An inexact primal–dual method with correction step for a saddle point problem in image deblurring

Changjie Fang\textsuperscript{1,2}, Liliang Hu\textsuperscript{1,2} · Shenglan Chen\textsuperscript{1,2}

Received: 31 August 2021 / Accepted: 6 July 2022 / Published online: 22 July 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

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 definite matrix $D$. 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\textsubscript{1} image deblurring problems. Numerical simulation results illustrate the efficiency of our method.

Keywords 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:

\[
\begin{aligned}
\min_{x \in X} \max_{y \in Y} L(x, y) &= f(x) + \langle Ax, y \rangle - g(y) \\
\end{aligned}
\tag{1.1}
\]

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

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

\[
L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*), \forall x \in X, y \in Y
\tag{1.2}
\]
Now we consider the primal problem
\[
\min_{x \in X} f(x) + h(Ax) \tag{1.3}
\]
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.,
\[
0 \in \partial f(x^*) + A^*y^*, \quad 0 \in \partial h(Ax^*) - y^*,
\]
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 \(f\), \(g\), 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 \(\lambda\) and \(\tau\) are small. In [4], a primal–dual method with inertial step \(\theta \in [0, 1]\) was proposed (denoted 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 \(\tau \lambda < 1/\|A^T A\|\). In [12], some prediction–correction contraction methods were presented in which the convergence was guaranteed with relaxed step sizes satisfying \(\lambda \tau < 4/(1 + \theta)^2\|A^T A\|^2\) 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 \(\lambda_i\) and \(\tau_i\) 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 corresponding parameters are required to be square summable. [18] introduced four different 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 $D(\tau)$. Applying these notions, we propose 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 $\lambda_i$ 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 $R - \tau_k \lambda_k S^{-1} > 0$, 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 $\{\delta_k\}$ and $\{\epsilon_k\}$ are required to decrease like $O(1/k^{2\alpha + 1})$ for some $\alpha > 0$; see Theorem 3.2.

In the numerical experiments part, we investigate the applications of our method in TV-L1 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 Sect. 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 Sect. 2, we introduce the concepts of inexact extended proximal operators and present some auxiliary lemmas. In Sect. 3, we describe our method and prove the convergence of our method. At the same time, we also analyze the convergence rate. Numerical experiment results are reported in Sect. 4. Some conclusions are presented in Sect. 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 $D > 0$ and given $y \in \mathbb{R}^n$, denote

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

and define the extended proximal operator of $h$ as

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

where $\|x\|_D^2 = \langle x, Dx \rangle$ and $D^{-1}$ denotes the inverse of $D$. Because $D$ is symmetric positive definite, $\text{Prox}^D_{\tau h}(y)$ 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 $\text{Prox}^D_{\tau h}(y)$ with precision $\varepsilon$ if

$$z \approx_0^{\varepsilon} \text{Prox}^D_{\tau h}(y) \iff \|z - \bar{z}\|_D \leq \sqrt{2\tau \varepsilon}$$

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 \( \text{ProxD}_{\tau h}(y) \) with precision \( \varepsilon \) if
\[
z \approx^1_\varepsilon \text{ProxD}_{\tau h}(y) \iff 0 \in \partial_\varepsilon G_y(z)
\]

Definition 2.3 Let \( \varepsilon \geq 0 \). \( z \in X \) is said to be a type-2 approximation of the extended proximal point \( \text{ProxD}_{\tau h}(y) \) with precision \( \varepsilon \) if
\[
z \approx^2_\varepsilon \text{ProxD}_{\tau h}(y) \iff \frac{1}{\tau} D(y - z) \in \partial_\varepsilon h(z)
\]

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 corresponding definitions in [18].

According to the above definitions, 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 \( z \in \text{dom} h \) and \( z \approx^1_0 \arg \min_{x \in X} \{ h(x) + \frac{1}{2\tau} \| x - y \|_D^2 \} \).

Proof According to Definition 2.2 and the definition of \( \partial_\varepsilon G_y \), we have
\[
G_y(x) \geq G_y(z) - \varepsilon, \quad \forall x \in X.
\]
Setting \( x = \bar{x} \) in (2.3) and using (2.2), from (2.3) we have
\[
\| \bar{x} - z \|_D^2 \leq 2\tau \varepsilon + 2\tau [h(\bar{x}) - h(z) + \langle \frac{1}{\tau} D(\bar{x} - y), \bar{x} - z \rangle]
\]
which implies that \( z \in \text{dom} f \). According to the optimality condition of (2.2), we have
\[
h(x) \geq h(z) + \langle \frac{1}{\tau} D(y - z), x - z \rangle, \quad \forall x \in X.
\]
Setting \( x = z \) in (2.5) and substituting the resulting inequality into (2.4), we get
\[
\| \bar{x} - z \|_D \leq \sqrt{2\tau \varepsilon}.
\]
In view of Definite 2.1, we obtain the conclusion. \qed

Lemma 2.2 Suppose that \( z \approx^1_\varepsilon \arg \min_{x \in X} \{ h(x) + \frac{1}{2\tau} \| x - y \|_D^2 \} \), 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
\[
h(x) \geq h(z) + (x - z, \frac{1}{\tau} D(y - z - \frac{x - z}{2})) - \varepsilon, \quad \forall x \in X.
\]
Set \( r = \frac{x - z}{2} \). By the definition of the \( \varepsilon \)-subdifferential of \( h \), the conclusion holds. \qed

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

\[
G_\gamma(x) = h(x) + \frac{1}{2\tau} \|x - y\|_D^2
\geq h(z) + \left( \frac{1}{\tau} D(y - z), x - z \right) - \varepsilon + \frac{1}{2\tau} \|x - y\|_D^2
= h(z) + \frac{1}{2\tau} \|z - y\|_D^2 + \frac{1}{2\tau} \|x - z\|_D^2 - \varepsilon
= G_\gamma(z) - \varepsilon + \frac{1}{2\tau} \|x - z\|_D^2
\geq G_\gamma(z) - \varepsilon,
\]

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

\[
\|a - b\|_D^2 = \|a - c\|_D^2 + \|b - c\|_D^2 - 2\langle D(a - c), b - c\rangle.
\]  

(2.6)

Hence, 0 \in \partial_\varepsilon G_\gamma(z). By Definition 2.2, the conclusion holds.

