsilon Prox_{\tau h}^D(y) \overset{def}{\Longleftrightarrow} \|z - \bar{z}\|_D \leq \sqrt{2\tau\varepsilon}$$

Next we recall the definition of $\varepsilon-$ subdifferential of h at z, denoted by $\partial_\varepsilon h(z)$:

$$\partial_\varepsilon h(z) = \{p \in X | h(x) \geq h(z) + \langle p, x - z\rangle - \varepsilon, \forall x \in X\}.$$

In the following ,we give the definition of $\varepsilon-$ subdifferential of G_y at z, denoted by $\partial_\varepsilon G_y(z)$:

$$\partial_\varepsilon G_y(z) := \{p \in X | G_y(x) \geq G_y(z) + \langle p, x - z\rangle - \varepsilon, \forall x \in X\}.$$

Definition 2.2. *Let $\varepsilon \geq 0$. $z \in X$ is said to be a type-1 approximation of the extended proximal point $Prox_{\tau h}^{D}(y)$ with precision ε if*

$$z \approx_1^{\varepsilon} Prox_{\tau h}^{D}(y) \overset{def}{\Longleftrightarrow} 0 \in \partial_{\varepsilon} G_y(z)$$

Definition 2.3. *Let $\varepsilon \geq 0$. $z \in X$ is said to be a type-2 approximation of the extended proximal point $Prox_{\tau h}^{D}(y)$ with precision ε if*

$$z \approx_2^{\varepsilon} Prox_{\tau h}^{D}(y) \overset{def}{\Longleftrightarrow} \frac{1}{\tau} D(y-z) \in \partial_{\varepsilon} h(z)$$

Remark 2.1. *If $D = I$, where I is the identity matrix, then the inexact extended proximal operators in Definitions 2.1-2.3 will reduce into the inexact proxima, for example, introduced in [18], respectively. Thus, Definitions 2.1-2.3 are the generalization of the correcponding definitions in [18].*

According to the above definition, we have the following lemmas.

Lemma 2.1. *Suppose $z \approx_1^{\varepsilon} \arg\min_{x \in X}\{h(x) + \frac{1}{2\tau}||x-y||_D^2\}$, then $z \in dom\, h$ and $z \approx_0^{\varepsilon} \arg\min_X\{h(x) + \frac{1}{2\tau}||x-y||_D^2\}$.*

Proof. According to Definition 2.2 and the definition of $\partial_{\varepsilon} G_y$, we have

$$G_y(x) \geq G_y(z) - \varepsilon, \quad \forall x \in X. \tag{2.3}$$

Setting $x = \bar{z}$ in (2.3) and using (2.2), from (2.3) we have

$$\|\bar{z} - z\|_D^2 \leq 2\tau\varepsilon + 2\tau[h(\bar{z}) - h(z) + \langle \frac{1}{\tau} D(\bar{z} - y), \bar{z} - z\rangle] \tag{2.4}$$

which implies that $z \in dom f$. According to the optimality condition of (2.2), we have

$$h(x) \geq h(\bar{z}) + \langle \frac{1}{\tau} D(y - \bar{z}), x - \bar{z}\rangle, \quad \forall x \in X. \tag{2.5}$$

Setting $x = z$ in (2.5) and substituting the resulting inequality into (2.4), we get

$$||\bar{z} - z||_D \leq \sqrt{2\tau\varepsilon}.$$

In view of Definite 2.1, we obtain the conclusion. □

Lemma 2.2. *Suppose thatz $\approx_1^{\varepsilon} \arg\min_{x \in X}\{h(x) + \frac{1}{2\tau}\|x-y\|_D^2\}$,then there exists $r \in X$ with$\|r\|_D \leq \sqrt{2\tau\varepsilon}$ such that*

$$\frac{1}{\tau} D(y - z - r) \in \partial_{\varepsilon} h(z)$$

.

Proof. According to Definition 2.2, from (2.1) we have

$$h(x) \geq h(z) + \langle x - z, \frac{1}{\tau} D(y - z - \frac{x-z}{2})\rangle - \varepsilon, \quad \forall x \in X.$$

Set $r = \frac{x-z}{2}$. By the definition of the $\varepsilon-$ subdifferential of h, the conclusion holds. □

Lemma 2.3. *Suppose $z \approx_2^{\varepsilon} \arg\min_{x \in X}\{h(x) + \frac{1}{2\tau}\|x-y\|_D^2\}$,then $z \approx_1^{\varepsilon} \arg\min_{x \in X}\{h(x) + \frac{1}{2\tau}\|x-y\|_D^2\}$*

Proof. According to Definition 2.3, from (2.1) we have

$$\begin{aligned}
G_y(x) &= h(x) + \frac{1}{2\tau}\|x-y\|_D^2 \\
&\geq h(z) + \langle \frac{1}{\tau}D(y-z), x-z\rangle - \varepsilon + \frac{1}{2\tau}\|x-y\|_D^2 \\
&= h(z) + \frac{1}{2\tau}\|z-y\|_D^2 + \frac{1}{2\tau}\|x-z\|_D^2 - \varepsilon \\
&= G_y(z) - \varepsilon + \frac{1}{2\tau}\|x-z\|_D^2 \\
&\geq G_y(z) - \varepsilon,
\end{aligned}$$

where the first inequality follows from the definition of $\partial_\varepsilon h(z)$ and the first equality follows from following identity equality

$$\|a-b\|_D^2 = \|a-c\|_D^2 + \|b-c\|_D^2 - 2\langle D(a-c), b-c\rangle. \tag{2.6}$$

Hence, $0 \in \partial_\varepsilon G_y(z)$. By Definition 2.2, the conclusion holds. □

The following lemma illustrate that the extended proximal operator (2.2) is well-defined.

Lemma 2.4.

$$\|\mathrm{Prox}_{\tau h}^D(y_1) - \mathrm{Prox}_{\tau h}^D(y_2)\|_D \leq \|y_1 - y_2\|_D.$$

Proof. Let $z_1 = \mathrm{Prox}_{\tau h}^D(y_1)$ and $z_2 = \mathrm{Prox}_{\tau h}^D(y_2)$. Since $z_1 = \mathrm{Prox}_{\tau h}^D(y_1)$, the optimality condition implies that $\frac{1}{\tau}D(y_1 - z_1) \in \partial h(z_1)$, and hence

$$h(x) \geq h(z_1) + \langle \frac{1}{\tau}D(y_1 - z_1), x - z_1\rangle, \quad \forall x \in X. \tag{2.7}$$

Set $x = z_2$ in (2.7) and get

$$h(z_2) \geq h(z_1) + \langle \frac{1}{\tau}D(y_1 - z_1), z_2 - z_1\rangle. \tag{2.8}$$

Similarly, we have

$$h(z_1) \geq h(z_2) + \langle \frac{1}{\tau}D(y_2 - z_2), z_1 - z_2\rangle. \tag{2.9}$$

Adding (2.8) and (2.9), and by a simple manipulation, we obtain

$$\|z_1 - z_2\|_D^2 \leq \langle D(y_2 - y_1), z_2 - z_1\rangle \leq \|y_2 - y_1\|_D\|z_2 - z_1\|_D,$$

i.e.,

$$\|\mathrm{Prox}_{\tau h}^D(y_1) - \mathrm{Prox}_{\tau h}^D(y_1)\|_D \leq \|y_2 - y_1\|_D.$$

□

The following lemma is crucial in proving the convergence of Algorithm 1.

Lemma 2.5. *Suppose that $g : X \mapsto \bar{R}$ is a convex function. For given $z_0, u, v \in X$ and $\tau, \varepsilon > 0$. Let*

$$z_1 \approx_2^\varepsilon \arg\min_{z\in X}\{g(z) + \frac{1}{2\tau}||z - (z_0 - D^{-1}u)||_D^2\}, \tag{2.10}$$

$$z_2 \approx_1^\varepsilon \arg\min_{z\in X}\{g(z) + \frac{1}{2\tau}||z - (z_0 - D^{-1}v)||_D^2\}, \tag{2.11}$$

then

(i)

$$0 \leq ||z_1 - z_2||_D \leq \frac{1}{2}(\sqrt{2\tau\varepsilon} + ||u-v||_{D^{-1}} + \sqrt{||u-v||^2_{D^{-1}} + 10\tau\varepsilon + 2\sqrt{2\tau\varepsilon}||u-v||_{D^{-1}}}) \quad (2.12)$$

(ii) $\forall z \in X$

$$g(z_1) - g(z) + \langle z_1 - z, \frac{1}{\tau}v\rangle \leq \frac{1}{2\tau}(||z - z_0||^2_D - ||z - z_2||^2_D) + \frac{1}{2\tau}(||D^{-1}(u-v)||^2_D - ||z_0 - z_1||^2_D) + 2\varepsilon + \sqrt{\frac{2\varepsilon}{\tau}}||z - z_2||_D \quad (2.13)$$

