Multi-Step Proximity Algorithms for Solving a Class of Convex Optimization Problems *

Qia Li † Lixin Shen‡ Yuesheng Xu †‡§ Na Zhang †

September 21, 2012

Abstract

We introduce in this paper a class of multi-step proximity algorithms for solving an optimization problem (in the context of image processing) whose objective function is the sum of two convex functions with one composed by an affine transformation. We are particularly interested in the scenario when the convex functions involved in the objective function have low regularity (not differentiable) since many problems encountered in image processing have this nature. We characterize the solutions of the optimization problem as fixed-points of a mapping defined in terms of the proximity operators of the two convex functions. A class of multi-step iterative schemes are developed based on the fixed-point equations appearing in the characterization of the solutions. For the purpose of studying convergence of the proposed algorithms, we introduce a notion of weakly firmly non-expansive mappings and establish under certain conditions that the sequence generated from a weakly firmly non-expansive mapping via a multi-step iteration is convergent. We use this general convergence result to conclude that the proposed multi-step algorithms converge. We in particular design a new two-step algorithm for solving the optimization problem which includes several existing algorithms as special examples. Moreover, we reinterpret the well-known alternating split Bregman iteration method as a special case of the proposed algorithm and modify it to improve its convergence result. Numerical experiments for the L1-TV image denoising model for impulsive noise removal are presented to compare the approximation accuracy and computational efficiency of the two-step algorithm with those of a benchmark algorithm of Chambolle and Pock.

1 Introduction

We consider in this paper solving an optimization problem whose objective function is the sum of two convex functions with one composed by an affine transformation. Such a problem has important applications in image and signal processing. For example, the Rudin-Osher-Fatemi (ROF) model of image denoising [38], the total-variation based impulsive noise removal model [31], high-resolution image reconstruction [10, 11], and the multiresolution sparse regularization model [17, 19] can all be considered as special cases of this optimization problem.

We are particularly interested in a scenario when both of the convex functions involved in the objective function of the optimization problem are not differentiable since the non-differentiability

*This research is supported in part by the US National Science Foundation under grant DMS-1115523, by Guangdong Provincial Government of China through the "Computational Science Innovative Research Team" program.

†Guangdong Province Key Laboratory of Computational Science, School of Mathematics and Computational Sciences, Sun Yat-sen University, Guangzhou 510275, P. R. China.

‡Department of Mathematics, Syracuse University, Syracuse, NY 13244, USA.

§Author to whom all correspondence should be sent. E-mail: yxu06@syr.edu.

presents computational challenges for solving the optimization problem. A number of numerical methods were proposed to address this issue. Taking the image deblurring model as an example, we classify these methods roughly into the two categories: direct optimization [1, 37, 38], solving the associated Euler-Lagrange equations [40, 41], and dual and primal-dual formulations [8, 13, 23]. Other methods including graph-based approaches [16, 33], spilt Bregman iteration [7, 20, 26], and Nesterov's first-order explicit [3, 30] schemes were proposed and the references cited therein. Similar methods have been proposed as well for the L1-TV model. These methods either smoothen the non-differentiable part of the model [14, 18], introduce intermediate variables which make the model separable, or consider the dual problem of the model [28, 32, 42]. The point of view that we adopt in this paper is that we solve the original problem directly (other than a smoothed or approximate problem) by characterizing the solutions of the optimization problem in terms of a system of fixed-point equations via *proximity operators*. We then develop efficient fixed-point algorithms for solving the equations which characterize the solutions.

Proximity operators provide a useful tool for the algorithmic development to be presented in this paper. When the proximity operators of the convex functions involved in the objective function have explicit expressions, the resulting algorithms are computationally efficient. Particularly, the proximity operators of the non-differentiable convex functions that we encounter in the area of image processing have close forms. Iterative algorithms developed from the characterization combined with the close forms of the proximity operators of the non-differentiable convex functions involved in it are extremely efficient in computation, since they bypass the difficulty caused by the non-differentiability of the convex functions. The computational efficiency of proximity algorithms was well demonstrated in the ROF model [27] for the ROF model and in [25, 28] for the L1-TV model.

In this paper, we first characterize the solutions by the proximity operators of the two convex functions. We find that a solution of the considered optimization problem can be viewed as a fixed-point of an operator. However, this operator is the composition of a nonexpansive operator with an expansive operator. Thus a simple explicit one-step iteration of this fixed-point equation might not be able to yield a solution for the problem. This phenomenon was numerically demonstrated in the L1-TV model for impulse removal[28]. Therefore, we resort to explore an implicit multi-step iteration to expect the sequence converges. We also give some conditions under which the implicit iteration is computable. We propose a notion weakly firmly nonexpansive to analyze the convergence of our algorithm. Eventually, we prove that under some condition, our established multi-step algorithm converges.

