*Research Article*

# A Preconditioning Technique for First-Order Primal-Dual Splitting Method in Convex Optimization

**Meng Wen,[1] Jigen Peng,[1] Yuchao Tang,[2,3] Chuanxi Zhu,[2] and Shigang Yue[4]**

[1]*School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China*
[2]*Department of Mathematics, Nanchang University, Nanchang 330031, China*
[3]*School of Management, Nanchang University, Nanchang 330031, China*
[4]*School of Computer Science, University of Lincoln, Lincoln LN6 7TS, UK*

Correspondence should be addressed to Jigen Peng; jgpeng@mail.xjtu.edu.cn

We introduce a preconditioning technique for the first-order primal-dual splitting method. The primal-dual splitting method offers a very general framework for solving a large class of optimization problems arising in image processing. The key idea of the preconditioning technique is that the constant iterative parameters are updated self-adaptively in the iteration process. We also give a simple and easy way to choose the diagonal preconditioners while the convergence of the iterative algorithm is maintained. The efficiency of the proposed method is demonstrated on an image denoising problem. Numerical results show that the preconditioned iterative algorithm performs better than the original one.

## 1. Introduction

Many real-world application problems arising in signal and image processing [1–3], machine learning [4–6], and medical image reconstruction [7, 8] can be modeled as solving some convex optimization problems (maybe nonsmooth). In recent years, the optimization of the sum of two convex functions has received much attention, which takes the form of

$$\min_{x \in R^n} \quad f(x) + h(Bx), \tag{1}$$

where $f \in \Gamma_0(R^n)$, $h \in \Gamma_0(R^m)$, and $B : R^n \to R^m$ is a linear transformation matrix. Here $\Gamma_0(R^n)$ denotes the set of proper, lower semicontinuous convex functions from $R^n$ to $(-\infty, +\infty]$. The functions $f$ and $g$ in (1) usually denote the data error term and the regularization term, respectively.

Chambolle and Pock [9] proposed a general primal-dual method to solve problem (1). Under the assumption that the functions $f$ and $h$ have a closed-form solution of proximity operator, they proved the convergence of the proposed iterative algorithms. He and Yuan [10] have shown that the primal-dual method of Chambolle and Pock [9] is equivalent to the proximity point algorithm (PPA). Thus, the original

convergence analysis can be easily obtained by the well-known PPA method. In order to accelerate the primal-dual method [9], Pock and Chambolle introduced in [11] a preconditioning technique for the primal-dual iterative algorithm for which the constant iterative parameters were replaced by some precondition iteration matrices. Then, the convergence of the preconditioned iterative algorithm followed directly from the PPA method. They also gave a practical way to choose the precondition iteration matrix. As an application, Sidky et al. [12] applied the primal-dual method of Chambolle and Pock [9] to solve a variety of application problems arising in medical image reconstruction. The primal-dual method and the corresponding precondition method presented the performance very well. Some related works can also be found in [13, 14].

If the objective function in (1) has a differentiable term with Lipschitz continuous gradient, such as the least squares loss function, the primal-dual method introduced by Chambolle and Pock [9] did not use the gradient operator of the function. In order to solve such a more general problem, Combettes and Pesquet [15], Condat [1], and Vũ [16] introduced a primal-dual method. For instance, Condat [1] considered an optimization problem involving the sum of three

convex functions, with a smooth function with Lipschitz continuous gradient, a nonsmooth proximable function, and linear composite functions. The problem is presented below:

$$\min_{x \in R^n} \quad f(x) + g(x) + h(Bx), \tag{2}$$

where $f \in \Gamma_0(R^n)$, $g \in \Gamma_0(R^n)$, $h \in \Gamma_0(R^m)$, $f : R^n \to R$ is differentiable, and its gradient $\nabla f$ is Lipschitz continuous with Lipschitz constant $\beta > 0$; $B : R^n \to R^m$ is a linear transformation matrix. To solve problem (2), he proposed a primal-dual splitting method and proved the convergence of the new iterative algorithm in an infinite-dimensional Hilbert space based on the fixed point theory of nonexpansive mappings.

The purpose of this paper is to introduce a preconditioned primal-dual splitting method to solve problem (2). The advantage of the preconditioning technique is that the iterative parameters will be updated self-adaptively. Furthermore, we give a family of preconditioners which are restricted to diagonal matrices and guarantee the convergence of the algorithm. To illustrate the efficiency of the proposed method, we compare it with the original method on image denoising problem. In addition, Combettes et al. [17] recently proposed a variable metric primal-dual method; they also referred to the preconditioning technique. But our algorithm is different from the method proposed by Combettes et al. In our algorithm, we use two different metrics: one for the primal variable; one for the dual variable(s). The other difference is in [17], which is a smooth term in the dual problem. When it is zero, the convergence conditions are stronger than the ones presented here. What is more, we explain how to choose the variable metric.

The rest of the paper is organized as follows. In Section 2, we briefly review the primal-dual splitting method proposed by Condat [1]. In Section 3, we present the preconditioning technique and provide a practical way to choose the iterative parameters matrices. In Section 4, we make several experiments in image denoising problems. Finally, we make a brief conclusion on this paper.

## 2. A Primal-Dual Splitting Method of Condat [1]

In this section, we briefly review the primal-dual splitting method introduced by Condat [1]. First, we introduce some definitions and notations. Let $H$ be a real Hilbert space, with its inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|^{1/2}$. We denote by $\Gamma_0(H)$ the set of proper, lower semicontinuous, convex functions from $H$ to $(-\infty, +\infty]$. Let $f \in \Gamma_0(H)$; its Fenchel conjugate $f^* \in \Gamma_0(H)$ is defined by $f^*(u) = \sup_x \langle x, u \rangle - f(x)$ and its proximity operator by $\text{prox}_{\lambda f}(u) = \arg\min_x \{(1/2)\|x - u\|^2 + \lambda f(x)\}$; $\lambda > 0$ is a positive constant. We define the subdifferential of $f$ as the set-valued operator $\partial f(u) = \{v \in H : f(u') \ge f(u) + \langle v, u' - u \rangle, \ \forall u' \in H\}$. If $f$ is differential at $u$, then $\partial f(u) = \{\nabla f(u)\}$. Let $H_1$ and $H_2$ be two real Hilbert spaces and $B : H_1 \to H_2$ be a bounded linear operator with adjoint $B^*$ and induced norm

$$\|B\| = \sup_{\|x\| \le 1} \{\|Bx\|, \ x \in H_1\} < +\infty. \tag{3}$$

Condat [1] considered solving the following optimization problem:

$$\text{Find } \widehat{x} \in \arg\min_x \quad f(x) + g(x) + h(Bx), \tag{4}$$

where $f : H_1 \to R$ is convex, differentiable on $H_1$, and its gradient $\nabla f$ is $\beta$-Lipschitz continuous, for some $\beta > 0$; that is, $\|\nabla f(x) - \nabla f(y)\| \le \beta\|x - y\|, \forall x, y \in H_1$. $g \in \Gamma_0(H_1)$ and $h \in \Gamma_0(H_2)$, and their proximity operator have a closed-form solution. $B : H_1 \to H_2$ is a bounded linear operator with adjoint $B^*$. The dual formulation of the primal problem (4) is

$$\text{Find } \widehat{y} \in \arg\min_y \quad (f + g)^* (-B^* y) + h^*(y). \tag{5}$$