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

Lemma 2.4

\[
\|\text{Prox}^D_{\tau h}(y_1) - \text{Prox}^D_{\tau h}(y_2)\|_D \leq \|y_1 - y_2\|_D.
\]

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

\[
h(x) \geq h(z_1) + \left( \frac{1}{\tau} D(y_1 - z_1), x - z_1 \right), \quad \forall x \in X.
\]  

(2.7)

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

\[
h(z_2) \geq h(z_1) + \left( \frac{1}{\tau} D(y_1 - z_1), z_2 - z_1 \right).
\]  

(2.8)

Similarly, we have

\[
h(z_1) \geq h(z_2) + \left( \frac{1}{\tau} D(y_2 - z_2), z_1 - z_2 \right).
\]  

(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.,

\[
\|\text{Prox}^D_{\tau h}(y_1) - \text{Prox}^D_{\tau h}(y_2)\|_D \leq \|y_1 - y_2\|_D.
\]

□

The following lemma is crucial in proving the convergence of Algorithm 1.
Lemma 2.5 Suppose that \( g : X \mapsto \tilde{R} \) is a convex function. For given \( z_0, u, v \in X \) and \( \tau, \varepsilon > 0 \). Let

\[
\begin{align*}
\forall z & \in X \quad \frac{1}{\tau} D((z_0 - D^{-1}u) - z_1) \in \partial_{\varepsilon} g(z_1). \\
\forall z & \in X \quad \frac{1}{\tau} D((z_0 - D^{-1}v) - z_2) \in \partial_{\varepsilon} g(z_2).
\end{align*}
\]

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

\[
\begin{align*}
\langle D((z_0 - D^{-1}u) - z_1) - r, z - z_1 \rangle & \geq \varepsilon, \quad \forall z \in X, \\
\langle D((z_0 - D^{-1}v) - z_2) - r, z - z_2 \rangle & \geq \varepsilon, \quad \forall z \in X.
\end{align*}
\]

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

\[
\begin{align*}
g(z) - g(z_1) & \geq \frac{1}{\tau} \langle D((z_0 - D^{-1}u) - z_1) - r, z - z_1 \rangle, \quad \forall z \in X, \\
g(z) - g(z_2) & \geq \frac{1}{\tau} \langle D((z_0 - D^{-1}v) - z_2) - r, z - z_2 \rangle, \quad \forall z \in X.
\end{align*}
\]

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

\[
\begin{align*}
\langle D((z_0 - D^{-1}v) - z_2), z_2 - z \rangle & \geq \tau (g(z_2) - g(z) - \varepsilon) + \langle Dr, z_2 - z \rangle, \forall z \in X.
\end{align*}
\]

From (2.14), we have

\[
\begin{align*}
\langle D((z_0 - D^{-1}u) - z_1), z_1 - z \rangle & \geq \tau (g(z_1) - g(z) - \varepsilon).
\end{align*}
\]

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

\[
\begin{align*}
\langle z_2 - z_1, D(z_1 - z_0 + D^{-1}u) \rangle & \geq \tau (g(z_1) - g(z_2) - \varepsilon).
\end{align*}
\]

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

\[
\begin{align*}
||z_1 - z_2||_D^2 & \leq (\sqrt{2\tau \varepsilon} + ||u - v||_{D^{-1}})||z_1 - z_2||_D + 2\tau \varepsilon.
\end{align*}
\]
where the equality follows from (2.6) and the inequality follows from (2.16). Hence, the inequality (2.20) is equivalent to the following one:

\[ t^2 \leq pt + q. \]

Thus,

\[ \left( t - \frac{p}{2} \right)^2 \leq \left( \frac{p}{2} \right)^2 + q. \]  

(2.22)

Since \( t := \| z_1 - z_2 \|_D \geq 0 \), from (2.22) we have

\[ 0 \leq t \leq \frac{p + \sqrt{p^2 + 4q}}{2}. \]  

(2.23)

Combining (2.21) with (2.23), we get

\[ 0 \leq \| z_1 - z_2 \|_D \]

\[ \leq \frac{1}{2} \left( \sqrt{2\tau \varepsilon + \| u - v \|_{D^{-1}}} + \sqrt{\| u - v \|^2_{D^{-1}} + 10\tau \varepsilon + 2\sqrt{2\tau \varepsilon} \| u - v \|_{D^{-1}}} \right). \]

(ii) Note that

\[ \frac{1}{2} \| z - z_0 \|^2_D - \frac{1}{2} \| z - z_2 \|^2_D \]

\[ = \frac{1}{2} \| z_2 \|^2_D - \frac{1}{2} \| z_0 \|^2_D - \langle z_2 - z_0, Dz_0 \rangle \]

\[ - \langle z - z_2, v \rangle + \langle z - z_2, D(z_2 - z_0 + D^{-1}v) \rangle \]

\[ \geq \frac{1}{2} \| z_2 \|^2_D - \frac{1}{2} \| z_0 \|^2_D - \langle z_2 - z_0, Dz_0 \rangle - \langle z_1 - z_2, v \rangle \]

\[ + \langle z_1 - z, v \rangle + \tau (g(z_2) - g(z) - \varepsilon) + \langle Dr, z_2 - z \rangle, \]

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

\[ \langle z_1 - z, v \rangle \leq \frac{1}{2} \| z - z_0 \|^2_D - \frac{1}{2} \| z - z_2 \|^2_D - \xi \]

\[ - \tau (g(z_2) - g(z) - \varepsilon) - \langle Dr, z_2 - z \rangle, \]  

(2.24)

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

Also,

\[ \xi = \frac{1}{2} \| z_2 \|^2_D - \frac{1}{2} \| z_0 \|^2_D - \langle z_2 - z_0, Dz_0 \rangle - \langle z_1 - z_2, v - u \rangle \]

\[ + \langle z_2 - z_1, D(z_0 - z_1) \rangle + \langle z_2 - z_1, D(D^{-1}u - z_0 + z_1) \rangle \]

\[ \geq \frac{1}{2} \| z_2 \|^2_D - \frac{1}{2} \| z_0 \|^2_D - \langle z_2 - z_0, Dz_0 \rangle - \langle z_1 - z_2, v - u \rangle \]

\[ + \langle z_2 - z_1, D(z_0 - z_1) \rangle + \tau (g(z_1) - g(z_2) - \varepsilon) \]

\[ = \frac{1}{2} \| z_2 - z_1 \|^2_D + \frac{1}{2} \| z_0 - z_1 \|^2_D \]

\[ - \langle z_1 - z_2, u - v \rangle + \tau (g(z_1) - g(z_2) - \varepsilon) \]
Remark 3.1 It is easy to see that, if the following Algorithm 3, which is the PDL method in [25],

\[
ri \text{ where } n \text{ naturally holds. Thus, our method relaxes the requirement on primal–dual step sizes in [25].}
\]

Algorithm 2: \( R \)

At the same time, we suppose that the matrices \( X \in \mathbb{R}^{m \times n} \), \( R \in \mathbb{R}^{m \times m} \) and \( S \in \mathbb{R}^{n \times n} \). Therefore, multiplying both sides of the above inequality by \( \frac{1}{\tau} \) yields

\[
g(z_1) - g(z) + < z_1 - z, \frac{1}{\tau} v > \leq \frac{1}{2\tau}(\|z - z_0\|^2_D - \|z - z_2\|^2_D) + \frac{1}{2}(\|D^{-1}(u - v)\|^2_D - \|D^{-1}(u - v)\|^2_D)
\]

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

Therefore, combining (2.24) with (2.25), we have

\[
< z_1 - z, v > \leq \frac{1}{2}(\|z - z_0\|^2_D - \|z - z_2\|^2_D) + \frac{1}{2}(\|D^{-1}(u - v)\|^2_D - \|D^{-1}(u - v)\|^2_D)
\]

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

\[
(Dp, q) \leq \frac{1}{2}\|p\|_D + \frac{1}{2}\|q\|_D.
\]

Combining (2.24) with (2.25), we have

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

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

3 Main results

In this section, we suppose that \( X = \mathbb{R}^m \), \( Y = \mathbb{R}^n \), \( A \in \mathbb{R}^{n \times m} \), \( R \in \mathbb{R}^{m \times m} \) and \( S \in \mathbb{R}^{n \times n} \). At the same time, we suppose that the matrices \( R \), \( S \) and \( A^TRA \) are symmetric positive definite.