Proof. (i)In view of Definition 2.3 and (2.10), we have

$$\frac{1}{\tau}D((z_0 - D^{-1}u) - z_1) \in \partial_\varepsilon g(z_1). \quad (2.14)$$

By Lemma 2.2, there exists $r \in X$ with$||r||_D \leq \sqrt{2\tau\varepsilon}$ such that

$$\frac{1}{\tau}D((z_0 - D^{-1}v) - z_2 - r) \in \partial_\varepsilon g(z_2). \quad (2.15)$$

From the definition of $\varepsilon-$subdifferential and (2.15), we have

$$g(z) - g(z_2) \geq \frac{1}{\tau}\langle D((z_0 - D^{-1}v) - z_2 - r), z - z_2\rangle - \varepsilon, \forall z \in X$$

i.e.,

$$\langle D((z_0 - D^{-1}v) - z_2), z_2 - z\rangle \geq \tau(g(z_2) - g(z) - \varepsilon) + \langle Dr, z_2 - z\rangle, \forall z \in X. \quad (2.16)$$

Taking $z = z_1$ in (2.16), we have

$$\langle D(z_2 - z_0 + D^{-1}v), z_1 - z_2\rangle \geq \tau(g(z_2) - g(z_1) - \varepsilon) + \langle Dr, z_2 - z_1\rangle. \quad (2.17)$$

From (2.14), we have

$$\langle D((z_0 - D^{-1}u) - z_1), z_1 - z\rangle\tau(\geq g(z_1) - g(z) - \varepsilon). \quad (2.18)$$

Setting $z = z_2$ in (2.18), we have

$$\langle z_2 - z_1, D(z_1 - z_0 + D^{-1}u)\rangle \geq \tau(g(z_1) - g(z_2) - \varepsilon) \quad (2.19)$$

Adding (2.17) and (2.19), and by a simple manipulation, we get

$$\|z_1 - z_2\|^2_D \leq (\sqrt{2\tau\varepsilon} + \|u - v\|_{D^{-1}})\|z_1 - z_2\|_D + 2\tau\varepsilon \quad (2.20)$$

Therefore,

$$0 \leq \|z_1 - z_2\|_D \leq \frac{1}{2}(\sqrt{2\tau\varepsilon} + \|u-v\|_{D^{-1}} + \sqrt{\|u-v\|^2_{D^{-1}} + 10\tau\varepsilon + 2\sqrt{2\tau\varepsilon}\|u-v\|_{D^{-1}}})$$

(ii)Note that

$$\begin{aligned}
&\frac{1}{2}\|z - z_0\|^2_D - \frac{1}{2}\|z - z_2\|^2_D \\
&= \frac{1}{2}\|z_2\|^2_D - \frac{1}{2}\|z_0\|^2_D - \langle z_2 - z_0, Dz_0\rangle \\
&\quad - \langle z - z_2, v\rangle + \langle z - z_2, D(z_2 - z_0 + D^{-1}v)\rangle \\
&\geq \frac{1}{2}||z_2||^2_D - \frac{1}{2}||z_0||^2_D - \langle z_2 - z_0, Dz_0\rangle - \langle z_1 - z_2, v\rangle \\
&\quad + \langle z_1 - z, v\rangle + \tau(g(z_2) - g(z) - \varepsilon) + \langle Dr, z_2 - z\rangle,
\end{aligned}$$

where the equality follows from (2.6) and the inequality follows from (2.16). Hence,

$$\langle z_1 - z, v\rangle \leq \frac{1}{2}||z - z_0||_D^2 - \frac{1}{2}||z - z_2||_D^2 - \xi \\ - \tau(g(z_2) - g(z) - \varepsilon) - \langle Dr, z_2 - z\rangle, \tag{2.21}$$

where $\xi = \frac{1}{2}||z_2||_D^2 - \frac{1}{2}||z_0||_D^2 - \langle z_2 - z_0, Dz_0\rangle - \langle z_1 - z_2, v\rangle$.

Also,

$$\begin{aligned}\xi &= \frac{1}{2}||z_2||_D^2 - \frac{1}{2}||z_0||_D^2 - \langle z_2 - z_0, Dz_0\rangle - \langle z_1 - z_2, v - u\rangle \\ &\quad + \langle z_2 - z_1, D(z_0 - z_1)\rangle + \langle z_2 - z_1, D(D^{-1}u - z_0 + z_1)\rangle \\ &\geq \frac{1}{2}||z_2||_D^2 - \frac{1}{2}||z_0||_D^2 - \langle z_2 - z_0, Dz_0\rangle - \langle z_1 - z_2, v - u\rangle \\ &\quad + \langle z_2 - z_1, D(z_0 - z_1)\rangle + \tau(g(z_1) - g(z_2) - \varepsilon) \\ &= \frac{1}{2}||z_2 - z_1||_D^2 + \frac{1}{2}||z_0 - z_1||_D^2 \\ &\quad - \langle z_1 - z_2, u - v\rangle + \tau(g(z_1) - g(z_2) - \varepsilon) \\ &\geq \frac{1}{2}||z_2 - z_1||_D^2 + \frac{1}{2}||z_0 - z_1||_D^2 - \frac{1}{2}||z_2 - z_1||_D^2 \\ &\quad - \frac{1}{2}||D^{-1}(u - v)||_D^2 + \tau(g(z_1) - g(z_2) - \varepsilon) \\ &= \frac{1}{2}||z_0 - z_1||_D^2 - \frac{1}{2}||D^{-1}(u - v)||_D^2 + \tau(g(z_1) - g(z_2) - \varepsilon),\end{aligned} \tag{2.22}$$

where the first inequality follows from (2.19) and the second one is due to the following inequality

$$\langle Dp, q\rangle \leq \frac{1}{2}\|p\|_D + \frac{1}{2}\|q\|_D.$$

Combining (2.21) with (2.22), we have

$$\begin{aligned}< z_1 - z, v > &\leq \frac{1}{2}(||z - z_0||_D^2 - ||z - z_2||_D^2) + \frac{1}{2}||D^{-1}(u - v)||_D^2 \\ &\quad - \frac{1}{2}||z_0 - z_1||_D^2 + \tau(g(z) - g(z_1) + \varepsilon) - \langle Dr, z_2 - z\rangle + 2\tau\varepsilon \\ &\leq \frac{1}{2}(||z - z_0||_D^2 - ||z - z_2||_D^2) + \frac{1}{2}||D^{-1}(u - v)||_D^2 \\ &\quad - \frac{1}{2}||z_0 - z_1||_D^2 + \tau(g(z) - g(z_1) + \varepsilon) + \sqrt{2\tau\varepsilon}||z - z_2||_D + 2\tau\varepsilon,\end{aligned}$$

where the second inequality follows from the Cauchy-Schwarz inequality.

Therefore, multiplying both sides of the above inequality by $\frac{1}{\tau}$ yields

$$\begin{aligned}g(z_1) - g(z) + < z_1 - z, \frac{1}{\tau}v > &\leq \frac{1}{2\tau}(||z - z_0||_D^2 - ||z - z_2||_D^2) \\ &\quad + \frac{1}{2\tau}(||D^{-1}(u - v)||_D^2 - ||z_0 - z_1||_D^2) + 2\varepsilon + \sqrt{\frac{2\varepsilon}{\tau}}||z - z_2||_D.\end{aligned}$$

□

3 Main results

Now we present the inexact primal-dual method for solving (1.1).

Algorithm 1 Inexact Primal-Dual Method with Correction Step

Initialization: $x^0 \in X, \bar{y}^0 \in Y, \tau_0, \lambda_0 > 0$.

Iteration:

$$y^{k+1} \approx_2^{\frac{\varepsilon_{k+1}}{2}} \arg\max_{y \in Y} L(x^k, y) - \frac{1}{2\tau_k}||y - \bar{y}^k||_S^2 \tag{3.1}$$

$$x^{k+1} \approx_2^{\delta_{k+1}} \arg\min_{x \in X} L(x, y^{k+1}) + \frac{1}{2\lambda_k}||A(x - x^k)||_R^2 \tag{3.2}$$

$$\bar{y}^{k+1} \approx_1^{\frac{\varepsilon_{k+1}}{2}} \arg\max_{y \in Y} L(x^{k+1}, y) - \frac{1}{2\tau_k}||y - \bar{y}^k||_S^2 \tag{3.3}$$

Until meet stopping criterion.

Next we will analyze the convergence of Algorithm 1. Firstly, we prove two important lemmas which will be used in the sequence.