The corresponding saddle point of primal problem (4) and dual problem (5) is as follows:

$$\text{Find } (\widehat{x}, \widehat{y}) \in \arg\min_x \max_y \quad \{f(x) + g(x) + \langle Bx, y \rangle - h^*(y)\}. \tag{6}$$

The pair $(\widehat{x}, \widehat{y})$ can be found via the following monotone variational inclusion:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial g(\widehat{x}) + B^* \widehat{y} + \nabla f(\widehat{x}) \\ -B\widehat{x} + \partial h^*(\widehat{y}) \end{pmatrix}, \tag{7}$$

where $\partial g$ and $\partial h^*$ are the subdifferential of $g$ and $h^*$, respectively.

Condat [1] proposed the following iterative algorithm to solve problem (4).

*Algorithm 1* (primal-dual splitting method (PDS) for solving problem (4)). Choose the proximal parameters $\tau > 0$, $\sigma > 0$ and the relaxation parameters $\{\rho_n\}$. Give an initial value $(x_0, y_0) \in H_1 \times H_2$. For $n = 0, 1, 2, \ldots$, the iterative $\{x_n\}$ and $\{y_n\}$ are updated by

(1) $\widetilde{x}_{n+1} = \text{prox}_{\tau g}(x_n - \tau(\nabla f(x_n) + e_{f,n}) - \tau B^* y_n) + e_{g,n}$;

(2) $\widetilde{y}_{n+1} = \text{prox}_{\sigma h^*}(y_n + \sigma B(2\widetilde{x}_{n+1} - x_n)) + e_{h,n}$;

(3) $(x_{n+1}, y_{n+1}) = \rho_n(\widetilde{x}_{n+1}, \widetilde{y}_{n+1}) + (1 - \rho_n)(x_n, y_n)$,

where the error terms $e_{f,n} \in H_1$, $e_{g,n} \in H_1$, and $e_{h,n} \in H_2$ model the inexact computation of the operators $\nabla f$, $\text{prox}_{\tau g}$, and $\text{prox}_{\sigma h}$, respectively.

If some stopping criteria have been reached, then the algorithm stops.

The convergence of Algorithm 1 was ensured by the following theorem.

**Theorem 2** (see [1]). *Let $\tau > 0$, $\sigma > 0$, and the sequences $\{\rho_n\}$, $\{e_{f,n}\}$, $\{e_{g,n}\}$, and $\{e_{h,n}\}$ be the parameters of Algorithm 1. Let $\beta$ be the Lipschitz constant of $\nabla f$. Suppose that $\beta > 0$ and the following conditions hold:*

(i) $1/\tau - \sigma\|B\|^2 > \beta/2$;

(ii) $\{\rho_n\} \subset (0, \delta)$, *where* $\delta = 2 - (\beta/2)(1/\tau - \sigma\|B\|^2)^{-1}$;

(iii) $\sum \rho_n(\delta - \rho_n) = +\infty$;

(iv) $\sum \rho_n \|e_{f,n}\| < +\infty$, $\sum \rho_n \|e_{g,n}\| < +\infty$, *and* $\sum \rho_n \|e_{h,n}\| < +\infty$. *Then the sequences $\{x_n\}$ and $\{y_n\}$ generated by Algorithm 1 converge weakly to $\widehat{x}$ of (4) and $\widehat{y}$ of (5), respectively.*
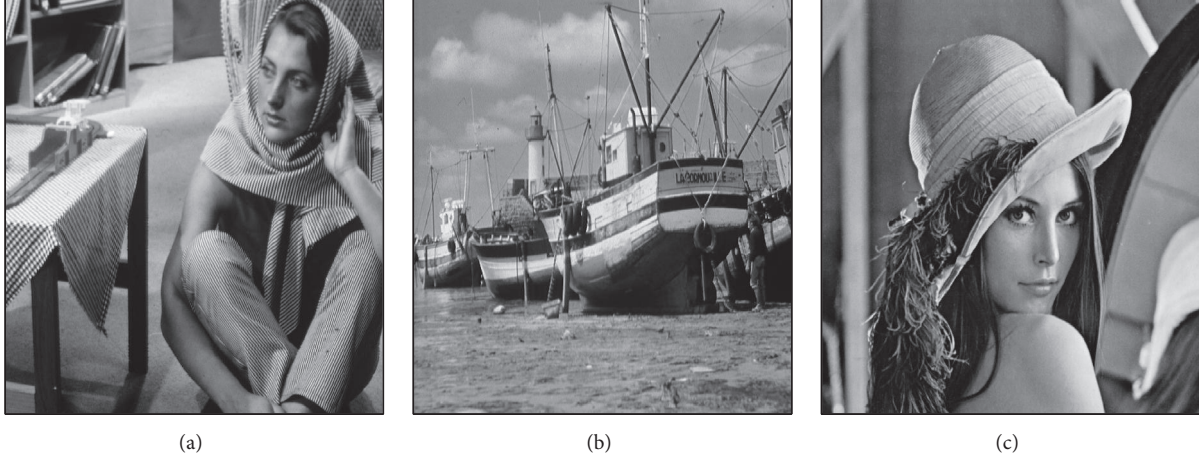
FIGURE 1: Original images. (a) "Barbara"; (b) "Boat"; (c) "Lena."

*Remark 3.* (1) Due to Moreau's identity $\text{prox}_{\sigma h^*}(u) = u - \sigma\text{prox}_{(1/\sigma)h}((1/\sigma)u)$, the proximity operator $\text{prox}_{\sigma h^*}$ can be computed from $\text{prox}_{(1/\sigma)h}$.