Next we present the inexact primal–dual method (see Algorithm 1) for solving (1.1). Now we consider the two special cases of Algorithm 1.

Remark 3.1 It is easy to see that, if \( r_i s_i < 1(i = 1, 2) \) as required in [25], then \( R - \tau^{-1}S^{-1} > 0 \) naturally holds. Thus, our method relaxes the requirement on primal–dual step sizes in [25].
# Algorithm 1 Inexact Primal–Dual Method with Correction Step

**Initialization:** $x^0 \in X, y^0 \in Y, r_0, \lambda_0 > 0$.

**Iteration:**

\[ y^{k+1} \approx \frac{1}{2} \underset{y \in Y}{\text{arg max}} L(x^k, y) - \frac{1}{2r_k} \| y - y^k \|^2_S \]  \hspace{1cm} (3.1)

\[ x^{k+1} \approx \frac{1}{2} \underset{x \in X}{\text{arg min}} L(x, y^{k+1}) + \frac{1}{2\lambda_k} \| A(x - x^k) \|^2_R \]  \hspace{1cm} (3.2)

\[ \bar{y}^{k+1} \approx \frac{1}{2} \underset{y \in Y}{\text{arg max}} L(x^{k+1}, y) - \frac{1}{2r_k} \| y - \bar{y}^k \|^2_S \]  \hspace{1cm} (3.3)

Until meet stopping criterion.

# Algorithm 2 Primal–Dual Method with Correction Step-A

**Initialization:** $x^0 \in X, y^0 \in Y, r_0, \lambda_0 > 0$.

**Iteration:**

\[ y^{k+1} = \underset{y \in Y}{\text{arg max}} L(x^k, y) - \frac{1}{2r_k} \| y - y^k \|^2_S \]  \hspace{1cm} (3.4)

\[ x^{k+1} = \underset{x \in X}{\text{arg min}} L(x, y^{k+1}) + \frac{1}{2\lambda_k} \| A(x - x^k) \|^2_R \]  \hspace{1cm} (3.5)

\[ \bar{y}^{k+1} = \underset{y \in Y}{\text{arg max}} L(x^{k+1}, y) - \frac{1}{2r_k} \| y - \bar{y}^k \|^2_S \]  \hspace{1cm} (3.6)

Until meet stopping criterion.

# Algorithm 3 Primal–Dual Method with Correction Step-B

**Initialization:** $x^0 \in X, y^0 \in Y, r_0, \lambda_0 > 0$.

**Iteration:**

\[ y^{k+1} = \underset{y \in Y}{\text{arg max}} L(x^k, y) - \frac{1}{2} \| y - y^k \|^2_S \]  \hspace{1cm} (3.7)

\[ x^{k+1} = \underset{x \in X}{\text{arg min}} L(x, y^{k+1}) + \frac{1}{2} \| A(x - x^k) \|^2_R \]  \hspace{1cm} (3.8)

\[ \bar{y}^{k+1} = \underset{y \in Y}{\text{arg max}} L(x^{k+1}, y) - \frac{1}{2} \| y - \bar{y}^k \|^2_S \]  \hspace{1cm} (3.9)

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 \((y^{k+1}, x^{k+1}, \overline{y}^{k+1}) \in X \times Y \times X\) be obtained from Algorithm 1, then for any pair \((x, y) \in X \times Y\) we have

\[
L(x^{k+1}, y) - L(x^{k+1}, y^{k+1}) \leq \frac{1}{2\tau_k} (\|y - \overline{y}\|_S^2 - \|y - y^{k+1}\|_S^2) + \frac{\epsilon_k + 1}{\tau_k} \|y - y^{k+1}\|_S
\]

\[
+ \frac{\tau_k}{2} \|A(x^{k+1} - x^k)\|_{S^{-1}}^2 + 2\epsilon_k \delta_k - \frac{1}{2\tau_k} \|y^{k+1} - y^{k+1}\|_S^2.
\]

(3.10)

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

\[
y^{k+1} \approx \frac{\epsilon_k + 1}{2} \arg \min_{y \in Y} \left\{ g(y) + \frac{1}{2\tau_k} \|y - \overline{y} - \tau_k S^{-1} Ax^{k}\|_S^2 \right\}
\]

and

\[
\overline{y}^{k+1} \approx \frac{\epsilon_k + 1}{2} \arg \min_{y \in Y} \left\{ g(y) + \frac{1}{2\tau_k} \|y - \overline{y} - \tau_k S^{-1} Ax^{k+1}\|_S^2 \right\},
\]

respectively.

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

\[
g(y^{k+1}) - g(y) + \langle y^{k+1} - y, Ax^{k+1} \rangle \leq \frac{1}{2\tau_k} (\|y - \overline{y}\|_S^2 - \|y - y^{k+1}\|_S^2) + \frac{\epsilon_k}{2} \|A(x^{k+1} - x^k)\|_{S^{-1}}^2 + 2\epsilon_k \delta_k
\]

\[
+ \sqrt{\frac{\epsilon_k + 1}{\tau_k} \|y - \overline{y}\|_S - \frac{1}{2\tau_k} \|y^{k+1} - y^{k+1}\|_S^2}.
\]

Hence,

\[
L(x^{k+1}, y) - L(x^{k+1}, y^{k+1}) = g(y^{k+1}) - g(y) - \langle y^{k+1} - y, Ax^{k+1} \rangle
\]

\[
\leq \frac{1}{2\tau_k} (\|y - \overline{y}\|_S^2 - \|y - y^{k+1}\|_S^2) + \frac{\epsilon_k}{2} \|A(x^{k+1} - x^k)\|_{S^{-1}}^2 + 2\epsilon_k \delta_k
\]

\[
+ \sqrt{\frac{\epsilon_k + 1}{\tau_k} \|y - \overline{y}\|_S - \frac{1}{2\tau_k} \|y^{k+1} - y^{k+1}\|_S^2}.
\]

This completes the proof. \(\square\)

**Lemma 3.2** Let \((y^{k+1}, x^{k+1}, \overline{y}^{k+1})\) be obtained from Algorithm 1, then for any \(x \in X\) we have

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

(3.11)
Proof By Definition 2.3, the optimal condition of (3.2) yields
\[
\frac{1}{\lambda_k} A^T RA(x^k - x^{k+1}) \in \partial \delta_{k+1} L(x^{k+1}, y^{k+1}).
\]
In view of the definition of \(\varepsilon\)-subdifferential, we have
\[
L(x^{k+1}, y^{k+1}) - L(x, y^{k+1}) \leq \frac{1}{\lambda_k} \langle A^T RA(x^k - x^{k+1}), x^{k+1} - x \rangle + \delta_{k+1}. \tag{3.12}
\]
Setting \(a := x^k, b := x, c := x^{k+1}, D := A^T RA\) in (2.6), we get
\[
\langle A^T RA(x^k - x^{k+1}), x^{k+1} - x \rangle = -\frac{1}{2} \left[ \|x^k - x^{k+1}\|_{A^T RA}^2 + \|x - x^k\|_{A^T RA}^2 - \|x - x^k\|_{A^T RA}^2 \right]. \tag{3.13}
\]
Combining (3.12) with (3.13), we know that (3.11) holds. \(\square\)