Lemma 3.1. *Let $(y^{k+1}, x^{k+1}, \bar{y}^{k+1})$ be obtained from Algorithm 1, then for any pair $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ we have*

$$\begin{aligned} L(x^{k+1}, y) - L(x^{k+1}, y^{k+1}) \leq{}& \frac{1}{2\tau_k}(||y - \bar{y}^k||_S^2 - ||y - \bar{y}^k||_S^2) + \sqrt{\frac{\varepsilon_{k+1}}{\tau_k}}||y - \bar{y}^{k+1}||_S \\ &+ \frac{\tau_k}{2}||A(x^{k+1} - x^k)||_{S^{-1}}^2 + 2\varepsilon_{k+1} - \frac{1}{2\tau_k}||\bar{y}^k - y^{k+1}||_S^2. \end{aligned} \tag{3.4}$$

Proof. From Algorithm 1, it is very easy to deduce that the formulas (3.1) and (3.2) are equivalent to the following ones

$$y^{k+1} \approx_2^{\frac{\varepsilon_{k+1}}{2}} \arg\min_{y \in Y}\{g(y) + \frac{1}{2\tau_k}\|y - \bar{y}^k - \tau_k S^{-1}Ax^k\|_S^2\}$$

and

$$\bar{y}^{k+1} \approx_1^{\frac{\varepsilon_{k+1}}{2}} \arg\min_{y \in Y}\{g(y) + \frac{1}{2\tau_k}\|y - \bar{y}^k - \tau_k S^{-1}Ax^{k+1}\|_S^2\},$$

respectively.

Setting $\tau = \tau_k, \varepsilon = \frac{\varepsilon_{k+1}}{2}, D = S, z = y, z_0 = \bar{y}^k, z_1 = y^{k+1}, z_2 = \bar{y}^{k+1}, u = -\tau_k Ax^k, v = -\tau_k Ax^{k+1}$ in Lemma 2.5(ii), we get

$$\begin{aligned} g(y^{k+1}) - g(y) + \langle y^{k+1} - y, Ax^{k+1}\rangle \leq{}& \frac{1}{2\tau_k}(\|y - \bar{y}^k\|_S^2 - \|y - \bar{y}^{k+1}\|_S^2) \\ &+ \frac{\tau_k}{2}\|A(x^k - x^{k+1})\|_{S^{-1}}^2 + 2\varepsilon_{k+1} \\ &+ \sqrt{\frac{\varepsilon_{k+1}}{\tau_k}}\|y - \bar{y}^{k+1}\|_S - \frac{1}{2\tau_k}\|\bar{y}^k - y^{k+1}\|_S^2. \end{aligned}$$

Hence

$$\begin{aligned}L(x^{k+1},y)-L(x^{k+1},y^{k+1})&=g(y^{k+1})-g(y)-\langle y^{k+1}-y,Ax^{k+1}\rangle\\&\leq\frac{1}{2\tau_k}(\|y-\bar{y}^k\|_S^2-\|y-\bar{y}^{k+1}\|_S^2)+\frac{\tau_k}{2}\|A(x^k-x^{k+1})\|_{S^{-1}}^2+2\varepsilon_{k+1}\\&\quad+\sqrt{\frac{\varepsilon_{k+1}}{\tau_k}}\|y-\bar{y}^{k+1}\|_S-\frac{1}{2\tau_k}\|\bar{y}^k-y^{k+1}\|_S^2.\end{aligned}$$

This completes the proof. □

Lemma 3.2. *Let $(y^{k+1},x^{k+1},\bar{y}^{k+1})$ be obtained from Algorithm 1, then for any $x\in\mathbb{R}^n$ we have*

$$L(x^{k+1},y^{k+1})-L(x,y^{k+1})\leq\frac{1}{2\lambda_k}[(\|x-x^k\|_{A^TRA}^2-\|x-x^{k+1}\|_{A^TRA}^2-\|x^{k+1}-x^k\|_{A^TRA}^2]+\delta_{k+1}. \tag{3.5}$$

Proof. By Definition 2.3, the optimal condition of (3.2) yields

$$\frac{1}{\lambda_k}A^TRA(x^k-x^{k+1})\in\partial_{\delta_{k+1}}L(x^{k+1},y^{k+1}).$$

In view of the definition of $\varepsilon-$ subdifferential, we have

$$L(x^{k+1},y^{k+1})-L(x,y^{k+1})\leq\frac{1}{\lambda_k}\langle A^TRA(x^k-x^{k+1}),x^{k+1}-x)\rangle+\delta_{k+1}. \tag{3.6}$$

Setting $a:=x^k,b:=x,c:=x^{k+1},D:=A^TRA$ in (2.6), we get

$$\langle A^TRA(x^k-x^{k+1}),x^{k+1}-x)\rangle=-\frac{1}{2}[\|x^k-x^{k+1}\|_{A^TRA}^2+\|x-x^{k+1}\|_{A^TRA}^2-\|x-x^k\|_{A^TRA}^2]. \tag{3.7}$$

Combining (3.6) with (3.7), we know that (3.5) holds. □

The following two lemmas play an important role in proving the convergence of Algorithm 1.

Lemma 3.3. *([21])Assume that the sequence $\{\mu_N\}$ is nonnegaive and satisfies the recursion*

$$\mu_N^2\leq T_N+\sum_{n=1}^{N}\sigma_n\mu_n$$

for all $N\geq 1$, where $\{T_N\}$ is an increasing sequence, $T_0\geq\mu_0^2$, and $\sigma_n\geq 0$ for all $n\geq 0$. Then for all $N\geq 1$

$$\mu_N\leq\frac{1}{2}\sum_{n=1}^{N}\sigma_n+(T_N+(\frac{1}{2}\sum_{n=1}^{N}\sigma_n)^2)^{\frac{1}{2}}$$

Set

$$\hat{x}^N:=\frac{1}{N}\sum_{k=0}^{N-1}x^{k+1}\quad\text{and}\quad\hat{y}^N:=\frac{1}{N}\sum_{k=0}^{N-1}y^{k+1}. \tag{3.8}$$

Lemma 3.4. *Let the sequence $\{(x^{k+1},y^{k+1},\bar{y}^{k+1})\}$ be obtained by Algorithm 1 and $(\hat{x}^N,\hat{y}^N)$ defined by (3.8). Suppose that $\{\tau_k\}$ and $\{\lambda_k\}$ are nondecreasing and $R-\tau_k\lambda_kS^{-1}\succ 0$. Then for every saddle point $(x^*,y^*)\in X\times Y$ of (1.1), we have*

$$L(\hat{x}^N,y^*)-L(x^*,\hat{y}^N)\leq\frac{1}{2N\tau_N}[\sqrt{\frac{\tau_N}{\tau_0}}||y^*-\bar{y}^0||_S+\sqrt{\frac{\tau_N}{\lambda_0}}||x^*-x^0||_{A^TRA}+2A_N+\sqrt{2B_N}]^2, \tag{3.9}$$

where $A_N:=\sum_{k=0}^{N-1}\tau_N\sqrt{\frac{\varepsilon_{k+1}}{\tau_k}}$ and $B_N:=\sum_{k=0}^{N-1}\tau_N(2\varepsilon_{k+1}+\delta_{k+1})$.

Proof. Adding (3.4) and (3.5) yields

$$L(x^{k+1},y)-L(x,y^{k+1})\le \frac{1}{2\tau_k}(||y-\bar{y}^k||_S^2-||y-\bar{y}^{k+1}||_S^2)+\frac{1}{2\lambda_k}(||x-x^k||^2_{A^TRA}-||x-x^{k+1}||^2_{A^TRA})$$
$$+\sqrt{\frac{\varepsilon_{k+1}}{\tau_k}}||y-\bar{y}^{k+1}||_S+2\varepsilon_{k+1}+\delta_{k+1}-\frac{1}{2\lambda_k}||A(x^k-x^{k+1})||^2_{R-\tau_k\lambda_kS^{-1}}-\frac{1}{2\tau_k}||\bar{y}^k-y^{k+1}||^2_S \quad (3.10)$$

Since $\{\tau_k\}$ and $\{\lambda_k\}$ are nondecreasing and $R-\tau_k\lambda_kS^{-1}\succ 0$,

$$\begin{aligned}L(x^{k+1},y)-L(x,y^{k+1})\le{}& \frac{1}{2\tau_k}||y-\bar{y}^k||^2_S-\frac{1}{2\tau_{k+1}}||y-\bar{y}^{k+1}||^2_S\\&+\frac{1}{2\lambda_k}||x-x^k||^2_{A^TRA}-\frac{1}{2\lambda_{k+1}}||x-x^{k+1}||^2_{A^TRA}\\&+\sqrt{\frac{\varepsilon_{k+1}}{\tau_k}}||y-\bar{y}^{k+1}||_S+2\varepsilon_{k+1}+\delta_{k+1}.\end{aligned}\quad (3.11)$$