Most of the existing methods for solving the optimization problem are one-step iteration schemes, that is, they compute the next step iteration using only the results from the current step. If not only the results from the current step but also those from the previous steps are used when computing the next iteration step, we anticipate to obtain a better numerical result. Along this direction, a two-step primal-dual algorithm (Algorithm CP) was proposed in [9, 34] to solve the optimization problem. Algorithm CP introduces an extrapolation of the current and previous steps, which makes it a two-step algorithm. However, it is shown in [21] that Algorithm CP is equivalent to a one-step method and this point was used there in coming up a simple proof for convergence of Algorithm CP from the one-step method perspective. The main purpose of this paper is to develop multi-step methods for solving the optimization problem and to study their convergence properties. Convergence analysis for the multi-step methods is much more involved than that for one-step methods. To this end, we introduce a notion of weakly firmly non-expensive mappings and establish under certain conditions that the sequence generated from a weakly firmly non-expansive mapping via a multi-step iteration is convergent. We then use this general convergence result to conclude that the proposed specific multi-step algorithms converge.

The general framework described in this paper provides a platform for us to better understand the existing algorithms. Specifically, the alternating Split Bregman iteration (Algorithm ASBI) [20], the fixed-point algorithm for the L2-TV denoising model [27] and for the L1-TV denoising model [25], the algorithm CP, and the preconditioned first-order primal-dual algorithm [34] can all be viewed as special cases of the proposed iterative scheme. As a result, convergence of these existing algorithms can be easily shown from the general result. Moreover, the general framework can provide us more insight of the algorithms and perhaps improve upon them. For example, using the general framework to understand Algorithm ASBI leads to an improved Bregman iteration scheme. There are two main problems with Algorithm ASBI. First, the iteration involves solving a subproblem which has no closed form solution. Thus, inner solver for the subproblem is required to perform the iteration, which not only takes up more computational cost, but also increases the difficulty of implementing the algorithm. Second, as it was well-known, there are two sequences, one for the primal problem and the other for the dual problem, generated by the alternating split Bregman iteration for model in [20]. In the existing literature, it was only shown that the sequence for the dual formulation of model is convergent. The convergence of the sequence for the primal formulation is still absent. With our convergence analysis for the general framework, for Algorithm ASBI, we show that the sequence for the primal problem converges with respect to a symmetric semi-positive definite matrix. Even more, with our general framework and the corresponding convergence analysis, we can propose a way to modify Algorithm ASBI so that the sequence associated with the primal problem converges to a solution of the model.

We organize this paper in ten sections. In section 2 we characterize the solutions of the optimization problem as a fixed-point of a mapping defined in terms of the proximity operators of the two convex functions. Based on this characterization, we develop in section 3 a multi-step iterative algorithm. For the purpose of analyzing convergence of the proposed algorithm, we study in section 4 weakly firmly non-expansive operators and establish a general theory for convergence of a sequence generated from such operators via a multi-step iteration. As an application of the general convergence theory, we prove in section 5 that under certain proper conditions the multi-step iterative scheme proposed in section 3 converges to a solution of the optimization problem. In section 6 we design a particular two-step method and provide special results for the method to converge. We also discuss the possibility of using the Gauss-Seidel iteration in conjunction with the proposed two-step method aiming at accelerating its convergence speed. We further identify in section 7 four existing algorithms as special cases of the proposed two-step algorithm. In section 8 we reinterpret the well-known alternating split Bregman iteration method as a special case of the proposed algorithm and modify it to improve its convergence result. We present in section 9 numerical results for the L1-TV image denoising model for impulsive noise removal with a comparison of the approximation accuracy and computational efficiency of the proposed two-step algorithm with those of an algorithm of Chambolle and Pock. We conclude this paper in section 10.

2 The Optimization Problem and a Characterization of its Solutions

In this section we first describe the optimization problem considered in this paper. We then present a characterization of its solutions in terms of a system of fixed-point equations via proximity operators. The system of fixed-point equations will serve as a basis for developing multi-step iterative schemes for solving the problem.

For a positive integer n, by $\mathbb{R}^n$ we denote the usual n-dimensional Euclidean space. The minimization problem we consider in this paper has the form

$$\min\{\varphi(x) + \psi(Bx) : x \in \mathbb{R}^n\}, \tag{1}$$

where B is an $m \times n$ real matrix, $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ and $\psi : \mathbb{R}^m \to \overline{\mathbb{R}}$ are two proper lower semicontinuous convex functions. In the context of image processing, at least one of the functions φ and ψ is not differentiable. As a result, gradients of these convex functions do not necessarily exist and solutions of the optimization problem (1) can not be characterized completely by gradients of these convex functions. In this case, solutions of the optimization problem (1) may be characterized by the proximity operators of the convex functions involved in its objective function.

We now recall the notion of the proximity operator of a convex function. For x and y in the space $\mathbb{R}^d$, we denote the inner product by $\langle x, y\rangle := \sum_{i\in\mathbb{N}_d} x_i y_i$, where $\mathbb{N}_d := \{1, 2, \ldots, d\}$, and ℓ_2-norm by $\|x\|_2 := \langle x, x\rangle^{\frac{1}{2}}$. The ℓ_1-norm of $x \in \mathbb{R}^d$ is denoted by $\|x\|_1 := \sum_{i\in\mathbb{N}_d} |x_i|$. By $\mathbb{S}^d$ (resp. $\mathbb{S}^d_+$) we denote the sets of symmetric positive semi-definite (resp. definite) matrices. For an $H \in \mathbb{S}^d_+$ the weighted inner product is defined by $\langle x, y\rangle_H := \langle x, Hy\rangle$ and the corresponding ℓ_2-norm is defined by $\|x\|_H := \langle x, x\rangle_H{}^{\frac{1}{2}}$. By $\Gamma_0(\mathbb{R}^d)$ we denote the class of all lower semicontinuous convex functions $f : \mathbb{R}^d \to (-\infty, +\infty]$ such that $\mathrm{dom}(f) := \{x \in \mathbb{R}^d : f(x) < +\infty\} \neq \emptyset$. For a function $f \in \Gamma_0(\mathbb{R}^d)$, the proximity operator of f with respect to a given matrix $H \in \mathbb{S}^d_+$, denoted by $\mathrm{prox}_{f,H}$, is a mapping from $\mathbb{R}^d$ to itself, defined for a given point $x \in \mathbb{R}^d$ by