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 nonnegative 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 \(N \geq 1\),
\[
\mu_N \leq \frac{1}{2} \sum_{n=1}^{N} \sigma_n + \left( T_N + \left( \frac{1}{2} \sum_{n=1}^{N} \sigma_n \right)^2 \right)^{\frac{1}{2}}.
\]
Set
\[
\hat{x}^N := \frac{1}{N} \sum_{k=0}^{N-1} x^{k+1} \quad \text{and} \quad \hat{y}^N := \frac{1}{N} \sum_{k=0}^{N-1} y^{k+1}. \tag{3.14}
\]

Lemma 3.4 Let the sequence \(\{(x^k, y^k, \hat{y}^k)^T\}\) be obtained by Algorithm 1 and \((\hat{x}^N, \hat{y}^N)\) defined by (3.14). Suppose that \(\{\tau_k\}\) and \(\{\lambda_k\}\) are nondecreasing and \(R - \tau_k \lambda_k S^{-1} > 0\). Then for every saddle point \((x^*, y^*) \in X \times Y\) of (1.1), we have
\[
L(\hat{x}^N, y^*) - L(x^*, \hat{y}^N) \leq \frac{1}{2N\tau_N} \left[ \frac{\tau_N}{\tau_0} \|y^* - \hat{y}^0\|_S \right. \nonumber \\
+ \sqrt{\frac{\tau_N}{\lambda_0} \|x^* - x^0\|_{A^T RA} + 2A_N + \sqrt{2B_N}^2}, \tag{3.15}
\]
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 (3.10) and (3.11) yields
\[
L(x^{k+1}, y) - L(x, y^{k+1}) \leq \frac{1}{2\tau_k} (\|y - \hat{y}^k\|_S^2 - \|y - \hat{y}^{k+1}\|_S^2)\]
Since \( \{\tau_k\} \) and \( \{\lambda_k\} \) are nondecreasing and \( R - \tau_k\lambda_k S^{-1} > 0 \),
\[
L(x^{k+1}, y) - L(x, y^{k+1}) \leq \frac{1}{2\tau_0} \| y - y^0 \|_S^2 - \frac{1}{2\tau N} \| y - y^N \|_S^2 \leq \frac{1}{2\lambda_0} \| x - x^0 \|_{ATRA}^2 + \frac{1}{2\lambda_k} \| x - x^{k+1} \|_{ATRA}^2
\]
\[+ \sqrt{\frac{\varepsilon_k + 1}{\tau_k}} \| y - y^{k+1} \|_S \leq \frac{1}{2\lambda_k} \| x - x^{k+1} \|_{ATRA}^2 + \frac{1}{2\lambda_k} \| x - x^{k+1} \|_{ATRA}^2
\]
\[+ \sqrt{\frac{\varepsilon_k + 1}{\tau_k}} \| y - y^{k+1} \|_S + 2\varepsilon_k + 1 + \delta_{k+1}.
\]

(3.17)

Since \( L(x, y) \) and \(-L(x, y)\) are convex with respect to \( x \) and \( y \) respectively, using Jensen inequality and (3.17), we have
\[
N(L(\hat{X}^N, y) - L(x, \hat{Y}^N))
\]
\[
\leq \sum_{k=0}^{N-1} L(x^{k+1}, y) - L(x, y^{k+1})
\]
\[
\leq \frac{1}{2\tau_0} \| y - y^0 \|_S^2 - \frac{1}{2\tau N} \| y - y^N \|_S^2 + \frac{1}{2\lambda_0} \| x - x^0 \|_{ATRA}^2
\]
\[+ \sum_{k=0}^{N-1} \sqrt{\frac{\varepsilon_k + 1}{\tau_k}} \| y - y^{k+1} \|_S \leq \frac{1}{2\lambda_k} \| x - x^{k+1} \|_{ATRA}^2 + \sum_{k=0}^{N-1} \sqrt{\frac{\varepsilon_k + 1}{\tau_k}} \| y - y^{k+1} \|_S + \sum_{k=0}^{N-1} (2\varepsilon_k + 1 + \delta_{k+1}).
\]

(3.18)

Setting \( x := x^* \) and \( y := y^* \) in (3.18) and using (1.2) we have
\[
\| y^* - \hat{Y}^N \|_S \leq \frac{\tau N}{\tau_0} \| y^* - y^0 \|_S^2 + \frac{\tau N}{\lambda_0} \| x^* - x^0 \|_{ATRA}^2
\]
\[+ \sum_{k=0}^{N-1} \sqrt{\frac{\varepsilon_k + 1}{\tau_k}} \| y^* - y^{k+1} \|_S \leq 2\sum_{k=0}^{N-1} \sqrt{\frac{\varepsilon_k + 1}{\tau_k}} (2\varepsilon_k + 1 + \delta_{k+1}).
\]

Set \( \mu_N = \| y^* - \hat{Y}^N \|_S, T_N = \frac{\tau N}{\tau_0} \| y^* - y^0 \|_S^2 + \frac{\tau N}{\lambda_0} \| x^* - x^0 \|_{ATRA}^2 + 2B_N, \sigma_k = 2\tau N \sqrt{\frac{\varepsilon_k + 1}{\tau_k}} \)
in Lemma 3.3. Obviously, \( T_0 \geq \mu_0^2 \) and \( \sigma_k \geq 0 \). Thus,
\[
\| y^* - \hat{Y}^N \|_S \leq A_N + \left( \frac{\tau N}{\tau_0} \| y^* - y^0 \|_S^2 + \frac{\tau N}{\lambda_0} \| x^* - x^0 \|_{ATRA}^2 + 2B_N + A_N^2 \right)^{\frac{1}{2}}.
\]

Since \( A_N, B_N, \tau_k \) and \( \lambda_k \) are nondecreasing, we have for all \( k \leq N \),
\[
\| y^* - \hat{y}^k \|_S \leq A_k + \left( \frac{\tau_k}{\tau_0} \| y^* - y^0 \|_S^2 + \frac{\tau_k}{\lambda_0} \| x^* - x^0 \|_{ATRA}^2 + 2B_k + A_k^2 \right)^{\frac{1}{2}}
\]
\[\leq A_N + \left( \frac{\tau N}{\tau_0} \| y^* - y^0 \|_S^2 + \frac{\tau N}{\lambda_0} \| x^* - x^0 \|_{ATRA}^2 + 2B_N + A_N^2 \right)^{\frac{1}{2}}
\]
\[\leq 2A_N + \sqrt{\frac{\tau N}{\tau_0} \| y^* - y^0 \|_S + \sqrt{\frac{\tau N}{\lambda_0} \| x^* - x^0 \|_{ATRA} + \sqrt{2B_N}}.
\]