Since $L(x,y)$ and $-L(x,y)$ are convex with respect to x and y respectively, using Jensen inequality and (3.11), we have

$$\begin{aligned}N(L(\hat{x}^N,y)-L(x,\hat{y}^N))\le{}&\sum_{k=0}^{N-1}L(x^{k+1},y)-L(x,y^{k+1})\\\le{}&\frac{1}{2\tau_0}||y-\bar{y}^0||^2_S-\frac{1}{2\tau_N}||y-\bar{y}^N||^2_S+\frac{1}{2\lambda_0}||x-x^0||^2_{A^TRA}\\&-\frac{1}{2\lambda_N}||x-x^N||^2_{A^TRA}+\sum_{k=0}^{N-1}\sqrt{\frac{\varepsilon_{k+1}}{\tau_k}}||y-\bar{y}^{k+1}||_S+\sum_{k=0}^{N-1}(2\varepsilon_{k+1}+\delta_{k+1}).\end{aligned}\quad (3.12)$$

Setting $x:=x^*$ and $y:=y^*$ in (3.12) and using (1.2) we have

$$\begin{aligned}||y^*-\bar{y}^N||^2_S\le{}&\frac{\tau_N}{\tau_0}||y^*-\bar{y}^0||^2_S+\frac{\tau_N}{\lambda_0}||x^*-x^0||^2_{A^TRA}\\&+\sum_{k=0}^{N-1}\tau_N\sqrt{\frac{\varepsilon_{k+1}}{\tau_k}}||y^*-\bar{y}^{k+1}||_S+2\sum_{k=0}^{N-1}\tau_N(2\varepsilon_{k+1}+\delta_{k+1}).\end{aligned}$$

Set $\mu_N=||y^*-\bar{y}^N||_S, T_N=\frac{\tau_N}{\tau_0}||y^*-\bar{y}^0||^2_S+\frac{\tau_N}{\lambda_0}||x^*-x^0||^2_{A^TRA}+2B_N, \sigma_k=2\tau_N\sqrt{\frac{\varepsilon_{k+1}}{\tau_k}}$ in Lemma 3.3. Obviously, $T_0\ge\mu_0^2$ and $\sigma_k\ge 0$. Thus,

$$||y^*-\bar{y}^N||_S\le A_N+(\frac{\tau_N}{\tau_0}||y^*-\bar{y}^0||^2_S+\frac{\tau_N}{\lambda_0}||x^*-x^0||^2_{A^TRA}+2B_N+A_N^2)^{\frac{1}{2}}$$

Since A_N,B_N,τ_k and λ_k are nondecreasing, we have for all $k\le N$,

$$\begin{aligned}||y^*-\bar{y}^k||_S\le{}&A_k+(\frac{\tau_k}{\tau_0}||y^*-\bar{y}^0||^2_S+\frac{\tau_k}{\lambda_0}||x^*-x^0||^2_{A^TRA}+2B_k+A_k^2)^{\frac{1}{2}}\\\le{}&A_N+(\frac{\tau_N}{\tau_0}||y^*-\bar{y}^0||^2_S+\frac{\tau_N}{\lambda_0}||x^*-x^0||^2_{A^TRA}+2B_N+A_N^2)^{\frac{1}{2}}\\\le{}&2A_N+\sqrt{\frac{\tau_N}{\tau_0}}||y^*-\bar{y}^0||_S+\sqrt{\frac{\tau_N}{\lambda_0}}||x^*-x^0||_{A^TRA}+\sqrt{2B_N}.\end{aligned}\quad (3.13)$$

Hence, setting $x:=x^*$ and $y=:y^*$ in(3.12), and using (3.13) we have

$$\begin{aligned}N(L(\hat{x}^N,y^*)-L(x^*,\hat{y}^N))\le{}&\frac{1}{2\tau_0}||y^*-\bar{y}^0||^2_S+\frac{1}{2\lambda_0}||x^*-x^0||^2_{A^TRA}+\frac{1}{\tau_N}B_N\\&+\frac{1}{\tau_N}A_N(2A_N+\sqrt{\frac{\tau_N}{\tau_0}}||y^*-\bar{y}^0||_S+\sqrt{\frac{\tau_N}{\lambda_0}}||x^*-x^0||_{A^TRA}+\sqrt{2B_N})\\\le{}&\frac{1}{2\tau_N}(\sqrt{\frac{\tau_N}{\tau_0}}||y^*-\bar{y}^0||_S+\sqrt{\frac{\lambda_N}{\tau_0}}||y^*-\bar{y}^0||_{A^TRA}+2A_N+\sqrt{2B_N})^2,\end{aligned}$$

which implies that (3.9) holds. This completes the proof. □

Remark 3.1. *If, in addition, A_N and B_N are summable and $\{\tau_k\}$ is bounded above, then from (3.9) we can establish the $O(1/N)$ convergence rate of our method in the ergodic sense.*

Theorem 3.1. *Let $\{x^{k+1}, y^{k+1}, \bar{y}^{k+1}\}$ be the sequence pair generated by Algorithm 1 and $\{\widehat{x}^N, \widehat{y}^N\}$ be defined by (3.8) in Theorem 3.4. Suppose that the assumptions of Theorem 3.4 hold and $\tau_k \le \overline{\tau}, \lambda_k \le \overline{\lambda}$ with $R - \overline{\tau}\overline{\lambda}S^{-1} \succ 0$. If the partial sums A_N and B_N in Theorem 3.4 are summable and A is of full column rank, then every weak cluster point $(\widehat{x}, \widehat{y})$ of $\{\widehat{x}^N, \widehat{y}^N\}$ is a saddle point of problem (1.1). Moreover, if the dimension of X and Y is finite, then there exists a saddle point $(\widehat{x}, \widehat{y}) \in X \times Y$ such that $x^k \to \widehat{x}$ and $y^k \to \widehat{y}$ as $k \to \infty$.*

Proof. Since A_N and B_N are summable,

$$
\begin{aligned}
&(\sqrt{\frac{\tau_N}{\tau_0}}\|y^* - \bar{y}^0\|_S + \sqrt{\frac{\lambda_N}{\tau_0}}\|y^* - \bar{y}^0\|_{A^TRA} + 2A_N + \sqrt{2B_N})^2 \\
&\le (\sqrt{\frac{\bar{\tau}}{\tau_0}}\|y^* - \bar{y}^0\|_S + \sqrt{\frac{\bar{\lambda}}{\tau_0}}\|y^* - \bar{y}^0\|_{A^TRA} + 2A_N + \sqrt{2B_N})^2 := C_1 < +\infty
\end{aligned}
$$

From (3.13),we know that for all $k \le N$, $\|y^* - \bar{y}^k\|_S \le C_2 < +\infty$. By the same argumentation as for $\bar{y}^k$, from (3.12) we obtain $\|x^* - x^k\|_{A^TRA} < \infty$ for all $k \le N$ and hence $\{x^k\}$ is bounded, which implies the boundedness of $\{\widehat{x}^N\}$. Let $x := x^*$ and $y = y^*$ in (3.10) and then sum the resulting inequality from $k = 0$ to $N - 1$ to obtain

$$
\begin{aligned}
\frac{1}{2\bar{\lambda}}\sum_{k=0}^{N-1}\|x^{k+1} - x^k\|^2_{A^T(R-\bar{\tau}\bar{\lambda}S^{-1})A} + \sum_{k=0}^{N-1}\frac{1}{2\bar{\tau}}\|\bar{y}^k - y^{k+1}\|_S^2 \le \frac{1}{2\tau_0}\|y^* - \bar{y}^0\|_S^2 - \frac{1}{2\tau_N}\|y^* - \bar{y}^N\|_S^2 \\
+ \frac{1}{2\lambda_0}\|x^* - x^0\|^2_{A^TRA} - \frac{1}{2\lambda_N}\|x^* - x^N\|^2_{A^TRA} + \frac{C_2}{\tau_N}A_N + \frac{1}{\tau_N}B_N := C_3 < +\infty \quad (3.14)
\end{aligned}
$$

Letting $N \to \infty$ in (3.14), we have

$$
x^k - x^{k+1} \to 0 \quad \text{and} \quad \bar{y}^k - y^{k+1} \to 0 \quad \text{as} \quad k \to \infty. \tag{3.15}
$$

Hence, $\{\bar{y}^k - y^{k+1}\}$ is bounded. Thus,

$$
\|y^* - y^{k+1}\|_S \le \|y^* - \bar{y}^k\|_S + \|\bar{y}^k - y^{k+1}\|_S < +\infty,
$$

i.e.,$\{y^k\}$ is bounded, and hence $\{\widehat{y}^N\}$ is also bounded. Hence there exists a subsequence $(\widehat{x}^{N_i}, \widehat{y}^{N_i})$ weakly converging to a cluster point$(\widehat{x}, \widehat{y})$. Since f and g are l.s.c.(thus also weakly l.s.c.), from (3.9) we have

$$
L(\widehat{x}, y^*) - L(x^*, \widehat{y}) = \lim_{i\to\infty} L(\widehat{x}^{N_i}, y^*) - L(x^*, \widehat{y}^{N_i}) \le \lim_{i\to\infty}\frac{C_1}{2\bar{\tau}N_i} = 0
$$

which implies that $(\widehat{x}, \widehat{y})$ is a saddle point of $L(x, y)$.