$$\mathrm{prox}_{f,H}(x) := \mathrm{argmin}\left\{\frac{1}{2}\|u - x\|_H^2 + f(u) : u \in \mathbb{R}^d\right\}.$$

This operator enjoys many interesting properties that make its role crucial in convex analysis (see, e.g., [29, 36]) and its applications (see, e.g., [4, 5, 15, 27]). In particular, we use prox_f for $\mathrm{prox}_{f,I}$.

The proximity operator of a function is intimately related to its subdifferential. The subdifferential of the function f at a given vector $x \in \mathbb{R}^d$ is the set defined by

$$\partial f(x) := \{y : \ y \in \mathbb{R}^d \quad \text{and} \quad f(z) \geq f(x) + \langle y, z - x\rangle \quad \text{for all} \quad z \in \mathbb{R}^d\}.$$

We remark that if a function $f \in \Gamma_0(\mathbb{R}^d)$ is differentiable at some point x then $\partial f(x) = \{\nabla f(x)\}$. It is shown that for any $H \in \mathbb{S}^d_+$, x in the domain of f and $y \in \mathbb{R}^d$,

$$Hy \in \partial f(x) \quad \text{if and only if} \quad x = \mathrm{prox}_{f,H}(x + y). \tag{2}$$

For a discussion of this relation, see, e.g., [2, Proposition 16.34] or [22].

We also need the notion of the conjugate function. The conjugate of $f \in \Gamma_0(\mathbb{R}^d)$ is the function $f^* \in \Gamma_0(\mathbb{R}^d)$ defined at $y \in \mathbb{R}^d$ by

$$f^*(y) := \sup\{\langle x, y\rangle - f(x) : x \in \mathbb{R}^d\}.$$

A characterization of the subdifferential of a function f in $\Gamma_0(\mathbb{R}^d)$ is that for $x \in \mathrm{dom}(f)$ and $y \in \mathrm{dom}(f^*)$

$$y \in \partial f(x) \quad \text{if and only if} \quad x \in \partial f^*(y). \tag{3}$$

Now, we are ready to characterize a solution of model (1) with the help of (2) and (3).

Proposition 2.1 *Let $\varphi \in \Gamma_0(\mathbb{R}^n)$, $\psi \in \Gamma_0(\mathbb{R}^m)$ and B an $m \times n$ matrix. If $x \in \mathbb{R}^n$ is a solution of problem (1), then for any $P \in \mathbb{S}^n_+$ and $Q \in \mathbb{S}^m_+$ there exists a vector $y \in \mathbb{R}^m$ such that*

$$x = \mathrm{prox}_{\varphi,P}(x - P^{-1}B^\top y), \tag{4}$$

$$y = \mathrm{prox}_{\psi^*,Q}(y + Q^{-1}Bx). \tag{5}$$

Conversely, if there exist $P \in \mathbb{S}^n_+$, $Q \in \mathbb{S}^m_+$, $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ satisfying equations (4) and (5), then x is a solution of problem (1).

Proof: We prove this proposition by applying the Fermat rule that a vector $x \in \mathbb{R}^n$ is a solution of model (1) if and only if the zero vector is in the subdifferential of the objective function of model (1) evaluated at x.

Let $x \in \mathbb{R}^n$ be a solution of model (1). From the Fermat rule and the chain rule of the subdifferential, we get that

$$0 \in \partial\varphi(x) + B^\top \partial\psi(Bx). \tag{6}$$

Thus, there exists $y \in \mathbb{R}^m$ such that $y \in \partial\psi(Bx)$ and $-B^\top y \in \partial\varphi(x)$. The last inclusion implies that for any $P \in \mathbb{S}^n_+$, $PP^{-1}(-B^\top y) \in \partial\varphi(x)$. Therefore, equation (4) follows from (2). By (3), from $y \in \partial\psi(Bx)$, we have that $Bx \in \partial\psi^*(y)$. Hence, for any $Q \in \mathbb{S}^m_+$, we obtain that $Q(Q^{-1}Bx) \in \partial\psi^*(y)$, which by (2) is equivalent to equation (5).