(2) The role of primal variable $x$ and dual variable $y$ in Algorithm 1 can be exchanged. The convergence of the new iterative algorithm can also be ensured, according to the same Theorem 2. In practice, the performance of the two iterative algorithms is nearly the same.

## 3. A Preconditioned Primal-Dual Splitting Method for Solving (4)

In this section, we give a precondition version of Algorithm 1. The main idea is motivated by the work of Pock and Chambolle [11]. The iterative parameters in Algorithm 1 are replaced by some symmetrical positive matrices. First, we give the detailed iterative algorithm below. Then we will prove its convergence.

*Algorithm 4* (preconditioned primal-dual splitting method (PPDS) for solving problem (4)). Choose the symmetrical positive define matrices $T$, $\Sigma$ and the relaxation parameters $\{\rho_n\}$. Give an initial value $(x_0, y_0) \in H_1 \times H_2$. For $n = 0, 1, 2, \ldots$, the iterative $\{x_n\}$ and $\{y_n\}$ are updated by

(1) $\tilde{x}_{n+1} = \text{prox}_{Tg}(x_n - T(\nabla f(x_n) + e_{f,n}) - TB^* y_n) + e_{g,n}$;

(2) $\tilde{y}_{n+1} = \text{prox}_{\Sigma h^*}(y_n + \Sigma B(2\tilde{x}_{n+1} - x_n)) + e_{h,n}$;

(3) $(x_{n+1}, y_{n+1}) = \rho_n(\tilde{x}_{n+1}, \tilde{y}_{n+1}) + (1 - \rho_n)(x_n, y_n)$,

where the error terms $e_{f,n} \in H_1$, $e_{g,n} \in H_1$, and $e_{h,n} \in H_2$ model the inexact computation of the operators $\nabla f$, $\text{prox}_{Tg}$, and $\text{prox}_{\Sigma h}$, respectively.

If some stopping criteria have been reached, then the algorithm stops.

**Theorem 5.** *Let $E$ be a symmetrical positive matrix and the sequences $\{\rho_n\}$, $\{e_{f,n}\}$, $\{e_{g,n}\}$, and $\{e_{h,n}\}$ be the parameters of Algorithm 4. Let $\beta$ be the Lipschitz constant of $\nabla f$. Suppose that $\beta > 0$ and the following conditions hold:*

(i) $\|T^{-1}\| - \|\Sigma\|\|B\|^2 > \|E\|/2$, $\|E\| \geq \beta$.

(ii) $\{\rho_n\} \subset (0, \delta)$, *where* $\delta = 2 - (\|E\|/2)(\|T^{-1}\| - \|\Sigma\|\|B\|^2)^{-1}$.

(iii) $\sum \rho_n(\delta - \rho_n) = +\infty$.

(iv) $\sum \rho_n\|e_{f,n}\| < +\infty$, $\sum \rho_n\|e_{g,n}\| < +\infty$, *and* $\sum \rho_n\|e_{h,n}\| < +\infty$. *Then the sequences $\{x_n\}$ and $\{y_n\}$ generated by Algorithm 4 converge weakly to $\hat{x}$ of (4) and $\hat{y}$ of (5), respectively.*

*Proof.* The main idea of the proof is based on the method of Theorem 2. First, we give some definitions and notations. Let $Z = H_1 \times H_2$ with inner product $\langle \cdot, \cdot \rangle_I$:

$$\langle z, z' \rangle_I = \langle x, x' \rangle + \langle y, y' \rangle,$$
$$\forall z = (x, y), \ z' = (x', y') \in Z. \quad (8)$$

Then, $Z$ is a real Hilbert space with the defined inner product.

Define $z_n = (x_n, y_n)$. It follows from the proof of Theorem 2; we know that the iterative sequences (1)–(3) in Algorithm 4 could be rewritten as follows:

$$z_{n+1} = \rho_n\left(\left(I + P^{-1} \circ A\right)^{-1}\left(z_n - P^{-1} \circ B(z_n) - e_{2,n}\right) + e_{1,n}\right) + (1 - \rho_n)z_n, \quad (9)$$

where $P = \begin{pmatrix} T^{-1} & -B^* \\ -B & \Sigma^{-1} \end{pmatrix}$, $A = \begin{pmatrix} \partial g & B^* \\ -B & \partial H^* \end{pmatrix}$, $B = \begin{pmatrix} \nabla f \\ 0 \end{pmatrix}$, $e_{1,n} = (e_{g,n}, e_{h,n})$, and $e_{2,n} = P^{-1}(e_{f,n}, -2Be_{g,n})$.

Notice that, from condition (i), it is easy to check that $P$ is bounded, self-adjoint, and strictly positive; that is, $\langle z, Pz \rangle_I > 0$, for every $z \neq 0$. Hence, we can define another inner product $\langle \cdot, \cdot \rangle_P$ and $\|\cdot\|_P = \langle \cdot, \cdot \rangle_P^{1/2}$ in $Z$ as