(3.19)
Hence, setting $x := x^*$ and $y := y^*$ in (3.18), and using (3.19) we have
\[
N(\mathcal{L}(\tilde{x}^N, y^*) - L(x^*, \tilde{y}^N)) \leq \frac{1}{2\tau_0} \|y^* - \tilde{y}^0\|^2_S + \frac{1}{2\lambda_0} \|x^* - x^0\|^2_{A^T RA} + \frac{1}{\tau N} B_N
\]
\[
+ \frac{1}{\tau N} A_N \left( 2A_N + \sqrt{\frac{\tau N}{\tau_0}} \|y^* - \tilde{y}^0\|_S + \sqrt{\frac{\lambda N}{\lambda_0}} \|x^* - x^0\|_{A^T RA} + \sqrt{2B_N} \right)
\]
\[
\leq \frac{1}{2\tau_0} \left( \sqrt{\frac{\tau N}{\tau_0}} \|y^* - \tilde{y}^0\|_S + \sqrt{\frac{\lambda N}{\lambda_0}} \|x^* - \tilde{y}^0\|_{A^T RA} + 2A_N + \sqrt{2B_N} \right)^2,
\]
which implies that (3.15) holds. This completes the proof. \hfill \Box

**Remark 3.2** If, in addition, $A_N$ and $B_N$ are summable and $\{\tau_k\}$ is bounded above, then from (3.15) we can establish the $O(1/N)$ convergence rate of our method in the ergodic sense.

**Theorem 3.1** Let $\{x^{k+1}, y^{k+1}, \tilde{y}^{k+1}\}$ be the sequence pair generated by Algorithm 1 and $\{\tilde{x}^N, \tilde{y}^N\}$ be defined by (3.14) in Lemma 3.4. Suppose that the assumptions of Lemma 3.4 hold and $\tau_k \leq \tau$, $\lambda_k \leq \lambda$ with $R - \tau \tau_S S^{-1} > 0$. If the partial sums $A_N$ and $B_N$ in Lemma 3.4 are summable and $A$ is of full column rank, then every cluster point $(\tilde{x}, \tilde{y})$ of $\{\tilde{x}^N, \tilde{y}^N\}$ is a saddle point of problem (1.1). Moreover, there exists a saddle point $(\tilde{x}, \tilde{y}) \in X \times Y$ such that $x^k \to \tilde{x}$ and $y^k \to \tilde{y}$ as $k \to \infty$.

**Proof** Since $A_N$ and $B_N$ are summable,
\[
\left( \sqrt{\frac{\tau N}{\tau_0}} \|y^* - \tilde{y}^0\|_S + \sqrt{\frac{\lambda N}{\lambda_0}} \|x^* - \tilde{y}^0\|_{A^T RA} + 2A_N + \sqrt{2B_N} \right)^2 \leq C_1 < +\infty.
\]
From (3.19), we know that for all $k \leq N$, $\|y^* - \tilde{y}^k\|_S \leq C_2 < +\infty$. By the same argumentation as for $\tilde{y}^k$, from (3.18) we obtain $\|x^* - x^k\|_{A^T RA} < \infty$ for all $k \leq N$ and hence $\{x^k\}$ is bounded, which implies the boundedness of $\{\tilde{x}^N\}$. Let $x := x^*$ and $y := y^*$ in (3.16) and then sum the resulting inequality from $k = 0$ to $N - 1$ to obtain
\[
\frac{1}{2\tau} \sum_{k=0}^{N-1} \|x^{k+1} - x^k\|^2_{A^T (R - \tau \tau_S S^{-1} A)^{-1} S} + \sum_{k=0}^{N-1} \frac{1}{2\tau} \|\tilde{y}^k - y^{k+1}\|^2_S
\]
\[
\leq \frac{1}{2\tau_0} \|y^* - \tilde{y}^0\|^2_S - \frac{1}{2\tau_N} \|y^* - \tilde{y}^N\|^2_S + \frac{1}{2\lambda_0} \|x^* - \tilde{x}^0\|^2_{A^T RA} - \frac{1}{2\lambda N} \|x^* - x^N\|^2_{A^T RA} + \frac{C_2}{\tau N} A_N + \frac{1}{\tau N} B_N := C_3 < +\infty. \quad (3.20)
\]
Letting $N \to \infty$ in (3.20) and applying the equivalence of $\|\cdot\|_M$ and $\|\cdot\|_2$, where $M$ denotes the symmetric positive definite matrix, we have
\[
x^k - x^{k+1} \to 0 \quad \text{and} \quad \tilde{y}^k - y^{k+1} \to 0 \quad \text{as} \quad k \to \infty. \quad (3.21)
\]
Hence, $\{\tilde{y}^k - y^{k+1}\}$ is bounded. Thus,
\[
\|y^* - y^{k+1}\|_S \leq \|y^* - \tilde{y}^k\|_S + \|\tilde{y}^k - y^{k+1}\|_S < +\infty.
\]
i.e., \( \{y^k\} \) is bounded, and hence \( \{\gamma^N_k\} \) is also bounded. Hence there exists a subsequence \( (\tilde{x}^N_k, \gamma^N_k) \) converging to a cluster point \( (\tilde{x}, \gamma) \). Since \( f \) and \( g \) are l.s.c., from (3.15) we have

\[
L(\tilde{x}, y^*) - L(x^*, \gamma) = \lim_{i \to \infty} L(\tilde{x}^N_i, y^*) - L(x^*, \gamma^N_i) \leq \lim_{i \to \infty} \frac{C_1}{2\tau N_i} = 0,
\]

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

Since the sequence pair \( (x^k, y^k) \) is bounded, there exists a subsequence \( (x^{ki}, y^{ki}) \) converging to a cluster point \( (\tilde{x}, \gamma) \). Since \( (\tilde{x}, \gamma) \) is a saddle point of \( L(x, y) \) by replacing \( (x^*, y^*) \) with \( (\tilde{x}, \gamma) \) in (3.16), we know that (3.21) holds. Hence, \( x^{ki} - x^{ki} \to 0 \) and \( y^{ki} - y^{ki} \to 0 \) as \( i \to \infty \). Hence,

\[
\|x^{ki} - \tilde{x}\|_{\tau RA} \leq \|x^{ki} - x^{ki}\|_{\tau RA} + \|x^{ki} - \tilde{x}\|_{\tau RA} \to 0 \quad \text{as} \quad i \to \infty,
\]

i.e., \( x^{ki} \to \tilde{x} \). Let now \( x^{k+1} = H(x^k) \) denote (3.5) in Algorithm 2 and \( x^{k+1} = H_{b_k+1}(x^k) \) denote (3.2). In view of the continuity of \( H \), we have

\[
\|\tilde{x} - H(\tilde{x})\|_{\tau RA} = \lim_{i \to \infty} \|x^{ki} - H(x^{ki})\|_{\tau RA} \leq \lim_{i \to \infty} (\|x^{ki} - H_{b_k}(x^{ki})\|_{\tau RA} + \|H_{b_k}(x^{ki}) - H(x^{ki})\|_{\tau RA}) \leq \lim_{i \to \infty} (\|x^{ki} - x^{ki}\|_{\tau RA} + \sqrt{2\delta_k}) = 0,
\]

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

By (3.21) and using the equivalence of the norms \( \|\cdot\|_2 \) and \( \|\cdot\|_S \) we have

\[
\lim_{i \to \infty} \|y^{ki} - y^{ki}\|_S = 0. \tag{3.22}
\]