Conversely, suppose that there exist $P \in \mathbb{S}^n_+$, $Q \in \mathbb{S}^m_+$, $y \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ satisfying the system of fixed-point equations (4) and (5). The relation (2) ensures that $y \in \partial\psi(Bx)$ and $-B^\top y \in \partial\varphi(x)$. Clearly, these inclusions together ensure that the relation (6) holds. That is, the zero vector is in the subdifferential of the objective function at x. Again, by the Fermat rule, x is a solution of model (1). □

Proposition 2.1 characterizes a solution of model (1) in terms of the system of fixed-point equations (4) and (5). It make it possible for us to compute a solution of model (1) by developing fixed-point iterative scheme. Introducing the matrices P and Q in Lemma 2.1 plays important roles in developing algorithms for finding solutions of the coupled equations (4) and (5). They may be chosen so that the resulting algorithms converge or have better convergence rate [24, 34]. Equations (4) and (5) involve the proximity operators $\mathrm{prox}_{\varphi,P}$ and $\mathrm{prox}_{\psi^*,Q}$, respectively. We next express the proximity operator of a convex function with respect to a matrix $H \in \mathbb{S}^d_+$ in terms of the proximity operator of another convex function with respect to the identity matrix.

Lemma 2.2 *If $f \in \Gamma_0(\mathbb{R}^d)$ and $H \in \mathbb{S}^d_+$, then*

$$\mathrm{prox}_{f,H} = H^{-1/2} \circ \mathrm{prox}_{f\circ H^{-1/2}} \circ H^{1/2}. \tag{7}$$

Proof: Let $x \in \mathbb{R}^d$. It follows from the definition of the proximity operator that

$$\mathrm{prox}_{f\circ H^{-1/2}}(H^{1/2}x) = \mathrm{argmin}\left\{\frac{1}{2}\|u - H^{1/2}x\|_2^2 + f(H^{-1/2}u) : u \in \mathbb{R}^d\right\}.$$

Note that

$$\|H^{1/2}(H^{-1/2}u - x)\|_2^2 = \|H^{-1/2}u - x\|_H^2.$$

By a change of variables $s := H^{-1/2}u$ in the right-hand side of the above equation, we have that

$$\mathrm{prox}_{f\circ H^{-1/2}}(H^{1/2}x) = H^{1/2} \cdot \mathrm{argmin}\left\{\frac{1}{2}\|s - x\|_H^2 + f(s) : s \in \mathbb{R}^d\right\},$$

from which the result of this lemma follows. □

We remark that the proximity operator $\mathrm{prox}_{f\circ H^{-1/2}}$ appearing in the above proposition may be re-expressed in terms of the proximity operator of f according to [27]. Indeed, for $H \in \mathbb{S}^d_+$ and $z \in \mathbb{R}^d$, we have that

$$\mathrm{prox}_{f\circ H^{-1/2}}(z) = z - (H^{-1/2})^\top y$$

where $y \in \mathbb{R}^d$ is a solution of the equation

$$y = \frac{1}{\beta}(I - \mathrm{prox}_{\beta f})(H^{-1/2}z + (\beta I - H^{-1})y).$$

When we have an efficient way to compute $\text{prox}_{\beta f}$, we may solve the above fixed-point equation to obtain y. We need the matrix norm to describe a possible convergence of an algorithm for solving the fixed-point equation. For a $d \times d$ square matrix A, we define $\|A\|_H := \max\{\|Ax\|_H : x \in \mathbb{R}^d \text{ with } \|x\|_H = 1\}$. Clearly, if H is the identity matrix I, $\|A\|_H$ is simply written as $\|A\|_2$ which is the largest singular value of A. If $\beta > \|H^{-1}\|_2/2$ then the solutions of the above equation can be computed as the limit point of the sequence $\mathbf{y} := \{y^k : k \in \mathbb{N}\}$, where $\mathbb{N}$ is the set of all natural numbers, generated by

$$y^{k+1} = \frac{1}{\beta}(I - \text{prox}_{\beta f})(H^{-1/2}z + (\beta I - H^{-1})y^k)$$

for an arbitrary point y^0, see [27].

3 Multi-Step Algorithms

We develop in this section multi-step iterative schemes for solving the optimization problem (1) by using the system of fixed-point equations presented in the last section for the characterization of a solution of the problem.

We begin with a rewriting equations (4) and (5) in a compact form. To this end, we first introduce an operator by integrating the two proximity operators involved in equations (4) and (5) together. Specifically, for given $\varphi \in \Gamma_0(\mathbb{R}^n)$, $\psi \in \Gamma_0(\mathbb{R}^m)$, $P \in \mathbb{S}^n_+$, $Q \in \mathbb{S}^m_+$, we define the operator $\mathcal{T} = T^{(\varphi,P)}_{(\psi,Q)} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ at a vector $v = (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ as follows:

$$\mathcal{T}(v) := (\text{prox}_{\varphi,P}(x), \text{prox}_{\psi^*,Q}(y)). \tag{8}$$

Operator $\mathcal{T}$ couples the two proximity operators $\text{prox}_{\varphi,P}$ and $\text{prox}_{\psi^*,Q}$. It depends on functions φ, ψ^* and matrices P, Q. For convenient presentation, we use the notation $\mathcal{T}$ suppress the superscript and the subscript in the original notation. We next show that the operator $\mathcal{T}$ is the proximity operator of a new function

$$\Phi(v) := \varphi(x) + \psi^*(y)$$

for $v = (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ with respect to the matrix $R := \text{diag}(P, Q)$.