$$\langle z, z' \rangle_P = \langle z, Pz' \rangle_I, \quad \text{for every } z, z' \in Z. \quad (10)$$
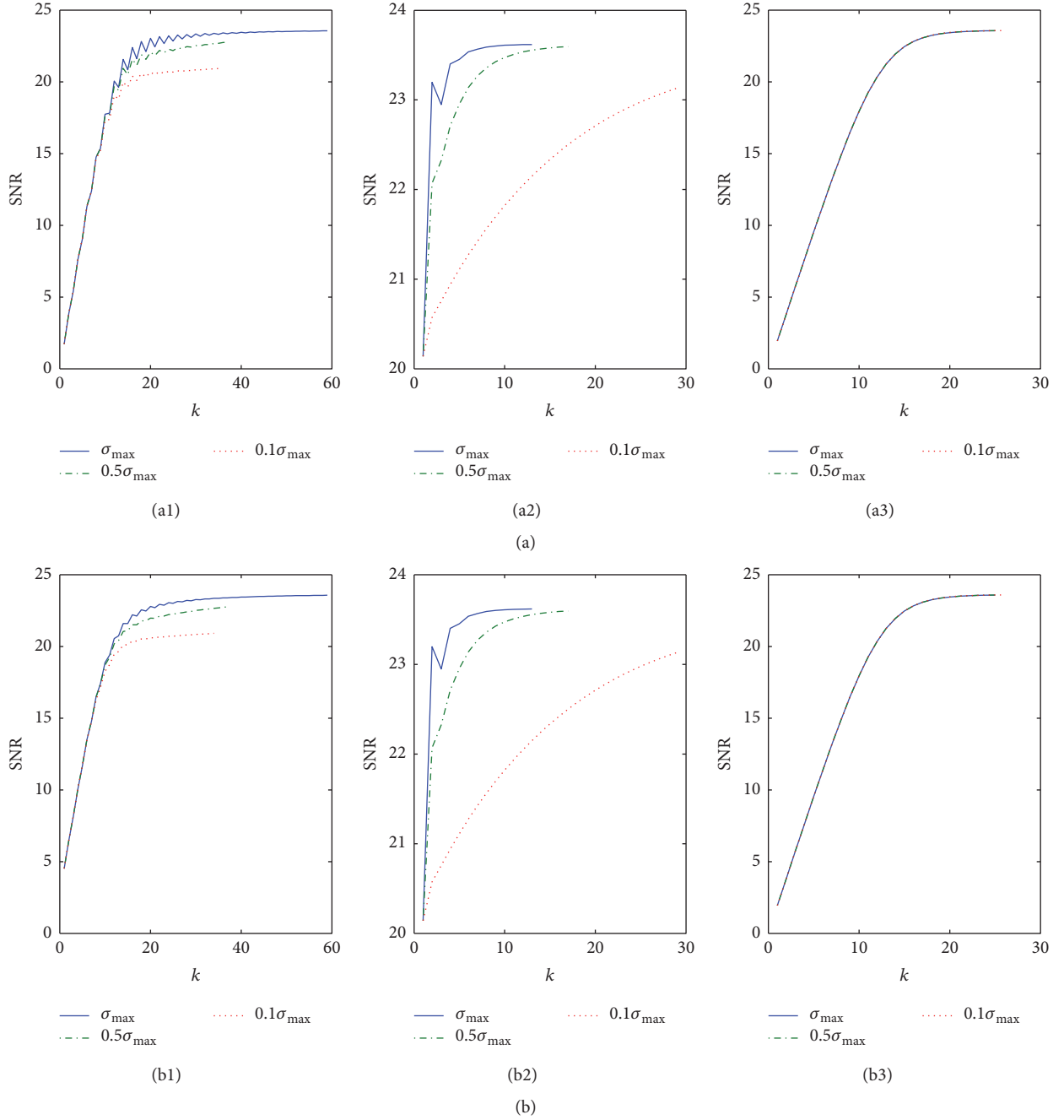
FIGURE 2: (a) The relation between SNR and iteration number with the nonnegative constraint, where $\tau = 0.9\tau_{\max}$ in (a1), $\tau = 0.5\tau_{\max}$ in (a2), and $\tau = 0.1\tau_{\max}$ in (a3). (b) The relation between SNR and iteration number with the bounded constraint, where $\tau = 0.9\tau_{\max}$ in (b1), $\tau = 0.5\tau_{\max}$ in (b2), and $\tau = 0.1\tau_{\max}$ in (b3).

Define $M_2 = P^{-1} \circ B$ and $k = (\|T^{-1}\| - \|\Sigma\|\|B\|^2)/\|E\|$; we will prove that $kM_2$ is 1-Lipschitz continuous. In fact, for any $z, z' \in Z$, we have

$$\left\| M_2(z) - M_2(z') \right\|_P^2 = \left\langle P^{-1} \circ B(z) - P^{-1} \right.$$

$$\circ B(z'), B(z) - B(z') \rangle_I$$

$$= \left\langle \Sigma^{-1} \left( T^{-1}\Sigma^{-1} - B^*B \right)^{-1} \right.$$

$$\cdot \left( \nabla f(x) - \nabla f(x') \right), \nabla f(x) \nabla f(x') \right\rangle \leq \left\| \Sigma^{-1} \right\|$$

$$\cdot \left( \left\| T^{-1}\Sigma^{-1} \right\| - \|B\|^2 \right)^{-1} \left\| \nabla f(x) - \nabla f(x') \right\|^2$$

$$\leq \beta^2 \left\| \Sigma^{-1} \right\| \left( \left\| T^{-1}\Sigma^{-1} \right\| - \|B\|^2 \right)^{-1} \left\| x - x' \right\|^2$$

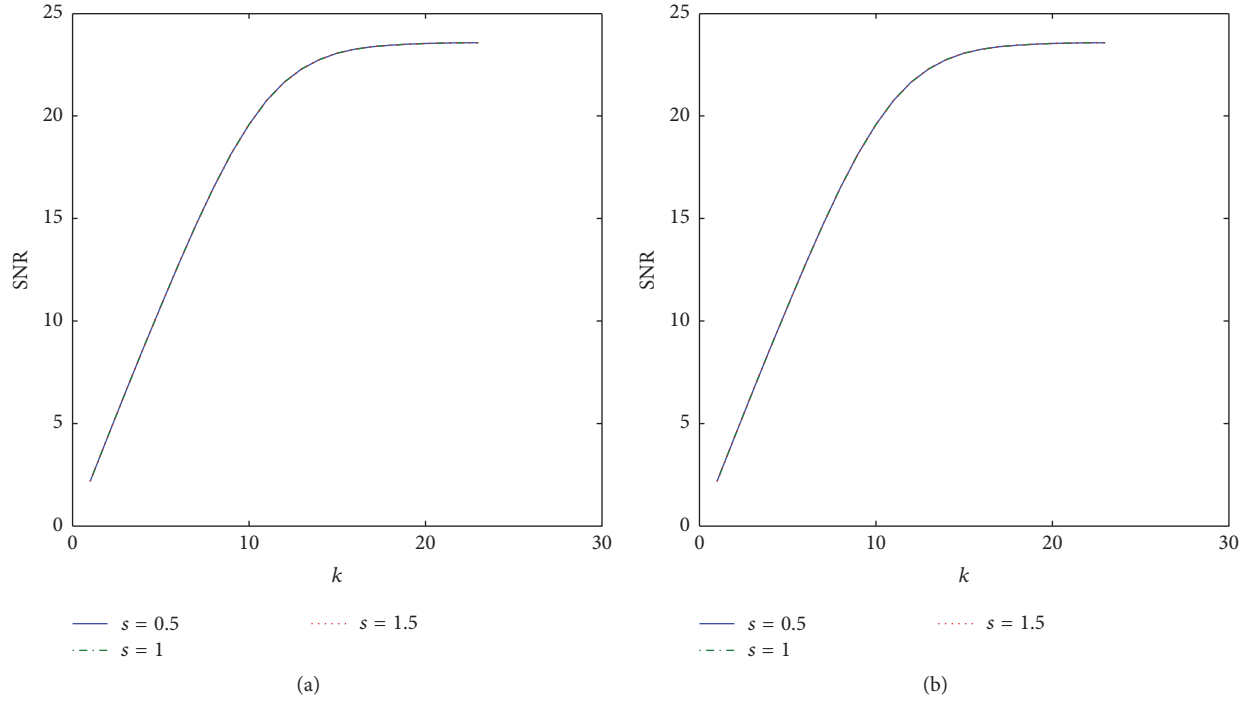$$\leq \frac{\|E\|}{k} \left\| x - x' \right\|^2.$$

$$(11)$$

FIGURE 3: (a) The relation between SNR and iteration number with the nonnegative constraint for different choice of $s$. (b) The relation between SNR and iteration number with the bounded constraint for different choice of $s$.