In view of Lemma 2.5(i), we have

\[
0 \leq \|y^k - \tilde{y}^k\|_S \leq \frac{1}{2} \left( \sqrt{2\tau \varepsilon_k} + \|x^k - x^{k-1}\|_{\tau S^{-1}A} + \sqrt{\|x^k - x^{k-1}\|_{\tau S^{-1}A}^2 + 10\tau \varepsilon_k + 2\sqrt{2\tau \varepsilon_k} \|x^k - x^{k-1}\|_{\tau S^{-1}A}^2} \right). \tag{3.23}
\]

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

\[
\lim_{k \to \infty} \|y^{ki} - y^{ki}\|_S = 0. \tag{3.24}
\]

Let \( y^{k+1} = \Gamma(y^k) \) and \( \tilde{y}^{k+1} = \Psi(y^k) \) in Algorithm 2, and \( y^{k+1} = \Gamma_{b_k+1}(y^k) \) and \( \tilde{y}^{k+1} = \Psi_{b_k+1}(y^k) \) in Algorithm 1. Thus,

\[
\|\tilde{y} - \Gamma \circ \Psi(y)\|_S = \lim_{i \to \infty} \|y^{ki} - \Gamma \circ \Psi(y^{ki})\|_S \leq \lim_{i \to \infty} (\|y^{ki} - \Gamma \circ \Psi_{b_k}(y^{ki})\|_S + \|\Gamma \circ \Psi_{b_k}(y^{ki}) - \Gamma \circ \Psi(y^{ki})\|_S) \leq \lim_{i \to \infty} (\|\Gamma_{b_k}(y^{ki}) - \Gamma(y^{ki})\|_S + \|\Gamma(y^{ki}) - \Gamma(y^{ki})\|_S) \leq \lim_{i \to \infty} (\|\Gamma_{b_k}(y^{ki}) - \Gamma(y^{ki})\|_S + \sqrt{2\tau \varepsilon_k}) \leq \lim_{i \to \infty} (\sqrt{2\tau \varepsilon_k} + \|y^{ki} - y^{ki}\|_S) = \lim_{i \to \infty} \|y^{ki} - y^{ki}\|_S.
\]
\[
\begin{align*}
&\leq \lim_{i \to \infty} (\| y^{k_i-1} - y^{k_i} \|_S + \| \tilde{y}^{k_i} - y^{k_i} \|_S) \\
&= \lim_{i \to \infty} \| \tilde{y}^{k_i-1} - y^{k_i} \|_S + \lim_{i \to \infty} \| \tilde{y}^{k_i} - y^{k_i} \|_S = 0, \\
\end{align*}
\]

where the last equality follows from (3.22) and (3.24).

Hence, from (3.25) we obtain \( \tilde{y} = \Gamma \circ \Psi(\tilde{y}) \). Since \( \tilde{x} = H(\tilde{x}) \), it follows that \((\tilde{x}, \tilde{y})\) is a fixed point of Algorithm 2 and hence a saddle point of problem (1.1). We now use \((\tilde{x}, \tilde{y})\) in (3.17) and sum from \(k = k_i, \ldots, N - 1\) to obtain
\[
\frac{1}{2\lambda} \| \tilde{x} - x^N \|_{AT RA}^2 + \frac{1}{2\tau} \| \tilde{y} - \tilde{y}^N \|_S^2
\leq \frac{1}{2\tau_1} \| \tilde{y} - \tilde{y}^{k_i} \|_S^2 + \frac{1}{2\lambda_1} \| \tilde{x} - x^{k_i} \|_{AT RA}^2
+ \sum_{k=k_i}^{N-1} \sqrt{\frac{\epsilon_{k+1}}{\tau_1}} \| \tilde{y} - \tilde{y}^{k+1} \|_S + \sum_{k=k_i}^{N-1} (2\epsilon_{k+1} + \delta_{k+1}).
\]

Since \( \epsilon_k \to 0 \) and \( \delta_k \to 0 \) as \( k \to \infty \), the right hand size in (3.26) tends to zero for \( i \to \infty \), which implies that also \( x^N \to \tilde{x} \) and \( \tilde{y}^N \to \tilde{y} \) for \( N \to \infty \). Since \( x^N \to \tilde{x} \) as \( N \to \infty \), it is easy to see that \( \lim_{N \to \infty} \| x^N - \tilde{x}^{N-1} \|_{AT RA} = 0 \). Taking \( k = N \) in (3.23) we have \( \| y^N - \tilde{y}^N \|_S \to 0 \) (\( N \to \infty \)). Therefore,
\[
\lim_{N \to \infty} \| y^N - \tilde{y}^N \|_S \leq \lim_{N \to \infty} (\| y^N - \tilde{y}^N \|_S + \| \tilde{y}^N - \tilde{y}^N \|_S) = 0.
\]

Thus, \( \| y^N - \tilde{y}^N \|_S \to 0 \) as \( N \to \infty \) and hence \( y^N \to \tilde{y} \) as \( N \to \infty \). This completes the proof.

Next we will establish convergence rates of our method, provided that \( \{\delta_k\} \) and \( \{\epsilon_k\} \) decrease like \( O(\sqrt{\frac{1}{k^{3/2}}}) \). We first review the following lemma.

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

Similar to Corollary 1 in [18] and Theorem 4 in [20], we have the following theorem.

**Theorem 3.2** Keep the assumptions stated in Theorem 3.1. If \( \alpha > 0 \) and \( \delta_k = O(\sqrt{\frac{1}{k^{3/2}}}) \), \( \epsilon_k = O(\sqrt{\frac{1}{k^{3/2}}}) \), then for a saddle point \((x^*, y^*) \in X \times Y\) of (1.1),
\[
L(\tilde{x}^N, y^*) - L(x^*, \tilde{y}^N) = \begin{cases} 
O\left(\frac{1}{N}\right), & \alpha > \frac{1}{2}, \\
O\left(\frac{\ln^2 N}{N}\right), & \alpha = \frac{1}{2}, \\
O(N^{-2\alpha}), & \alpha \in (0, \frac{1}{2}).
\end{cases}
\]

**Proof** If \( \alpha > \frac{1}{2} \), then the sequences \( \{\epsilon_k\}, \{\sqrt{\delta_k}\} \) and \( \{\delta_k\} \) are summable and hence \( A_N := \sum_{k=0}^{N-1} \tau_N \sqrt{\frac{\epsilon_{k+1}}{\epsilon_k}} \) and \( B_N := \sum_{k=0}^{N-1} \tau_N(2\epsilon_{k+1} + \delta_{k+1}) \) in (3.15) are bounded. Since the sequence \( \{\tau_k\} \) is bounded, from (3.15) we have
\[
L(\tilde{x}^N, y^*) - L(x^*, \tilde{y}^N) = O\left(\frac{1}{N}\right).
\]
If $\alpha = \frac{1}{2}$, then $\delta_k = \mathcal{O}(\frac{1}{k^2})$ and $\varepsilon_k = \mathcal{O}(\frac{1}{k^2})$. Hence, $B_N$ is bounded. In view of the assumption on $\varepsilon_k$, we have
\[\sqrt{\varepsilon_{k+1}} \leq \frac{c}{k+1} \quad \text{for some } c > 0.\]
Thus,
\[
\sum_{k=0}^{N-1} \sqrt{\varepsilon_{k+1}} \leq \sum_{k=0}^{N-1} \frac{c}{k+1} = c \left( 1 + \sum_{k=2}^{N} \frac{1}{k} \right) \leq c \left( 1 + \int_{1}^{N} \frac{1}{t} dt \right) = c (1 + \ln N).
\]
Hence, from the boundedness of $\{\tau_k\}$ we know that $A_N = \mathcal{O}(\ln N)$. Therefore, from (3.15) we get
\[L(\hat{x}^N, y^*) - L(x^*, \hat{y}^N) = \mathcal{O}\left(\frac{\ln^2 N}{N}\right).\]