Now suppose that the dimensions of X and Y are finite. Since the sequence pair (x^k, y^k) is bounded, there exists a subsequence (x^{k_i}, y^{k_i}) strongly converging to a cluster point $(\widehat{x}, \widehat{y})$. Since $(\widehat{x}, \widehat{y})$ is a saddle point of $L(x, y)$, by replacing (x^*, y^*) with $(\widehat{x}, \widehat{y})$ in (3.10), we know that (3.15) holds. Hence, $x^{k_i-1} - x^{k_i} \to 0$ and $\bar{y}^{k_i-1} - y^{k_i} \to 0$ as $i \to \infty$. Hence,

$$
||x^{k_i-1} - \widehat{x}||_{A^TRA} \le ||x^{k_i-1} - x^{k_i}||_{A^TRA} + ||x^{k_i} - \widehat{x}||_{A^TRA} \to 0 \text{ as } i \to \infty
$$

i.e., $x^{k_i-1} \to \widehat{x}$. Let now $x^{k+1} = H(x^k)$ denote (3.20) in Algorithm 2 and $x^{k+1} = H_{\delta_{k+1}}(x^k)$ denote (3.2). In view of the continuity of H , we have

$$\begin{aligned}||\widehat{x} - H(\widehat{x})||_{A^TRA} &= \lim_{i\to\infty} ||x^{k_i-1} - H(x^{k_i-1})||_{A^TRA} \\ &\le \lim_{i\to\infty} (||x^{k_i-1} - H_{\delta_{k_i}}(x^{k_i-1})||_{A^TRA} + ||H_{\delta_{k_i}}(x^{k_i-1}) - H(x^{k_i-1})||_{A^TRA}) \\ &\le \lim_{i\to\infty} (||x^{k_i-1} - x^{k_i}||_{A^TRA} + \sqrt{2\bar{\lambda}\delta_{k_i}}) = 0\end{aligned}$$

where the last inequality follows from Lemma 2.3, Lemma 2.1 and Definition 2.1.

Let $y^{k+1} = \Gamma(\bar{y}^k)$ and $\bar{y}^{k+1} = \Psi(\bar{y}^k)$ in Algorithm 1, and $y^{k+1} = \Gamma_{\varepsilon_{k+1}}(\bar{y}^k)$ and $\bar{y}^{k+1} = \Psi_{\varepsilon_{k+1}}(\bar{y}^k)$ in Algorithm 2. Hence

$$\begin{aligned}||\widehat{y} - \Gamma\circ\Psi(\widehat{y})||_S &= \lim_{i\to\infty} \|y^{k_i} - \Gamma\circ\Psi(\bar{y}^{k_i-1})\|_S \\ &\le \lim_{i\to\infty} (||y^{k_i} - \Gamma\circ\Psi_{\varepsilon_{k_i}}(\bar{y}^{k_i-1})||_S + ||\Gamma\circ\Psi_{\varepsilon_{k_i}}(\bar{y}^{k_i-1}) - \Gamma\circ\Psi(y^{k_i-1})||_S \\ &\le \lim_{i\to\infty} (||\Gamma_{\varepsilon_{k_i}}(\bar{y}^{k_i-1}) - \Gamma(\bar{y}^{k_i})||_S + ||\Psi_{\varepsilon_{k_i}}(\bar{y}^{k_i-1}) - \Psi(\bar{y}^{k_i-1})||_S \\ &\le \lim_{i\to\infty} (||\Gamma_{\varepsilon_{k_i}}(\bar{y}^{k_i-1}) - \Gamma(\bar{y}^{k_i-1})||_S + ||\Gamma(\bar{y}^{k_i-1}) - \Gamma(\bar{y}^{k_i})||_S + +\sqrt{2\bar{\tau}\varepsilon_{k_i}}) \\ &\le \lim_{i\to\infty} [(2\sqrt{2\bar{\tau}\varepsilon_{k_i}}) + ||\bar{y}^{k_i-1} - \bar{y}^{k_i}||_S] = \lim_{i\to\infty} ||\bar{y}^{k_i-1} - \bar{y}^{k_i}||_S \\ &\le \lim_{i\to\infty} (||\bar{y}^{k_i-1} - y^{k_i}||_S + ||\bar{y}^{k_i} - y^{k_i}||_S) \le \lim_{i\to\infty} ||\bar{y}^{k_i} - y^{k_i}||_S)\end{aligned}$$

i.e.

$$||\widehat{y} - \Gamma\circ\Psi(\widehat{y})||_S \le \lim_{i\to\infty} ||\bar{y}^{k_i} - y^{k_i}||_S. \tag{3.16}$$

In view of Lemma 2.5(i), we have

$$\begin{aligned}0 \le ||y^k - \bar{y}^k||_S \le \frac{1}{2}(&\sqrt{2\bar{\tau}\varepsilon_k} + ||x^k - x^{k-1}||_{A^TS^{-1}A} \\ &+ \sqrt{||x^k - x^{k-1}||^2_{A^TS^{-1}A} + 10\bar{\tau}\varepsilon_k + 2\sqrt{2\bar{\tau}\varepsilon_k}||x^k - x^{k-1}||_{A^TS^{-1}A}})\end{aligned} \tag{3.17}$$

Since $x^{k_i-1} - x^{k_i} \to 0 (i \to \infty)$,taking $k = k_i$ and letting $i \to \infty$ in the above formula, we get

$$\lim_{k\to\infty} ||y^{k_i} - \bar{y}^{k_i}||_S = 0$$

Hence, from (3.16) we obtain $\widehat{y} = \Gamma\circ\Psi(\widehat{y})$. Since $x^* = H(x^*)$,it follows that $(\widehat{x}, \widehat{y})$ is a fixed point of Algorithm 1 and hence a saddle point of problem (1.1).We now use $(x, y) = (\widehat{x}, \widehat{y})$ in (3.11) and sum from $k = k_i, ..., N-1$ to obtain

$$\begin{aligned}\frac{1}{2\bar{\lambda}}||\widehat{x} - x^N||^2_{A^TRA} + \frac{1}{2\bar{\tau}}||\widehat{y} - \bar{y}^N||^2_S &\le \frac{1}{2\tau_1}||\widehat{y} - \bar{y}^{k_i}||^2_S + \frac{1}{2\lambda_1}||\widehat{x} - x^{k_i}||^2_{A^TRA} \\ &+ \sum_{k=k_i}^{N-1}\sqrt{\frac{\varepsilon_{k+1}}{\tau_1}}||\widehat{y} - \bar{y}^{k+1}||_S + \sum_{k=k_i}^{N-1}(2\varepsilon_{k+1} + \delta_{k+1})\end{aligned} \tag{3.18}$$

Since $\varepsilon_k \to 0$ and $\delta_k \to 0$ as $k \to \infty$, the right hand size in (3.18) tends to zero for $i \to \infty$, which implies that also $x^N \to \widehat{x}$ and $\bar{y}^N \to \widehat{y}$ for $N \to \infty$. Since $x^N \to \widehat{x}$ as $N \to \infty$, it is easy to see that $\lim_{N\to\infty} \|x^N - x^{N-1}\|_{A^TS^{-1}A} = 0$. Taking $k = N$ in(3.17) we have $\|y^N - \bar{y}^N\|_S \to 0(N \to \infty)$. Therefore,

$$\lim_{N\to\infty} \|y^N - \widehat{y}\|_S \le \lim_{N\to\infty} (\|y^N - \bar{y}^N\|_S + \|\bar{y}^N - \widehat{y}\|_S) = 0$$

Thus,$x^N \to \widehat{x}(N \to \infty)$. This completes the proof. □

Next we will establish convergence rates of our method, provided that $\{\delta_n\}$ and $\{\varepsilon_n\}$ decrease like $\mathcal{O}(\frac{1}{n^{\alpha+\frac{1}{2}}})$. We first review the following lemma.