Let us define a linear operator $Q : (x, y) \rightarrow ((T^{-1} - \Sigma \|B\|^2)x, 0)$. Then, $P - Q$ is positive definite. So,

$$\|E\| k \|x - x'\|^2 \leq \left\langle z - z', Q\left(z - z'\right)\right\rangle_I$$
$$\leq \left\langle z - z', P\left(z - z'\right)\right\rangle = \|z - z'\|_P^2. \tag{12}$$

Substituting (12) into (11), we have

$$k \|M_2(z) - M_2(z')\|_P \leq \|z - z'\|_P. \tag{13}$$

The rest of the convergence proof follows the same argument of Theorem 2. So we omitted it here. This completes the proof. $\square$

As a matter of fact, $T$ and $\Sigma$ in Theorem 5 could be any symmetric, positive definite maps. In order to ensure that the proximity operator of $G$ and $F^*$ has a closed-form solution, it is sufficient to choose diagonal matrices for both of them. In the following, we give a practical way to choose the symmetric positive matrices and ensure the convergence of Theorem 5. To facilitate our proof, we need the following lemmas, which were obtained by [11].

**Lemma 6** (see [11]). *Let $K$ be a well-defined matrix and $T$ and $\Sigma$ be symmetric positive definite maps satisfying $\|\Sigma^{-1/2}KT^{-1/2}\|^2 < 1$. Define the matrix $M$ as*

$$M := \begin{pmatrix} T & -K^T \\ -K & \Sigma \end{pmatrix}. \tag{14}$$

*Then the matrix $M$ is symmetric and positive definite.*

Therefore, we are in the position to give our way to choose the matrices $T$ and $\Sigma$ accordingly.

**Lemma 7.** *Assume that $E$ is a diagonal matrix with $E = \text{diag}(\beta, \beta, \ldots, \beta)$, where $\beta$ is the Lipschitz constant of $\nabla f$. Fix $s \in [0, 2]$; let diagonal matrices $T = \text{diag}(\tau_1, \ldots, \tau_n)$ and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_m)$ with*

$$\tau_j = \frac{1}{\beta/2 + \sum_{i=1}^{m} \left|B_{i,j}\right|^{2-s}}, \quad j = 1, 2, \ldots, n,$$
$$\sigma_i = \frac{1}{\sum_{j=1}^{n} \left|B_{i,j}\right|^s}, \quad i = 1, 2, \ldots, m, \tag{15}$$

*and then it holds that*

$$T^{-1} - \frac{1}{2}E \succ 0, \quad \Sigma \succ 0, \tag{16}$$

$$\|T^{-1}\| - \|\Sigma\| \|B\|^2 > \frac{\|E\|}{2}. \tag{17}$$

*Proof.* Conclusion (16) follows from the definition of diagonal matrices $T$ and $\Sigma$. In fact,

$$\frac{1}{\tau_j} - \frac{1}{2}\beta = \sum_{i=1}^{m} \left|B_{i,j}\right|^{2-s} > 0, \quad \forall j = 1, 2, \ldots, n,$$
$$\sigma_i = \frac{1}{\sum_{j=1}^{n} \left|B_{i,j}\right|^s} > 0, \quad \forall i = 1, 2, \ldots, m. \tag{18}$$

(a1)                                                      (a2)                                                      (a3)

(b1)                                                      (b2)                                                      (b3)

(c1)                                                      (c2)                                                      (c3)

Figure 4: (a1)–(a3) are the noisy image of "Barbara" with noise levels 0.01, 0.05, and 0.1, respectively; (b1)–(b3) are the denoised image by Algorithm 1; (c1)–(c3) are the denoised image by our proposed Algorithm 4.

We will prove (17). It is easy to see that the proof of (17) is equivalent to the positive definite map of the following matrix $M$:

$$M := \begin{pmatrix} T^{-1} - \dfrac{1}{2}E & -B^T \\ -B & \Sigma^{-1} \end{pmatrix}. \tag{19}$$

From Lemma 6, it is sufficient to prove

$$\left\| \Sigma^{1/2} B \left( T^{-1} - \frac{1}{2}E \right)^{-1/2} \right\| < 1, \tag{20}$$

which ensure that the matrix $M$ is positive definite. For any $j = 1, 2, \ldots, n$, let us define $\widetilde{\tau}_j = \sum_{i=1}^{m} |B_{i,j}|^{2-s}$. We have

$$\left\| \Sigma^{1/2} B \left( T^{-1} - \frac{1}{2}E \right)^{-1/2} x \right\|^2$$

$$= \left\langle \Sigma^{1/2} B \left( T^{-1} - \frac{1}{2}E \right)^{-1/2} \right.$$

$$\left. \cdot x, \Sigma^{1/2} B \left( T^{-1} - \frac{1}{2}E \right)^{-1/2} x \right\rangle = \sum_{i=1}^{m} \sigma_i \left( \sum_{j=1}^{n} B_{i,j} \right.$$

(a1)

(a2)

(a3)