If $\alpha \in (0, \frac{1}{2})$, then $2\alpha + 1 > 1$ and hence by assumptions $\delta_k = \mathcal{O}(\frac{1}{k^{2\alpha+1}})$ and $\varepsilon_k = \mathcal{O}(\frac{1}{k^{2\alpha+1}})$ we deduce that $B_N$ is bounded. Since $\sqrt{\varepsilon_k} = \mathcal{O}(k^{-(\alpha + \frac{1}{2})})$ and $-\alpha - \frac{1}{2} > -1$, from Lemma 3.5 we obtain $A_N = \mathcal{O}(N^{1/2-\alpha})$. Thus, from (3.15) we have
\[L(\hat{x}^N, y^*) - L(x^*, \hat{y}^N) = \mathcal{O}(N^{-2\alpha}).\]

\[\square\]

4 Numerical experiments

In this section, we study the numerical solution of the TV-L1 model for image deblurring
\[
\min_{x \in X} F(x) = \|Kx - f\|_1 + \mu \|Dx\|_1,
\]
where $f \in Y$ is a given (noisy) image, $K : X \rightarrow Y$ is a known linear (blurring) operator, $D : X \rightarrow Y$ denotes the gradient operator and $\mu$ is a regularization parameter. Now we introduce the variables $y_1, y_2$, which satisfy $y_1, y_2 > 0$ and $y_1 + y_2 = \mu$. Then, (4.1) can be written as
\[
\min_{x \in X} \|Kx - f\|_1 + y_1 \|Dx\|_1 + y_2 \|Dx\|_1.
\]

Further, the above formula can be rewritten as [5]
\[
\min_{x \in X} \max_{y \in Y} L(x, y) := y_1 \|Dx\|_1 + \langle Ax, y \rangle - \delta_{C_{1}}(u) - \delta_{C_{2}}(v) - \langle f, u \rangle,
\]
where $C_\lambda = \{ y \in Y \mid \|y\|_\infty \leq \lambda \}$, $y = \begin{bmatrix} u \\ v \end{bmatrix}$, and $A = \begin{bmatrix} K \\ y_2 D \end{bmatrix}$. Next we suppose that $\mathcal{N}(K) \cap \mathcal{N}(D) = \{0\}$, 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 $\tau_k = \lambda_k = 1, R = \begin{bmatrix} \frac{1}{\tau_1} I_1 & 0 \\ 0 & \frac{1}{\tau_2} I_2 \end{bmatrix}$ and $S = \begin{bmatrix} \frac{1}{s_1} I_1 & 0 \\ 0 & \frac{1}{s_2} I_2 \end{bmatrix}$.

Algorithm 1. Setting $\varepsilon_{k+1} = 0$, and we can compute $y^{k+1}$ by the following formula:
\[u^{k+1} = P_{C_1}(\bar{u}^{k} + s_1 (Kx^k - f)),\]
$v^{k+1} = P_{C_{y^2}}(v^k + s_2 y D x^k),$

where $y^{k+1} = (u^{k+1}, v^{k+1})^T$. Similarly, we can get $\bar{v}^{k+1} = (u^{k+1}, \bar{v}^{k+1})$ where $u^{k+1}$ and $\bar{v}^{k+1}$ 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:

$$x^{k+1} \approx_{\frac{\delta}{2}} \min_{x \in X} L(x, y^{k+1}) + \frac{1}{2} \|A(x - x^k)\|_R^2$$

$$\approx_{\frac{\delta}{2}} \min_{x \in X} y_1 \|Dx\|_1 + \langle Ax, y^{k+1} \rangle + \frac{1}{2} \|A(x - x^k)\|_R^2.$$  \hspace{1cm} (4.3)

The above formula can be equivalently rewritten as

$$x^{k+1} \approx_{\frac{\delta}{2}} \min_{x \in X} y_1 \|Dx\|_1 + \frac{1}{2} \|Bx - \xi\|^2,$$  \hspace{1cm} (4.4)

where $B = \left[ \frac{1}{\sqrt{\gamma_1}} K \right]$ and $\xi = \left[ \frac{1}{\sqrt{\gamma_1}} K x^k - \sqrt{\gamma_1} u^{k+1} \right]$. We note that $B^T B = \frac{1}{\gamma_1} K^T K + \frac{\gamma_2}{\gamma_1} D^T D$ is symmetrically positive definite because $B$ is of full column rank. Therefore, there exists $z \in X$ such that $B z = \xi$. Further, (4.3) can be rewritten equivalently as

$$x^{k+1} \approx_{\frac{\delta}{2}} \min_{x \in X} y_1 \|Dx\|_1 + \frac{1}{2 y_1} \|x - z\|_{B^T B}^2.$$  \hspace{1cm} (4.5)

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

Setting $H(x) := h(Dx) = \|Dx\|_1, \varphi(x) = \frac{1}{2 y_1} \|x - z\|_{B^T B}^2$, we consider the following primal problem:

$$\min_{x \in X} h(Dx) + \varphi(x) := \|Dx\|_1 + \frac{1}{2 y_1} \|x - z\|_{B^T B}^2.$$  \hspace{1cm} (4.6)

together with its dual problem:

$$\min_{v \in Y} \varphi^*(-D^T v) + h^*(v) = \frac{1}{2 y_1} \|y_1 D^T v - B^T \xi\|_{(B^T B)^{-1}} - \frac{1}{2 y_1} \|\xi\|_2^2 + \delta_\Omega(v)$$  \hspace{1cm} (4.7)

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

$$\tilde{x} = (B^T B)^{-1}(B^T \bar{v} - y_1 D^T \bar{v})$$  \hspace{1cm} (4.8)

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

$$\psi(\bar{x}, \tilde{v}) = h(D\bar{x}) + \varphi(\bar{x}) + \varphi^*(-D^T \bar{v}) + h^*(\tilde{v})$$

$$= H(\bar{x}) + H^*(D^T \bar{v}) + \langle -\bar{x}, D^T \bar{v} \rangle$$

$$= \sup_{v \in Y} \{\langle v, \bar{x} \rangle - H^*(v)\} - \langle \bar{x}, D^T \bar{v} \rangle + H^*(D^T \bar{v})$$

$$\geq H^*(D^T \bar{v}) - H^*(v) + \langle v - D^T \bar{v}, \bar{x} \rangle, \forall v \in Y.$$  \hspace{1cm} (4.9)