Lemma 3.1 *If the operator $\mathcal{T}$ is defined by (8), then $\mathcal{T}$ is the proximity operator of the function Φ with respect to the matrix R, that is, $\mathcal{T} = \text{prox}_{\Phi,R}$.*

Proof: Let $v := (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$. By the definition of the proximity operator and the structure of $\Phi(v) = \varphi(x) + \psi^*(y)$, we have that

$$\begin{aligned}
\text{prox}_{\Phi,R}(v) &= \text{argmin}\left\{\frac{1}{2}\|\widetilde{v} - v\|^2_R + \Phi(\widetilde{v}) : \widetilde{v} = (\widetilde{x}, \widetilde{y}) \in \mathbb{R}^n \times \mathbb{R}^m\right\} \\
&= \text{argmin}\left\{\frac{1}{2}\|\widetilde{x} - x\|^2_P + \varphi(\widetilde{x}) + \frac{1}{2}\|\widetilde{y} - y\|^2_Q + \psi^*(\widetilde{y}) : \widetilde{x} \in \mathbb{R}^n, \widetilde{y} \in \mathbb{R}^m\right\} \\
&= (\text{prox}_{\varphi,P}(x), \text{prox}_{\psi^*,Q}(y))
\end{aligned}$$

which by definition is $\mathcal{T}$. □

Recall that an operator $J : \mathbb{R}^d \to \mathbb{R}^d$ is called firmly nonexpansive (resp. nonexpansive) with respect to a given matrix $H \in \mathbb{S}^d_+$ if for all $x, y \in \mathbb{R}^d$

$$\|Jx - Jy\|^2_H \le \langle Jx - Jy, x - y\rangle_H \quad (\text{resp.} \quad \|Jx - Jy\|_H \le \|x - y\|_H).$$

Lemma 3.1 ensures that the operator $\mathcal{T}$ is firmly nonexpansive with respect to the matrix $R \in \mathbb{S}_+^{m+n}$.

For the $m \times n$ matrix B in model (1), we define the $(m+n) \times (m+n)$ skew-symmetric matrix S_B by

$$S_B := \begin{bmatrix} 0 & -B^\top \\ B & 0 \end{bmatrix}. \tag{9}$$

We also introduce the $(m+n) \times (m+n)$ matrix

$$E := I + \operatorname{diag}(P^{-1}, Q^{-1})S_B. \tag{10}$$

With the help of the above notation, equations (4) and (5) can be rewritten by a single equation of the form

$$v = (\mathcal{T} \circ E)(v). \tag{11}$$

Equation (11) indicates that finding a solution of model (1) essentially amounts to computing a fixed-point of the operator $\mathcal{T} \circ E$ for proper $P \in \mathbb{S}_+^n$ and $Q \in \mathbb{S}_+^m$. Lemma 2.1 ensures the existence of fixed-points of the operator. We shall then focus on developing efficient schemes for finding a fixed-point of the operator. We first show the expansivity of the operator E which explains that a simple iteration $v^{k+1} = (\mathcal{T} \circ E)(v^k)$ for a given initial guess v^0, would not yield a convergent sequence $\mathbf{v} := \{v^k : k \in \mathbb{N}\}$.

Lemma 3.2 *There holds that* $\|E\|_R > 1$.

Proof: Let $v := (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ with $\|v\|_R = 1$. By the definition of the matrix S_B, we have that

$$Ev = \begin{bmatrix} x - P^{-1}B^\top y \\ Q^{-1}Bx + y \end{bmatrix}.$$

It follows that

$$\|Ev\|_R^2 = \|x - P^{-1}B^\top y\|_P^2 + \|Q^{-1}Bx + y\|_Q^2.$$

Noting that

$$\|x - P^{-1}B^\top y\|_P^2 = \|x\|_P^2 - 2\langle x, B^\top y\rangle + \|P^{-1}B^\top y\|_P^2,$$
$$\|Q^{-1}Bx + y\|_Q^2 = \|Q^{-1}Bx\|_Q^2 + 2\langle x, B^\top y\rangle + \|y\|_Q^2$$

and $\|x\|_P^2 + \|y\|_Q^2 = \|v\|_R^2$, we observe that

$$\|Ev\|_R^2 \geq 1 + \|P^{-1}B^\top y\|_P^2 + \|Q^{-1}Bx\|_Q^2.$$

Since B is a non-zero matrix, there must exist a vector $v := (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ with $\|v\|_R = 1$ such that the last two terms on the right-hand side of the above equation is positive. This implies that the norm $\|E\|_R$ is strictly greater than 1. □

From Lemmas 3.1 and 3.2, we see that $\mathcal{T} \circ E$ is the composition of a nonexpansive operator $\mathcal{T}$ with an expansive operator E with respect to the matrix R. Consequently, a simple explicit iteration for equation (11) may not be able to yield a solution to the equation. This phenomenon was observed numerically in the L1-TV model for impulsive noise removal [28]. For this reason, we resort to exploring implicit (instead of explicit) iterations. We point out that the meaning of implicit iterations mentioned here is implicit iterations with respect to the entire operator $\mathcal{T} \circ E$. Under certain circumstances, the iterations may be solved explicitly. Moreover, when computing a new update, we use not only the value of a current step but also values of several previous steps.