(b1)
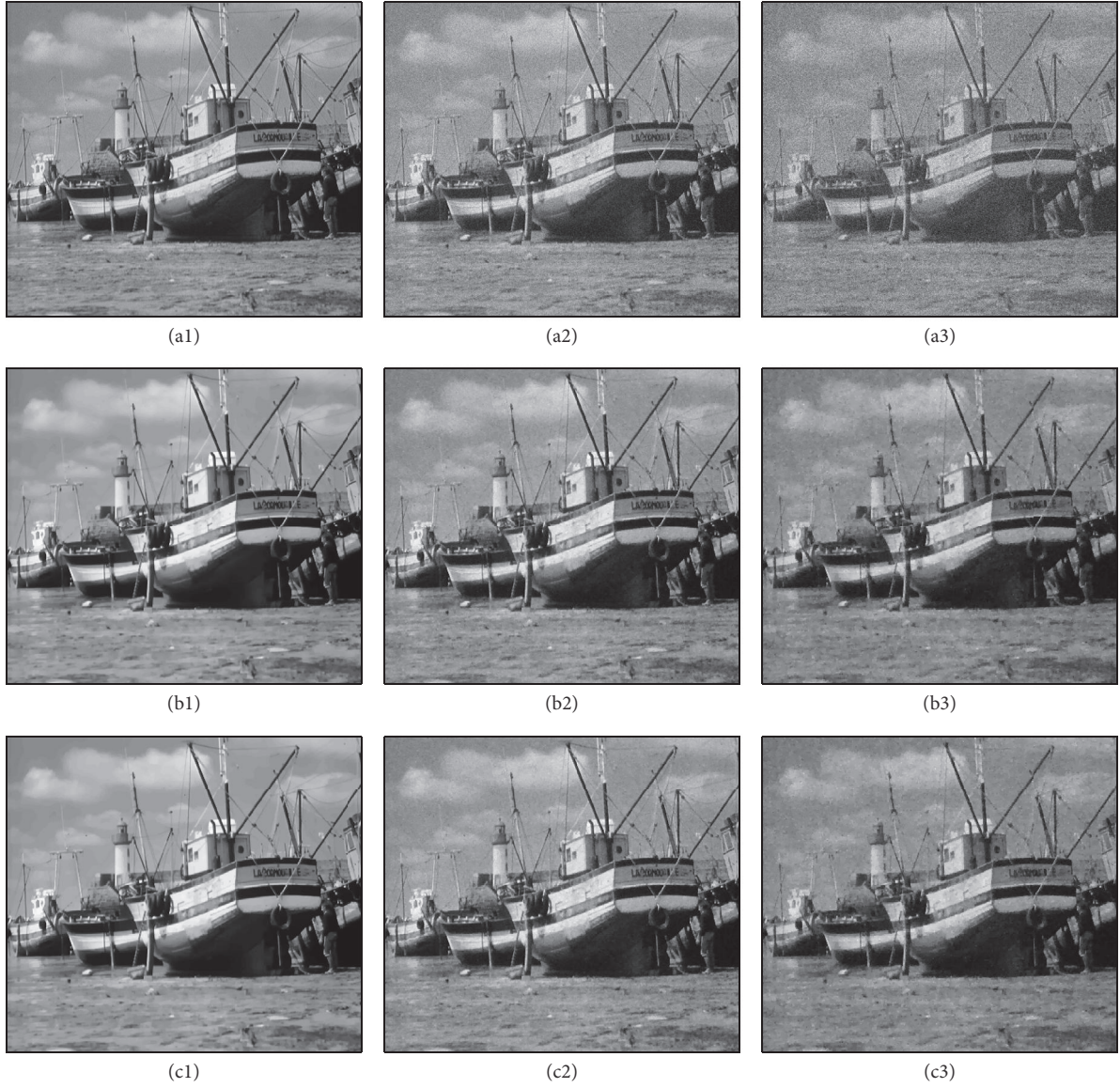
(b2)

(b3)

(c1)

(c2)

(c3)

Figure 5: (a1)–(a3) are the noisy image of "Boat" with noise levels 0.01, 0.05, and 0.1, respectively; (b1)–(b3) are the denoised image by Algorithm 1; (c1)–(c3) are the denoised image by our proposed Algorithm 4.

$$
\cdot \widetilde{\tau}_j^{-1/2} x_j \Bigg)^2
$$

$$
\leq \sum_{i=1}^m \sigma_i \left( \sum_{j=1}^n \left| B_{i,j} \right|^{s/2} \left| B_{i,j} \right|^{1-s/2} \widetilde{\tau}_j^{-1/2} x_j \right)^2
$$

(due to the Cauchy-Schwartz inequality)

$$
\leq \sum_{i=1}^m \sigma_i \left( \sum_{j=1}^n \left| B_{i,j} \right|^s \right) \left( \sum_{j=1}^n \left| B_{i,j} \right|^{2-s} \widetilde{\tau}_j^{-1} x_j^2 \right)
$$

(notice the condition of $\sigma_i$)

$$
= \sum_{i=1}^m \left( \sum_{j=1}^n \left| B_{i,j} \right|^{2-s} \widetilde{\tau}_j^{-1} x_j^2 \right)
$$

$$
= \sum_{j=1}^n \left( \sum_{i=1}^m \left| B_{i,j} \right|^{2-s} \widetilde{\tau}_j^{-1} x_j^2 \right)
$$

(from the definition of $\widetilde{\tau}_j$)

$$
= \|x\|.
$$

(21)

By the definition of the operator norm, we obtain

$$
\left\| \Sigma^{1/2} B \left( T^{-1} - \frac{1}{2} E \right)^{-1/2} \right\|^2 \leq 1.
$$

(22)

Figure 6: (a1)–(a3) are the noisy image of "Lena" with noise levels 0.01, 0.05, and 0.1, respectively; (b1)–(b3) are the denoised image by Algorithm 1; (c1)–(c3) are the denoised image by our proposed Algorithm 4.

Finally, the strict inequality of (20) can be obtained from the above proof process with one of the above inequalities becoming strictly smaller. □

## 4. Applications

In this section, we present an application of our proposed iterative algorithm. We aim at solving the following constrained total variation (TV) denoising problem:

$$\min_{x \in C} \quad \frac{1}{2} \|x - u\|_2^2 + \lambda \|x\|_{\mathrm{TV}}, \tag{23}$$

where $u \in R^n$ is a noisy image which was contaminated by Gaussian noise, $\lambda > 0$ is the regularization parameter, and $C$ is a closed convex set representing the prior information of the denoised image. By using the indicator function, constrained (TV) denoising problem (23) could be formulated by the following unconstrained optimization problem as follows:

$$\min_{x} \quad \frac{1}{2} \|x - u\|_2^2 + \lambda \|x\|_{\mathrm{TV}} + \iota_C(x), \tag{24}$$

where the indicator function $\iota_C \in \Gamma_0(R^n) : x \in R^n \mapsto$ {0, if $x \in C$; $+\infty$, otherwise}, since the total variation term

$\|x\|_{\mathrm{TV}}$ can be represented by a combination of convex function $\varphi$ and linear transformation matrix $B$; that is, $\|x\|_{\mathrm{TV}} = \varphi(Bx)$. See, for example, [18]. Then, optimization problem (24) is actually a special case of the general optimization problem (2) with $f(x) = (1/2)\|x - u\|_2^2$, $g(x) = \iota_C(x)$, and $h(x) = \lambda\varphi(x)$. Note that the gradient of the function $f(x)$ is $\nabla f(x) = x - u$ and the Lipschitz constant $\beta = 1$.