Thus, if $\psi(\bar{x}, \tilde{v}) \leq \delta$, 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 $\bar{x} \approx_{\frac{\delta}{2}} \text{Prox}_{y_1 H}(z)$. Therefore,

$$\psi(\bar{x}, \tilde{v}) \leq \delta \Rightarrow \bar{x} \approx_{\frac{\delta}{2}} \text{Prox}_{y_1 H}(z),$$  \hspace{1cm} (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:

$$
\Psi(x^{k+1}, v^{k+1}) \leq \delta_{k+1},
$$

where $\delta_{k+1} = O(1/(k + 1)^{2\alpha + 1})(\alpha > 0)$. 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 Figs. 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).
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 Fig. 3) (256 × 256) by fspecial(average,9), and 20% salt-pepper noise was added in. According to Theorem 3.2, the convergence rate of Algorithm 1 depends on the value of parameter \( \alpha \). At the same time, from (4.11) we know that the iteration number of inner loop closely relates to the parameter \( \alpha \). Hence, we first study the sensitivity of \( \alpha \). 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 cameraman.png as the test picture and take \( \alpha \in (0, 2.3) \). The iteration number of outer loop is fixed as 100. In Fig. 5, the ordinate denotes the iteration number of the inner loop while the abscissa denotes the value of \( \alpha \). From Fig. 5, we can see that, when \( \alpha \) is not very large, the iteration number of inner loop is very little; However, as \( \alpha \) increases, the iteration number of inner loop also increases rapidly. Similar results can also be found in [18].

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 Figs. 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 Figs. 6 and 7, we can see that, the running time decreases as $s_1$ or $s_2$ increases.

Further, we consider the sensitivity of parameters $\gamma_1$ and $\gamma_2$ 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 Fig. 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 $\gamma_1$ as four different values $\frac{1}{2} \mu, \frac{1}{3} \mu, \frac{1}{4} \mu$ and $\frac{1}{5} \mu$ which correspond to four different curves in Fig. 8, respectively. These curves indicate that, when the value of $\gamma_1$ or $\gamma_2$ varies, the number of inner iterations changes remarkably. Hence, the CPU time increases markedly as
Fig. 7 Sensitivity of $s_2$

Fig. 8 Sensitivity of $\gamma$

$\gamma_1$ decreases. By testing Fig. 3, the CPU time corresponding to the above four choices of $\gamma_1$ is 95.8s, 149.7s, 210.9s and 279.2s, respectively.

Finally, we analyze the variation of Algorithm 1’s numerical performance with respect to various choices of $\alpha$ when $\gamma_1$ is fixed as $\frac{1}{3}\mu$. Besides, average blur with hsize=9 was applied to the original image man.png (see Fig. 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 iteration number of outer and inner loops, respectively. Testing Figs. 3 and 4 yields the following results:

From Table 1 we can see that, as $\alpha$ increases, the CPU time and the number of inner iterations increase while the number of outer iterations keeps invariant basically.
Table 1 | Algorithm 1’s numerical results corresponding to various choices of $\alpha$

| $\alpha$ | 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 |

Table 2 | The parameters corresponding to four algorithms

| Algorithm | Parameters |
|-----------|------------|
| CP        | $\tau = \sigma = \frac{0.99}{\sqrt{8}}$ |
| iCP       | $\tau = \sigma = 0.99, \alpha = 1$ |
| PDL       | $s_1 = 2, r_1 = \frac{0.99}{s_1}, s_2 = 1, r_2 = \frac{0.99}{s_2}$ |
| iPDL      | $s_1 = 2, r_1 = \frac{0.99}{s_1}, s_2 = 1, r_2 = \frac{0.99}{s_2}, \alpha = 1, \gamma_1 = \gamma_2 = \frac{1}{2\mu}$ |

Fig. 9 | Restored images

4.2 Image denoising

In this section, we apply our Algorithm 1 to image deblurring 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 Figs. 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.

The restored images by the above four Algorithms are displayed in Fig. 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 needs fewer iterations than other three ones.

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-L1 image deblurring problems and verify their efficiency numerically.

It is worth mentioning that our method has 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. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imaging Sci. 2(1), 183–202 (2009)
2. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3, 1–122 (2010)
3. Cai, J.-F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim. 20(4), 1956–1982 (2010)
4. Chambolle, A., Pock, T.: A first-order primal–dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. 40(1), 120–145 (2011)
5. Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. Acta Numer. 25, 161–319 (2016)
6. Combettes, Patrick L., Reyes, Noli N.: Moreau’s decomposition in Banach spaces. Math. Program. (Ser. B) 139, 103–114 (2013)
7. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. 55, 293–318 (1992)
8. Ehrhardt, M., Becteke, M.: Multicontrast MRI reconstruction with structure-guided total variation. SIAM J. Imaging Sci. 9(3), 1084–1106 (2016)
9. Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal–dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. 3(4), 1015–1046 (2010)
10. Fadili, J.M., Peyre, G.: Total variation projection with first order schemes. IEEE Trans. Image Process. 20(3), 657–669 (2011)
11. He, B.S., Tao, M., Yuan, X.: Alternating direction method with Gaussian back substitution for separable convex programming. SIAM J. Optim. 22(2), 313–340 (2012)
12. He, B.S., Yuan, X.: Convergence analysis of primal–dual algorithms for a saddle-point problem: from contraction perspective. SIAM J. Imaging Sci. 5(1), 119–149 (2012)
13. He, B.S., You, Y.F., Yuan, X.M.: On the convergence of primal–dual hybrid gradient algorithm. SIAM J. Imaging Sci. 7, 2526–2537 (2014)
14. Liu, Y., Xu, Y., Yin, W.: Acceleration of primal–dual methods by preconditioning and simple subproblem procedures. J. Sci. Comput. 86, 21 (2021). https://doi.org/10.1007/s10915-020-01371-1
15. Ma, S., Goldfarb, D., Chen, L.: Fixed point and Bregman iterative methods for matrix rank minimization. Math. Program. 128(1), 321–353 (2011)
16. Ng, M.K., Wang, F., Yuan, X.: Inexact alternating direction methods for image recovery. SIAM J. Sci. Comput. 33, 1643–1668 (2011)
17. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the Mumford-Shah functional. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 1133–1140. IEEE (2009)
18. Rasch, J., Chambolle, A.: Inexact first-order primal–dual algorithms. Comput. Optim. Appl. 76, 381–430 (2020)
19. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D 60, 227–238 (1992)
20. Salzo, S., Villa, S.: Inexact and accelerated proximal point algorithms. J. Convex Anal. 19(4), 1167–1192 (2012)
21. Schmidt, M., Roux, N.L., Bach, F.R.: 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. Valkonen, T.: A primal–dual hybrid gradient method for nonlinear operators with applications to MRI. Inverse Probl. 30, 055012 (2014)
23. Villa, S., Salzo, S., Baldassarre, L., Verri, A.: Accelerated and inexact forward-backward algorithms. SIAM J. Optim. 23(3), 1607–1633 (2013)
24. Wang, Y., Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. SIAM J. Imaging Sci. 1(3), 248–272 (2008)
25. Xie, Z.: A primal-dual method with linear mapping for a saddle point problem in image deblurring. J. Vis. Commun. Image Represent. 42, 112–120 (2017)
26. Zalinescu, C.: Convex Analysis in General Vector Spaces. World Scientific, Singapore (2002)
27. Zhu, M., Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration, UCLA CAM Report, 34 (2008)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.