Lemma 3.5. *([18]) For $\omega > -1$, let $s_N := \sum_{n=1}^{N} n^{\omega}$. Then*

$$s_N = \mathcal{O}(N^{1+\omega})$$

Theorem 3.2. *If $\alpha > 0$ and $\delta_n = \mathcal{O}(\frac{1}{n^{\alpha+\frac{1}{2}}}), \varepsilon_n = \mathcal{O}(\frac{1}{n^{\alpha+\frac{1}{2}}})$,then*

$$L(x^N, y^*) - L(x^*, y^N) = \begin{cases} \mathcal{O}(\frac{1}{N}), & \alpha > \frac{1}{2} \\ \mathcal{O}(\frac{\ln^2(N)}{N}), & \alpha = \frac{1}{2} \\ \mathcal{O}(N^{-2\alpha}), & \alpha \in (0, \frac{1}{2}) \end{cases}$$

Proof. If $\alpha > \frac{1}{2}$, then A_N and B_N in(3.9) are bounded. Hence,

$$L(x^N, y^*) - L(x^*, y^N) = \mathcal{O}(\frac{1}{k}).$$

If $\alpha = \frac{1}{2}$, then B_N is bounded and $A_N = \mathcal{O}(\ln(N))$. Hence,

$$L(x^N, y^*) - L(x^*, y^N) = \mathcal{O}(\frac{\ln^2(N)}{N}).$$

If $\alpha \in (0, \frac{1}{2})$, from Lemma 3.5 we have

$$L(x^N, y^*) - L(x^*, y^N) = \mathcal{O}(N^{-2\alpha}).$$

□

Next we consider the two special cases of Algorithm 1.

If we take $\varepsilon_{k+1} = \delta_{k+1} \equiv 0$ in Algorithm 1, then Algorithm 1 reduces to the following one:

Algorithm 2 Primal-Dual Method with Correction Step-A

Initialization: $x^0 \in X, \bar{y}^0 \in Y, \tau_0, \lambda_0 > 0$.

Iteration:

$$y^{k+1} = \arg\max_{y \in Y} L(x^k, y) - \frac{1}{2\tau_k}||y - \bar{y}^k||_S^2, \tag{3.19}$$

$$x^{k+1} = \arg\min_{x \in X} L(x, y^{k+1}) + \frac{1}{2\lambda_k}||A(x - x^k)||_R^2, \tag{3.20}$$

$$\bar{y}^{k+1} = \arg\max_{y \in Y} L(x^{k+1}, y) - \frac{1}{2\tau_k}||y - \bar{y}^k||_S^2, \tag{3.21}$$

Until meet stopping criterion.

If, in Algorithm 2, we take $\tau_k = \lambda_k = 1$ and

$$R = \begin{bmatrix} \frac{1}{r_1} I_1 & \mathbf{0} \\ \mathbf{0} & \frac{1}{r_2} I_2 \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} \frac{1}{s_1} I_1 & \mathbf{0} \\ \mathbf{0} & \frac{1}{s_2} I_2 \end{bmatrix}$$

where $r_i, s_i > 0, i = 1, 2$, $I_i (i = 1, 2)$ are identity matrices, then Algorithm 2 reduces to the following one, which is the PDL method in [25].

Remark 3.2. *It is easy to see that, if $r_i s_i < 1 (i = 1, 2)$ as required in [25], then $R - \bar{\tau}\bar{\lambda} S^{-1} \succ \mathbf{0}$ naturally holds. Thus, our method relaxes the requirement on primal-dual step sizes in [25].*

Algorithm 3 Primal-Dual Method with Correction Step-B

Initialization: $x^0 \in X, \bar{y}^0 \in Y, \tau_0, \lambda_0 > 0$.

Iteration:

$$y^{k+1} = \arg\max_{y \in Y} L(x^k, y) - \frac{1}{2}||y - \bar{y}^k||_S^2, \tag{3.22}$$

$$x^{k+1} = \arg\min_{x \in X} L(x, y^{k+1}) + \frac{1}{2}||A(x - x^k)||_R^2, \tag{3.23}$$

$$\bar{y}^{k+1} = \arg\max_{y \in Y} L(x^{k+1}, y) - \frac{1}{2}||y - \bar{y}^k||_S^2, \tag{3.24}$$

Until meet stopping criterion.

4 Numerical experiments

In this section, we study the numerical solution of the $TV - L^1$ model for image deblurring

$$\min_{x \in X} \quad F(x) = ||Kx - f||_1 + \mu||Dx||_1 \tag{4.1}$$

where $f \in Y$ is a given (noisy) image , $K : X \to Y$ is a known linear (blurring) operator, $D : X \to Y$ denotes the gradient operator and μ is a regularization parameter. Now we introduce the variables γ_1, γ_2, which satisfy $\gamma_1, \gamma_2 > 0$ and $\gamma_1 + \gamma_2 = \mu$. Then,(4.1) can be written as

$$\min_{x \in X} ||Kx - f||_1 + \gamma_1||Dx||_1 + \gamma_2||Dx||_1$$

Further, the above formula can be rewritten as [5]

$$\min_{x \in X} \max_{y \in Y} L(x, y) := \gamma_1||Dx||_1 + \langle Ax, y\rangle - \delta_{C_1}(u) - \delta_{C_{\gamma_2}}(v) - \langle f, u\rangle \tag{4.2}$$

where $C_\lambda = \{y \in Y | \, \|y\|_\infty \le \lambda\}$, $y = \begin{bmatrix} u \\ v \end{bmatrix}$, and $A = \begin{bmatrix} K \\ \gamma_2 D \end{bmatrix}$. Next we suppose that $\mathcal{N}(K) \bigcap \mathcal{N}(D) = \{0\}$, where $\mathcal{N}(Q)$ represents the null space of the matrix Q, and this assumption has been used in many references including similar types of problems, for example, in [24]. Under this assumption, A is of full column rank.

For simplicity we set $\tau_k = \lambda_k = 1$, $R = \begin{bmatrix} \frac{1}{r_1} I_1 & \mathbf{0} \\ \mathbf{0} & \frac{1}{r_2} I_2 \end{bmatrix}$ and $S = \begin{bmatrix} \frac{1}{s_1} I_1 & \mathbf{0} \\ \mathbf{0} & \frac{1}{s_2} I_2 \end{bmatrix}$ in Algorithm 1. Setting $\varepsilon_{k+1} = 0$, and we can compute y^{k+1} by the following formula:

$$u^{k+1} = P_{C_1}(\bar{u}^k + s_1(Kx^k - f)),$$
$$v^{k+1} = P_{C_{\gamma_2}}(\bar{v}^k + s_2\gamma_2 Dx^k),$$

where $y^{k+1} = (u^{k+1}, v^{k+1})^T$. Similarly, we can get $\bar{y}^{k+1} = (\overline{u}^{k+1}, \overline{v}^{k+1})$ where $\overline{u}^{k+1}$ and $\overline{v}^{k+1}$ can be obtained by replacing x^k with x^{k+1} in u^{k+1} and v^{k+1}, respectively.

Now we consider the computation of the following subproblem:

$$\begin{aligned} x^{k+1} &\approx_2^{\delta_{k+1}} \arg\min_{x \in X} L(x, y^{k+1}) + \frac{1}{2}\|A(x - x^k)\|_R^2 \\ &\approx_2^{\delta_{k+1}} \arg\min_{x \in X} \gamma_1\|Dx\|_1 + \langle Ax, y^{k+1}\rangle + \frac{1}{2}\|A(x - x^k)\|_R^2. \end{aligned} \tag{4.3}$$

The above formula can be equivalently rewritten as

$$x^{k+1} \approx_2^{\delta_{k+1}} \arg\min_{x\in X} \gamma_1\|Dx\|_1 + \frac{1}{2}\|Bx-\xi\|^2, \tag{4.4}$$

where $B = \begin{bmatrix} \frac{1}{\sqrt{r_1}}K \\ \frac{\gamma_2}{\sqrt{r_2}}D \end{bmatrix}$ and $\xi = \begin{bmatrix} \frac{1}{\sqrt{r_1}}Kx^k - \sqrt{r_1}u^{k+1} \\ \frac{\gamma_2}{\sqrt{r_2}}Dx^k - \sqrt{r_2}v^{k+1} \end{bmatrix}$. We note that $B^TB = \frac{1}{r_1}K^TK + \frac{\gamma_2^2}{r_2}D^TD$ is symmetrically positive definite because B is of full column rank. Therefore, there exists $z \in X$ such that $Bz = \xi$. Further, (4.3) can be rewritten equivalently as

$$x^{k+1} \approx_2^{\frac{\delta_{k+1}}{\gamma_1}} \arg\min_{x\in X} ||Dx||_1 + \frac{1}{2\gamma_1}||x-z||^2_{B^TB} \tag{4.5}$$

Next we will show that the subproblem (4.5) can be computed by approximately minimizing duality gap.