If the constrained set $C = R^n$, then the constrained TV denoising problem reduces to the unconstrained TV denoising problem:

$$\min_x \quad \frac{1}{2}\|x - u\|_2^2 + \lambda\|x\|_{\mathrm{TV}}. \tag{25}$$

The above TV denoising problem is often referred to as the ROF denoising model which was first introduced in computer vision by Rudin et al. [19].

*4.1. Numerical Experiments.* In the following, we present some preliminary numerical results and show the efficiency of our proposed methods. All the experiments are run on a personal Lenovo computer with Pentium(R) Dual-Core CPU @ 2.8 GHz and RAM 4 GB.

For all the tested iterative algorithms the stopping criterion is

$$\frac{\left\|x^{k+1} - x^k\right\|}{\left\|x^k\right\|} \le \epsilon, \tag{26}$$

where $\epsilon > 0$ is a given small constant, or the maximum iteration numbers 4000 are reached. The reconstructed image is evaluated in terms of the signal to noise (SNR) defined by

$$\mathrm{SNR} = 10\log_{10}\frac{\left\|x_{\mathrm{ori}}\right\|^2}{\left\|x_{\mathrm{rec}} - x_{\mathrm{ori}}\right\|^2}, \tag{27}$$

where $x_{\mathrm{ori}}$ is the original clear image and $x_{\mathrm{rec}}$ is the reconstructed image. The reconstructed time is denoted by "$T$ (s)" and the iteration number "Iter" is recorded when the stopping criteria satisfied. The tested images are the well-known "Barbara," "Boat," and "Lena". These images have the same size of $512 \times 512$ and are displayed in Figure 1.

*Experiment 1.* We present how the iterative parameters are chosen. First, we choose different combination of parameters $\sigma$ and $\tau$ and then apply Algorithm 1 to solve image denoising problem (23). We choose "Barbara" as the test image and add it by random Gaussian noise with zero mean and standard deviation 0.05. For convenience, we define $\tau_{\max} = 2/\beta$, $\sigma_{\max} = (1/\|B\|^2)(1/\tau - \beta/2)$. Here, the Lipschitz constant $\beta = 1$ and $\|B\|^2 = 8$. The numerical results are reported in Table 1.

Meanwhile, we plot the SNR versus the iteration numbers in Figure 2. We can see from Figure 2 that when the iterative parameter $\tau = 0.9\tau_{\max}$ and $\tau = 0.5\tau_{\max}$, the greater $\sigma$, the faster the convergence speed. For the small $\tau$, it has no apparent difference by the choice of parameter $\sigma$. So we select $\tau = 0.1\tau_{\max}$ and $\sigma = 0.1\sigma_{\max}$ for Algorithm 1 for the comparative experiments.

For our proposed Algorithm 4, the corresponding preconditioned iterative matrices are chosen according to

TABLE 1: Numerical results obtained by Algorithm 1.

| Constraint set $C$ | Parameter $\tau =$ | Parameter $\sigma =$ | $\epsilon = 10^{-3}$ | | |
|---|---|---|---|---|---|
| | | | $T$ (s) | Iter | SNR |
| $\{x \mid x \ge 0\}$ | $0.9\tau_{\max}$ | $\sigma_{\max}$ | 2.17 | 59 | 23.56 |
| | | $0.5 * \sigma_{\max}$ | 1.29 | 37 | 22.76 |
| | | $0.1 * \sigma_{\max}$ | 1.22 | 35 | 20.93 |
| | $0.5\tau_{\max}$ | $\sigma_{\max}$ | 0.46 | 13 | 23.61 |
| | | $0.5 * \sigma_{\max}$ | 0.60 | 17 | 23.59 |
| | | $0.1 * \sigma_{\max}$ | 1.02 | 29 | 23.13 |
| | $0.1\tau_{\max}$ | $\sigma_{\max}$ | 0.88 | 25 | 23.57 |
| | | $0.5 * \sigma_{\max}$ | 0.88 | 25 | 23.58 |
| | | $0.1 * \sigma_{\max}$ | 0.92 | 26 | 23.59 |
| $\{x \mid 0 \le x \le 1\}$ | $0.9\tau_{\max}$ | $\sigma_{\max}$ | 2.17 | 59 | 23.56 |
| | | $0.5 * \sigma_{\max}$ | 1.33 | 37 | 22.76 |
| | | $0.1 * \sigma_{\max}$ | 1.21 | 34 | 20.91 |
| | $0.5\tau_{\max}$ | $\sigma_{\max}$ | 0.48 | 13 | 23.62 |
| | | $0.5 * \sigma_{\max}$ | 0.62 | 17 | 23.60 |
| | | $0.1 * \sigma_{\max}$ | 1.07 | 29 | 23.13 |
| | $0.1\tau_{\max}$ | $\sigma_{\max}$ | 0.91 | 25 | 23.57 |
| | | $0.5 * \sigma_{\max}$ | 0.91 | 25 | 23.58 |
| | | $0.1 * \sigma_{\max}$ | 0.94 | 26 | 23.59 |

TABLE 2: Numerical results obtained by Algorithm 4.

| Constraint set $C$ | Parameter $s =$ | $\epsilon = 10^{-3}$ | | |
|---|---|---|---|---|
| | | $T$ (s) | Iter | SNR |
| $\{x \mid x \ge 0\}$ | 0.5 | 0.80 | 23 | 23.58 |
| | 1 | 0.81 | 23 | 23.58 |
| | 1.5 | 0.81 | 23 | 23.58 |
| $\{x \mid 0 \le x \le 1\}$ | 0.5 | 0.90 | 23 | 23.58 |
| | 1 | 0.82 | 23 | 23.58 |
| | 1.5 | 0.82 | 23 | 23.58 |

Lemma 7. There is only one parameter $s$ that needs to be set, and the experiments results are reported in Table 2. Figure 3 shows the SNR value versus the iteration number. We can see from it that the performance of the three choices of $s$ is nearly the same. So we choose $s = 1$ for Algorithm 4 in the following test.

*Experiment 2.* We show the performance comparison between Algorithm 1 and our proposed Algorithm 4. The constraint set $C$ is set as nonnegative set; that is, $C = \{x \mid x \ge 0\}$. To perform fair comparison, we add each of these images with random Gaussian noise with 0 mean value and different level of standard variation $\sigma$. The regularization parameter $\lambda = 0.02$ when the noise level $\sigma \le 0.05$ and $\lambda = 0.05$ when $\sigma > 0.05$.