For this purpose, we choose a set of $(m+n)\times(m+n)$ matrices M_i, $i \in \mathbb{N}_\ell^0 := \{0,1,\ldots,\ell\}$, satisfying the relation $M_0 = \sum_{i\in\mathbb{N}_\ell} M_i$ and rewrite the matrix E as

$$E = E_0 + E_1, \quad \text{where} \quad E_0 := I + R^{-1}(S_B - M_0), \quad E_1 := R^{-1}\sum_{i\in\mathbb{N}_\ell} M_i.$$

We then rewrite equation (11) in an equivalent form

$$v = \mathcal{T}(E_0 v + E_1 v). \tag{12}$$

Consequently, based on equation (12), we propose multi-step implicit iterative schemes that for given $v^1,\ldots,v^\ell$, the vectors v^{k+1}, for $k \geq \ell$, are calculated by

$$v^{k+1} = \mathcal{T}\left(E_0 v^{k+1} + R^{-1}\sum_{i\in\mathbb{N}_\ell} M_i v^{k-i+1}\right). \tag{13}$$

In particular, when $\ell = 2$ and $M_2 = 0$ (in this case, $M_0 = M_1$), equation (13) reduces to

$$v^{k+1} = \mathcal{T}\left(E_0 v^{k+1} + R^{-1} M_0 v^k\right). \tag{14}$$

Many efficient algorithms including several well-known existing algorithms can be formulated in the form of (14). We shall provide in-depth discussion for this important special case in Sections 6 and 8.

We next consider the solvability of equation (13) when values $v^1,\ldots,v^\ell$ are given. For a $p\times q$ matrix A, the null space of A is the set $\mathcal{N}(A) := \{x \in \mathbb{R}^q : Ax = 0\}$, and the range of A is the set $\mathcal{R}(A) := \{Ax : x \in \mathbb{R}^q\}$. We denote by $A^\dagger$ the Moore-Penrose pseudoinverse of A.

Proposition 3.3 *Let $\varphi \in \Gamma_0(\mathbb{R}^n)$ and $\psi \in \Gamma_0(\mathbb{R}^m)$, B an $m\times n$ matrix, $P \in \mathbb{S}_+^n$ and $Q \in \mathbb{S}_+^m$, $a := (a_1, a_2) \in \mathbb{R}^n \times \mathbb{R}^m$, and M a block matrix having the form*

$$A = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix},$$

where the square matrices M_{11} and M_{22} are of size $n\times n$ and $m\times m$, respectively. If the following conditions are satisfied:

(i) $M_{11} \in \mathbb{S}^n$, $M_{22} \in \mathbb{S}^m$,

(ii) both ψ^* and φ are coercive, that is, $\psi^*(y) \to +\infty$ as $\|y\|_2 \to +\infty$ and $\varphi(x) \to +\infty$ as $\|x\|_2 \to +\infty$,

(iii) $\mathcal{R}(M_{21} - B) \subseteq (\mathcal{N}(M_{22}))^\perp$ and $\mathcal{R}(M_{12} + B^\top) \subseteq (\mathcal{N}(M_{11}))^\perp$,

(iv) $a_1 \in (\mathcal{N}(M_{11}))^\perp$ and $a_2 \in (\mathcal{N}(M_{22}))^\perp$,

(v) either $M_{12} = -B^\top$ or $M_{21} = B$,

then there is a vector $v \in \mathbb{R}^{n+m}$ satisfying the equation

$$v = \mathcal{T}\left(v + R^{-1}((S_B - A)v + a)\right). \tag{15}$$

Proof: Write $v := (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ and set $\Phi(v) := \varphi(x) + \psi^*(y)$. By relation (2) and Lemma 3.1, v is a solution of equation (15) if and only if

$$(S_B - A)v + a \in \partial\Phi(v),$$

which, by exploiting the structures of M and S_B, is equivalent to the following two inclusions

$$-(M_{11}x + B^\top y + M_{12}y + a_1) \in \partial\varphi(x), \tag{16}$$

$$Bx - M_{21}x - M_{22}y + a_2 \in \partial\psi^*(y). \tag{17}$$

Hence, it suffices to show that there exists a pair $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying (16) and (17).

The rest of this proof focuses on the existence of such a pair (x, y). To this end, we first establish that for any fixed y in $\mathbb{R}^m$ there exists a $x \in \mathbb{R}^n$ that satisfies (16). We then show that for any fixed $x \in \mathbb{R}^n$ there exists a $y \in \mathbb{R}^m$ that satisfies (17). Finally, we prove that equations (16) and (17) can be simultaneously satisfied by a pair $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$.

For any fixed y in $\mathbb{R}^m$, we consider the minimization problem

$$\min\left\{\varphi(x) + \frac{1}{2}\left\|M_{11}^{\frac{1}{2}}x + (M_{11}^\dagger)^{\frac{1}{2}}(B^\top y + M_{12}y + a_1)\right\|_2^2 : x \in \mathbb{R}^n\right\}. \tag{18}$$

From Items (i) and (ii), we conclude that the set of minimizers of problem (18) is nonempty. Suppose that x is a minimizer of problem (18). By the Fermat rule and the chain rule of the subdifferential, we get that

$$0 \in \partial\varphi(x) + M_{11}^{\frac{1}{2}}\left(M_{11}^{\frac{1}{2}}x + (M_{11}^\dagger)^{\frac{1}{2}}(B^\top y + M_{12}y + a_1)\right). \tag{19}$$

From the second inclusion in Item (iii) and from the first condition in Item (iv) we obtain that $B^\top y + M_{12}y + a_1 \in (\mathcal{N}(M_{11}))^\perp$. Hence, $M_{11}^{\frac{1}{2}}(M_{11}^\dagger)^{\frac{1}{2}}(B^\top y + M_{12}y + a_1) = B^\top y + M_{12}y + a_1$. This together with (19) leads to inclusion (16).