Setting $H(x) := h(Dx) = ||Dx||_1, \varphi(x) = \frac{1}{2\gamma_1}||x-z||^2_{B^TB}$, we consider the folowing primal problem:

$$\min_{x\in X} h(Dx) + \varphi(x) := ||Dx||_1 + \frac{1}{2\gamma_1}||x-z||^2_{B^TB}. \tag{4.6}$$

together with its dual problem:

$$\min_{v\in Y} \varphi^*(-D^Tv) + h^*(v) = \frac{1}{2\gamma_1}||\gamma_1 D^Tv - B^T\xi||^2_{(B^TB)^{-1}} - \frac{1}{2\gamma_1}||\xi||_2^2 + \delta_\Omega(v) \tag{4.7}$$

which $\Omega = \{v | ||v||_\infty \le 1\}$. If $\bar{v}$ is a solution of Problem (4.7), then

$$\bar{x} = (B^TB)^{-1}(B^T\xi - \gamma_1 D^T\bar{v}) \tag{4.8}$$

is a solution of Problem (4.6). Since h is positively homogeneous, from Remark 1 of [23] we get $h^*(v) = H^*(D^Tv)$. Now we consider the dual gap

$$\begin{aligned}\Psi(\bar{x},\bar{v}) &= h(D\bar{x}) + \varphi(\bar{x}) + \varphi^*(-D^T\bar{v}) + h^*(\bar{v}) \\ &= H(\bar{x}) + H^*(D^T\bar{v}) + \langle -\bar{x}, D^T\bar{v}\rangle \\ &= \sup_{v\in Y}\{\langle v,\bar{x}\rangle - H^*(v)\} - \langle \bar{x}, D^T\bar{v}\rangle + H^*(D^T\bar{v}) \\ &\ge H^*(D^T\bar{v}) - H^*(v) + \langle v - D^T\bar{v}, \bar{x}\rangle, \ \forall\, v \in Y. \end{aligned} \tag{4.9}$$

Thus, if $\Psi(\bar{x},\bar{v}) \le \delta$, where $\delta > 0$ is some given tolerance, then $\bar{x} \in \partial_\delta H^*(D^T\bar{v})$, which by Theorem 2.4.4 of [26] is equivalent to $D^T\bar{v} \in \partial_\delta H(\bar{x})$, and hence from Definition 2.3 this implies $\bar{x} \approx_2^\delta \mathrm{Prox}_{\gamma_1 H}^{B^TB}(z)$. Therefore,

$$\Psi(\bar{x},\bar{v}) \le \delta \Rightarrow \bar{x} \approx_2^\delta \mathrm{Prox}_{\gamma_1 H}^{B^TB}(z), \tag{4.10}$$

where $\bar{x}$ and $\bar{v}$ satisfy (4.8). Thus, we use FISTA method ([1]) to solve the dual problem (4.7) so as to better evaluate the gap. In view of (4.10), we adopt the following inequality as the stopping criterion of inner loop:

$$\Psi(x^{k+1}, v^{k+1}) \le \delta_{k+1}, \tag{4.11}$$

where $\delta_{k+1} = O(1/(k+1)^{\alpha+\frac{1}{2}}(\alpha > 0)$. In the following we will report the numerical experiment results.

The MATLAB codes are run on a PC (with CPU Intel i5-5200U) under MATLAB Version 8.5.0.197613 (R2015a) Service Pack 1. We report numerical results of the proposed methods. We test the images cameraman.png(256×256) and man.png(1024×1024), as presented in Figures 1 and 2. At the same time, we adopt the following stopping rule:

$$\frac{F(x^k) - F(x^*)}{F(x^*)} < 10^{-5},$$

where x^* is a solution of the $TV - L^1$ model (4.1).

Figure 1: Original cammeraman.png(256×256)

Figure 2: Original man.png(1024×1024)

4.1 Sensitivity of parameters

In this section, we will analyze the sensitivity of parameters. In this test, average blur with hsize=9 was applied to the original image cameraman.png(see Figure 3) (256×256) by fspecial(average,9), and 20% salt-pepper noise was added in. According to Theorem 3.2, the convergence rat e of Algorithm 1 depends on the value of parameter α. At the same time, from (4.11) we know that the iteration number of inner loop closely relates to the parameter α. Hence, we first study the sensitivity of α. In the following experiment, we take $\mu = 0.05$, $s_1 = 1$, $s_2 = 2$, $r_1 = \frac{0.99}{s_1}$, $r_2 = \frac{0.99}{s_2}$, $\gamma_1 = \gamma_2 = \frac{1}{2}\mu$. We choose the 256×256 cammeraman.png as the test picture and take $\alpha \in (0, 2.3)$. The iteration number of outer loop is fixed as 100. In Figure 5,the ordinate denotes the iteration number of the inner loop while the abscissa denotes the value of α. From Figure 5, we can see that, when α is not very large, the iteration number of inner loop is very little; However, as α increases, the iteration number of inner loop also increases rapidly. Similar results can also be found in [18].

Figure 3: Cammeraman.png with noise

Figure 4: Man.png with noise

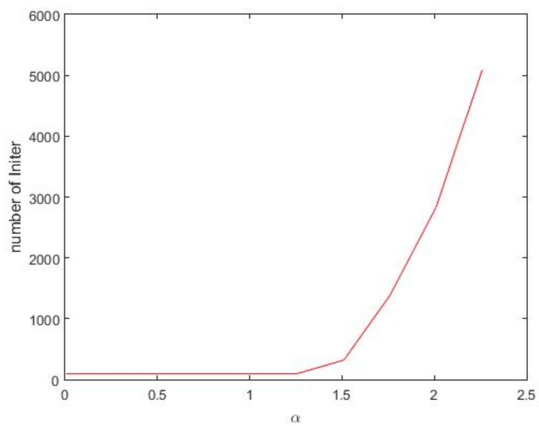

Figure 5: Sensitivity of α

Next we investigate the sensitivity of parameters s_1 and s_2. We still take $\mu = 0.05$, $r_1 = \frac{0.99}{s_1}$, $r_2 = \frac{0.99}{s_2}$, $\gamma_1 = \gamma_2 = \frac{1}{2}\mu$ and fix $\alpha = 1$. If s_2 is fixed as 2, we take $s_1 \in (0.8, 2.5)$; If s_1 is fixed as 1, we take $s_2 \in (0.8, 2.3)$. The iteration number of outer loop is fixed as 100. In Figures 6 and 7, the ordinate denotes the running time of Algorithm 1 while the abscissa denotes the value of s_1 or s_2. From Figures 6 and 7, we can see that, the running time decreases as s_1 or s_2 increases.

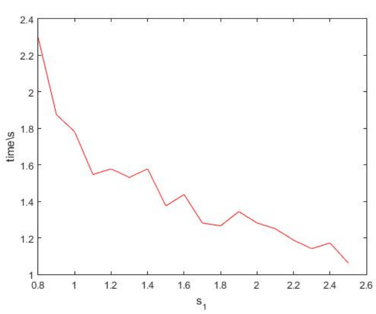

Figure 6: Sensitivity of s_1

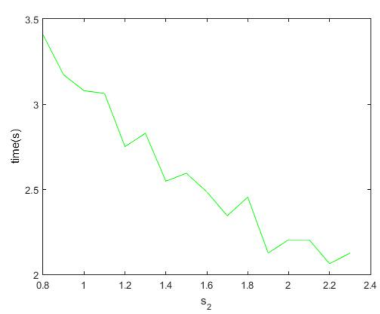

Figure 7: Sensitivity of s_2

Further, we consider the sensitivity of parameters γ_1 and γ_2 which satisfy $\gamma_1 + \gamma_2 = \mu$. We take $\alpha = 1$, $\mu = 0.05$ and fix the iteration number of outer loop as 200. In Figure 8, the ordinate denotes the value of $F(x^k) - F(x^*)$ over $F(x^*)$ while the abscissa denotes the value after the number of inner iterations is taken $\log_{10}$. In addition, "initr" denotes the number of total iterations of inner loop(i.e., the second subproblem). At the same time, we choose γ_1 as four different values $\frac{1}{2}\mu$, $\frac{1}{2.5}\mu$, $\frac{1}{3}\mu$ and $\frac{1}{3.5}\mu$ which correspond to four different curves in Figure 8, respectively. These curves indicate that, when the value of γ_1 or γ_2 varies, the number of inner iterations changes remarkably. Hence, the CPU time increases markedly as γ_1 decreases. By testing Figure 3, the CPU time corresponding to the above four choices of γ_1 is $95.8s, 149.7s, 210.9s$ and $279.2s$ respectively.