From Table 3, we can see that our proposed Algorithm 4 converges faster than Algorithm 1 in terms of iteration numbers and iteration time in CPU time. For the large noise level, Algorithm 4 reaches higher SNR value than Algorithm 1 with less iteration numbers. Both of the algorithms get the

TABLE 3: Performance comparison between Algorithms 1 and 4.

| Image | Noise level | Methods | $\epsilon = 10^{-3}$ | | | $\epsilon = 10^{-4}$ | | |
|---|---|---|---|---|---|---|---|---|
| | | | $T$ (s) | Iter | SNR | $T$ (s) | Iter | SNR |
| "Barbara" | 0.01 | Algorithm 1 | 0.87 | 26 | 26.06 | 2.18 | 64 | 26.07 |
| | | Algorithm 4 | 0.80 | 23 | 26.03 | 1.17 | 34 | 26.08 |
| | 0.05 | Algorithm 1 | 0.88 | 26 | 23.59 | 1.86 | 55 | 23.60 |
| | | Algorithm 4 | 0.80 | 23 | 23.58 | 1.12 | 32 | 23.60 |
| | 0.1 | Algorithm 1 | 1.37 | 32 | 19.67 | 3.20 | 94 | 19.71 |
| | | Algorithm 4 | 0.81 | 23 | 19.70 | 1.26 | 36 | 19.71 |
| "Boat" | 0.01 | Algorithm 1 | 0.88 | 26 | 28.63 | 2.06 | 62 | 28.62 |
| | | Algorithm 4 | 0.81 | 23 | 28.56 | 1.18 | 34 | 28.63 |
| | 0.05 | Algorithm 1 | 1.02 | 26 | 25.70 | 1.82 | 54 | 25.72 |
| | | Algorithm 4 | 0.89 | 23 | 25.70 | 1.12 | 32 | 25.72 |
| | 0.1 | Algorithm 1 | 1.03 | 31 | 22.39 | 3.05 | 92 | 22.45 |
| | | Algorithm 4 | 0.79 | 23 | 22.44 | 1.23 | 36 | 22.45 |
| "Lena" | 0.01 | Algorithm 1 | 0.93 | 26 | 30.55 | 2.19 | 65 | 30.48 |
| | | Algorithm 4 | 0.79 | 23 | 30.39 | 1.21 | 35 | 30.48 |
| | 0.05 | Algorithm 1 | 0.87 | 26 | 26.47 | 1.92 | 57 | 26.51 |
| | | Algorithm 4 | 0.79 | 23 | 26.48 | 1.09 | 32 | 26.50 |
| | 0.1 | Algorithm 1 | 1.05 | 31 | 23.21 | 3.23 | 96 | 23.34 |
| | | Algorithm 4 | 0.80 | 23 | 23.33 | 1.26 | 37 | 23.34 |

cleared image, which have nearly the same SNR value finally. To visualize the reconstructed images, we present the noised image and the final denoising image in Figures 4, 5, and 6, respectively.

## 5. Conclusion

In this paper, we have studied the general optimization problem with the sum of three convex functions which is composed of a differential function with Lipschitz continuous gradient, a proximable function, and a linear composition function. Many interesting problems arising in image restoration and image reconstruction are special case of this problem. Inspired by the preconditioning technique proposed by Pock and Chambolle, we have introduced a primal-dual splitting algorithm with self-adaptive step-size to solve such problem. We also proposed a practical way to choose these step-sizes with a proof of convergence. Numerical results on image denoising problem showed that the precondition iterative algorithm performs better than the original one with constant step-size.

## Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this article.

## Acknowledgments

## References

[1] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460–479, 2013.

[2] J. M. Bioucas-Dias and M. A. Figueiredo, "A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2992–3004, 2007.

[3] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[4] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[5] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 564–574, 2007.

[6] P. Combettes and J. Pesquet, "Fixed-point algorithm for inverse problems in science and engineering," in *Proximal Splitting Methods in Signal Processing*, H. Bauschke, R. Burachik, P. Combettes, V. Elser, D. Luke, and H. Wolkowicz, Eds., pp. 185–212, Springer, New York, NY, USA, 2010.

[7] K. Choi, J. Wang, L. Zhu, T.-S. Suh, S. Boyd, and L. Xing, "Compressed sensing based cone-beam computed tomography reconstruction with a first-order method," *Medical Physics*, vol. 37, no. 9, pp. 5113–5125, 2010.

 [8] Y. Chen, W. Hager, F. Huang, D. Phan, X. Ye, and W. Yin, "Fast algorithms for image reconstruction with application to partially parallel MR imaging," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 90–118, 2012.

 [9] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[10] B. He and X. Yuan, "Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 119–149, 2012.

[11] T. Pock and A. Chambolle, "Diagonal preconditioning for first order primal-dual algorithms in convex optimization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 1762–1769, Barcelona, Spain, November 2011.

[12] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm," *Physics in Medicine and Biology*, vol. 57, no. 10, pp. 3065–3091, 2012.

[13] Y. Tang, J. Peng, S. Yue, and J. Xu, "A primal dual proximal point method of chambolle-pock algorithms for $\ell 1$-TV minimization problems in image reconstruction," in *Proceedings of the 5th International Conference on Biomedical Engineering and Informatics (BMEI '12)*, pp. 12–16, IEEE, Chongqing, China, October 2012.

[14] Y. Tang, Y. Cai, X. Wang, and J. Peng, "A primal dual proximal point method of Chambolle-Pock algorithm for total variation image reconstruction," in *Proceedings of the IEEE International Conference on Medical Imaging Physics and Engineering (ICMIPE '13)*, pp. 6–10, IEEE, October 2013.

[15] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued and Variational Analysis*, vol. 20, no. 2, pp. 307–330, 2012.

[16] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Advances in Computational Mathematics*, vol. 38, no. 3, pp. 667–681, 2013.

[17] P. L. Combettes, L. Condat, J.-C. Pesquet, and B. C. Vũ, "A forward-backward view of some primal-dual optimization methods in image recovery," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '14)*, pp. 4141–4145, IEEE, Paris, France, October 2014.

[18] C.-A. Micchelli, L. Shen, and Y. Xu, "Proximity algorithms for image models: denoising," *Inverse Problems*, vol. 27, no. 4, Article ID 045009, 30 pages, 2011.

[19] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.