For any fixed x in $\mathbb{R}^n$, we consider the minimization problem

$$\min\left\{\psi^*(y) + \frac{1}{2}\left\|M_{22}^{\frac{1}{2}}y + (M_{22}^\dagger)^{\frac{1}{2}}(-Bx + M_{21}x - a_2)\right\|_2^2 : y \in \mathbb{R}^m\right\}. \tag{20}$$

In a way similar to how equation (16) is derived, we obtain inclusion (17) from (20).

If the first case of Item (v) holds, then the variable y in equations (18) and (16) disappears. For a solution x of problem (18), we can find y that satisfies the inclusion (17). If the second case of Item (v) holds, then the variable x in equations (20) and (17) disappears. For a solution y of problem (20), we can find x that satisfies the inclusion (18). The proof of this proposition is complete. $\square$

To close this section, we make several comments on Proposition 3.3. If we assume that $M_{11} \in \mathbb{S}_+^n$, then the assumption of the coerciveness for φ can be deleted. Furthermore, the condition $M_{11} \in \mathbb{S}_+^n$ implies $(\mathcal{N}(M_{11}))^\perp = \mathbb{R}^n$. Therefore, the second inclusion in Item (iii) and the first condition in Item (iv) are automatically satisfied. Similarly, if $M_{22} \in \mathbb{S}_+^m$, the assumption of ψ^* being coercive is not necessary and the first inclusion in Item (iii) and the second condition in Item (iv) are automatically satisfied. Item (v) also plays a role of computing $v^{k+1} := (x^{k+1}, y^{k+1})$ via decoupling the x^{k+1} and y^{k+1} in a way of computing x^{k+1} first then followed by y^{k+1} if $M_{12} = -B^\top$ or vise versa if $M_{21} = B$.

4 Weakly Firmly Nonexpansive Operators

We study in this section a class of set-valued operators whose definition is motivated from iterative scheme (13), aiming at analyzing convergence of the sequence generated from the iteration. Specifically, we let $2^{\mathbb{R}^d}$ denote the collection of all sets of vectors in $\mathbb{R}^d$ and investigate operators $T : \mathbb{R}^{\ell d} \to 2^{\mathbb{R}^d}$. We introduce the notion of weakly firmly nonexpansive operators and prove under some conditions the sequence generated by a weakly firmly nonexpansive operator converges to a fixed-point of the operator.

We first motivate the definition of the set-valued operator from iterative scheme (13). Associated with scheme (13) we define a block matrix by letting $M := \begin{bmatrix} M_1 & M_2 & \dots & M_\ell \end{bmatrix}$. Suppose that for any $u \in \mathbb{R}^{\ell(m+n)}$ there exists $w \in \mathbb{R}^{m+n}$ such that

$$w = \mathcal{T}\left(E_0 w + R^{-1} M u\right). \tag{21}$$

Proposition 3.3 ensures the existence of a solution of equation (21) under certain conditions. Equation (21) defines implicitly a nonlinear map from $u \in \mathbb{R}^{\ell(m+n)}$ to $w \in \mathbb{R}^{m+n}$. We describe below precisely the nonlinear operator. Let $T_{\mathcal{M}} : \mathbb{R}^{\ell(m+n)} \to 2^{\mathbb{R}^{m+n}}$ be defined as

$$T_{\mathcal{M}} : u \to \{w : w \in \mathbb{R}^{m+n}, (w, u) \text{ satisfies equation (21)}\}. \tag{22}$$

The sequence $\mathbf{v}$ generated by scheme (13), for given $v^1, v^2, \dots, v^\ell$, can be viewed as the sequence generated by the following iteration associated with the operator $T_{\mathcal{M}}$

$$v^{k+1} \in T_{\mathcal{M}}(v_\ell^k), \tag{23}$$

where $v_\ell^k := (v^k, v^{k-1}, \dots, v^{k-\ell+1})^\top$. The convergence analysis of the iteration (23) demands a study of the set-valued operator $T_{\mathcal{M}}$.

Motivated from the above discussion, we consider a general set-valued operator $T : \mathbb{R}^{\ell d} \to 2^{\mathbb{R}^d}$. By $\mathrm{dom}(T)$ we denote the domain of T, which is defined by $\mathrm{dom}(T) := \{u : u \in \mathbb{R}^{\ell d}, T(u) \neq \emptyset\}$. The graph $\mathrm{gra}(T)$ of T is defined by

$$\mathrm{gra}\ (T) := \left\{(u, w) : (u, w) \in \mathbb{R}^{\ell d} \times \mathbb{R}^d : w \in T(u)\right\}.$$

We say T is continuous if the graph of T is a closed set, that is, for any sequence $\{(u^k, w^k) \in \mathrm{gra}(T) : k \in \mathbb{N}\}$ converging to (u, w) we have that $(u, w) \in \mathrm{gra}(T)$. We say w is a fixed-point of T if $w \in T(1_\ell \otimes w)$, where $\otimes$ is Kronecker product and 1_ℓ is an ℓ dimensional vector whose components are all 1.