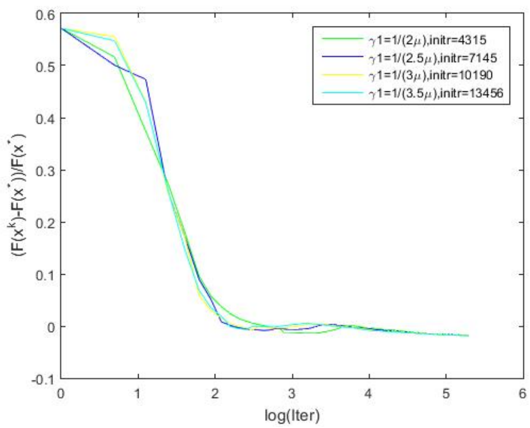

Figure 8: Sensitivity of γ

Finally, we analyze the variation of Algorithm 1's numerical performance with respect to various choices of α when γ_1 is fixed as $\frac{1}{3}\mu$. Besides, average blur with hsize=9 was applied to the original image man.png(see Figure 4) (1024×1024) by fspecial(average,9), and 20% salt-pepper noise was added in. In Table 1, CPU, "iter-out" and "iter-in" denote the CPU time in seconds,the iteraion number of outer and inner loops, respectively. Testing Figures 3 and 4 yields the following results:

Table 1

	Figure 3			Figure 4		
α	CPU	iter-out	iter-in	time(s)	iter-out	iter-in
0.1	0.8438	10	18	26.0900	10	18
0.3	1.0625	10	34	35.2031	10	32
0.5	1.2500	9	35	40.4688	10	41
0.8	1.5625	10	51	48.3594	10	50
1	1.8281	11	63	58.3906	11	63

From Table 1 we can see that, as α increases, the CPU time and the number of inner iterations increase while the number of outer iterations keeps invariant bascially.

4.2 Image denoising

In this section, we apply our Algorithm 1 to image debluring of $TV - L^1$ model (4.1). In the following tables and figures, "CP", iCP, "PDL" and "iPDL" denote Algorithm 1 in [4], Algorithm (4.2) in [18], Algorithm PDL in [25] and our Algorithm 1, respectively. In this experiment, we test Figures 3 and 4. We fixed the number of iterations as 200 and the penalty coefficient $\mu = 0.1$. When the above four Algorithms are implemented, their respective parameters are given in Table 2.

Table 2

CP	$\tau = \sigma = \frac{0.99}{\sqrt{8}}$
iCP	$\tau = \sigma = 0.99, \alpha = 1$
PDL	$s_1 = 2, r_1 = \frac{0.99}{s_1}, s_2 = 1, r2 = \frac{0.99}{s_2}$
$iPDL$	$s_1 = 2, r_1 = \frac{0.99}{s_1}, s_2 = 1, r2 = \frac{0.99}{s_2}, \alpha = 1, \gamma_1 = \gamma_2 = \frac{1}{2\mu}$

The restored images by the above four Algorithms are displayed in Figure 9. Obviously, our algorithm and CP algorithm get better restoration quality compared with the iCP and PDL methods. In our experiment, we find that, if we increase the number of iterations to 1000 or more, all four algorithms can restore the image with almost the same quality, but our algorithm need fewer iterations than other three ones.

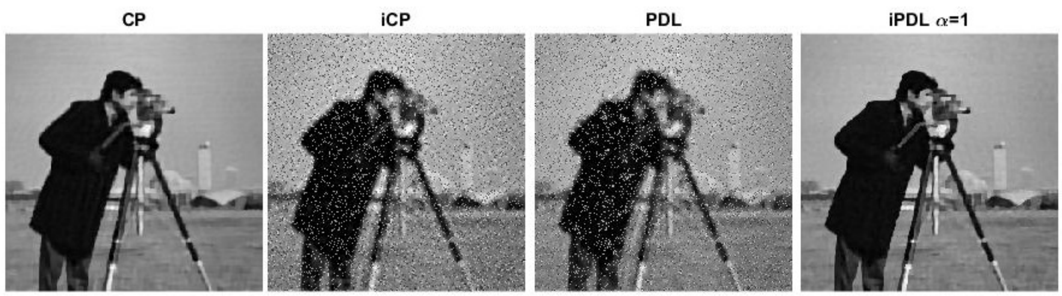

Figure 9: Restored images

5 Conclusions

In this paper, we propose an inexact primal-dual method for the saddle point problem by applying inexact extended proximal operators. We show the convergence of our Algorithm 1, provided that the partial sums A_N and B_N are summable. The $O(1/N)$ convergence rate in the ergodic sense is also established. We also apply our method to solve TV-L_1 image deblurring problems and verify their efficiency numerically.

It is worth mentioning that our method have some existing algorithms as special cases by the appropriate choices of parameters. Besides, our method also relaxes the requirement on primal-dual step sizes, for example, in [25]. At present, however, we are not able to provide the accelerated versions of our method , for example, under the assumption that f or g is strongly convex. Hence, this will be the subject of future research.

References

[1] A. Beck,and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imaging Sciences, 2(1)(2009), 183-202.

[2] S. Boyd , N. Parikh , E. Chu , B. Peleato, and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers[J], Foundations and Trends in Machine Learning, 3(2010), 1-122.

[3] J.-F. Cai, E.J. Cands, and Z. Shen, A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization, 20(4) (2010), 1956-1982.

[4] A. Chambolle, and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision, 40(1)(2011), 120-145.

[5] A. Chambolle, and T. Pock. An introduction to continuous optimization for imaging. Acta Numerica, 25(2016), 161-319.

[6] Patrick L. Combettes, and Noli N. Reyes, Moreau's decomposition in Banach spaces, Mathematical Programming(Ser. B), 139(2013), 103-114.

[7] J. Eckstein, and D. P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, Mathematical Programming, 55 (1992), 293-318.

[8] M. Ehrhardt, M.Betcke, Multicontrast MRI reconstruction with structure-guided total variation. SIAM Journal on Imaging Sciences, 9(3)(2016), 1084-1106.

[9] E. Esser, X. Zhang, and T. F. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM Journal on Imaging Sciences, 3(4)(2010), 1015-1046.

[10] J. M. Fadili, and G. Peyre, Total variation projection with frst order schemes. IEEE Trans. Image Process. 20(3)(2011), 657-669.

[11] B.S. He, M. Tao, and X. Yuan, Alternating direction method with Gaussian back substitution for separable convex programming, SIAM Journal on Optimization, 22(2)(2012), 313-340.

[12] B.S. He, X. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective, SIAM Journal on Imaging Sciences, 5(1)(2012), 119-149.

[13] B.S. He, Y.F. You, and X.M. Yuan, On the convergence of primal-dual hybrid gradient algorithm, SIAM Journal on Imaging Sciences, 7 (2014), 2526-2537.

[14] Y. Liu, Y. Xu and W. Yin, Acceleration of primal-dual methods by preconditioning and simple subproblem procedures, Journal of Scientific Computing, 86, 21 (2021). https://doi.org/10.1007/s10915-020-01371-1

[15] S. Ma, D. Goldfarb, L. Chen, Fixed point and Bregman iterative methods for matrix rank minimization. Mathematical Programming, 128(1)(2011), 321-353.

[16] M. K. Ng, F. Wang, and X. Yuan, Inexact alternating direction methods for image recovery, SIAM Journal on Scientific Computing, 33 (2011), 1643-1668.

[17] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, An algorithm for minimizing the Mumford-Shah functional, in Computer Vision, 2009 IEEE 12th International Conference on, IEEE, 2009, 1133-1140.

[18] J. Rasch J, and A. Chambolle, Inexact first-order primal-dual algorithms. Computational Optimization and Applications, 76(2020), 381-430.

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

[20] S. Salzo, and S. Villa. Inexact and accelerated proximal point algorithms[J]. Journal of Convex Analysis,19(4)(2012), 1167-1192.

[21] M.Schmidt, N.L. Roux, and F.R. Bach, Convergence rates of inexact proximal-gradient methods for convex optimization. In: Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems, vol. 24, pp. 1458-1466. Curran Associates Inc, Red Hook (2011)

[22] T. Valkonen, A primalCdual hybrid gradient method for nonlinear operators with applications to mri, Inverse Problems, 30 (2014), 055012.

[23] S. Villa, S. Salzo, L. Baldassarre, and A. Verri. Accelerated and inexact forward-backward algorithms.SIAM Journal on Optimization, 23(3)(2013), 1607-1633.

[24] Y. Wang, J. Yang, W. Yin, and Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction. SIAM Journal on Imaging Sciences, 1(3)(2008), 248-272.

[25] Z. Xie, A primal-dual method with linear mapping for a saddle point problem in image deblurring. J. Vis. Commun. Image R., 42(2017), 112-120.

[26] C. Zalinescu. Convex Analysis in General Vector Spaces. World Scientific, 2002

[27] M. Zhu, and T. Chan, An efficient primal-dual hybrid gradient algorithm for total variation image restoration, UCLA CAM Report, 34 (2008).