We consider the sequence w^k, for $k \geq \ell + 1$, generated by the iteration

$$w^{k+1} \in T(w_\ell^k), \tag{24}$$

when $w^1, w^2, \dots, w^\ell$ are given. A central issue for this iteration is the convergence analysis of the sequence generated from (24). It requires the availability of a mathematical tool. Recall that if $T : \mathbb{R}^d \to \mathbb{R}^d$ is firmly nonexpansive with respect to a $d \times d$ symmetric positive definite matrix A, then the Picard sequence generated by the operator T converges to a fixed-point of the operator. Motivated by the notion of the firmly nonexpansive operator, we define below the weakly firmly nonexpansive operator for the class of the set-valued operators. For a set of $d \times d$ matrices $\mathcal{M} := \{M_i : i \in \mathbb{N}_\ell\}$, we define the block matrix M by

$$M := \begin{bmatrix} M_1 & M_2 & \dots & M_\ell \end{bmatrix}. \tag{25}$$

Definition 4.1 *We say that an operator* $T : \mathbb{R}^{\ell d} \to 2^{\mathbb{R}^d}$ *is weakly firmly nonexpansive with respect to* $\mathcal{M}$ *if for any* $(u_i, w_i) \in \operatorname{gra}(T)$, *for* $i = 1, 2$,

$$\langle w_2 - w_1, M_0(w_2 - w_1)\rangle \le \langle w_2 - w_1, M(u_2 - u_1)\rangle. \tag{26}$$

In Definition 4.1, the set $\mathcal{M}$ of matrices may be considered as "weights" of the "weighted inner products" used in inequality (26). They need to satisfy certain conditions. We next describe these conditions. A set of $d \times d$ matrices $\{M_i : i \in \mathbb{N}_\ell\}$ for $\ell \ge 2$ is said to satisfy *Semi-Condition-M* if the following five hypotheses are satisfied:

(i) $M_0 = \sum_{i=1}^{\ell} M_i$

(ii) $M_1 = M_2 = \cdots = M_{\ell-1}$,

(iii) $H := M_0 + M_\ell$ is in $\mathbb{S}^d$,

(iv) $\mathcal{N}(H) \subseteq \mathcal{N}(M_\ell) \cap \mathcal{N}(M_\ell^\top)$

(v) $\left\|(H^\dagger)^{\frac{1}{2}} M_\ell (H^\dagger)^{\frac{1}{2}}\right\|_2 < \frac{1}{2}$.

Furthermore, if $\mathbb{S}^d$ in Item (iii) is replace by $\mathbb{S}^d_+$, we call these five conditions *Condition-M*.

We remark that when $\ell = 2$, $M_2 = 0$ and $\mathcal{M}$ satisfies Condition-M ($M_0 = M_1$ is symmetric positive definite), the statement that T is weakly firmly nonexpansive with respect to $\mathcal{M}$ is equivalent to that T is firmly nonexpansive with respect to M_0. We assume $\ell \ge 2$ in the rest of this paper. The first three hypotheses implies that the matrices M_i, $i \in \mathbb{N}_{\ell-1}$, can be represented by M_ℓ and H. More precisely,

$$M_0 = H - M_\ell \quad \text{and} \quad M_i = \frac{1}{\ell - 1}(H - 2M_\ell), \tag{27}$$

for $1 \le i \le \ell - 1$. With relation (27) and the fourth hypothesis of Condition-M, we have that

$$\mathcal{N}(H) \subseteq \mathcal{N}(M_i) \cap \mathcal{N}(M_i^\top), \tag{28}$$

for $i \in \mathbb{N}_\ell$.

The main purpose of this section is to prove that if a continuous operator T is weakly firmly nonexpansive with respect to $\mathcal{M}$ satisfying Condition-M, then the sequence $\mathbf{w} := \{w^k : k \in \mathbb{N}\}$, generated by (24), converges to a fixed point of T. This result is proved by a number of technical lemmas.

We first present an implication of relation (28).

Lemma 4.2 *For* $G \in \mathbb{S}^d$ *and a* $d \times d$ *matrix* A, *if*

$$\mathcal{N}(G) \subseteq \mathcal{N}(A) \cap \mathcal{N}(A^\top) \tag{29}$$

then for any $w, s \in \mathbb{R}^d$,

$$\langle w, As\rangle = \langle G^{\frac{1}{2}} w, G^{\dagger\frac{1}{2}} A G^{\dagger\frac{1}{2}} G^{\frac{1}{2}} s\rangle \tag{30}$$

and

$$A = G^{\frac{1}{2}} G^{\dagger\frac{1}{2}} A G^{\dagger\frac{1}{2}} G^{\frac{1}{2}}. \tag{31}$$

Proof: For any $w, s \in \mathbb{R}^d$, there exist unique decompositions $w = w_1 + w_2$ and $s = s_1 + s_2$ where $w_1, s_1 \in \mathcal{N}(G)$ and $w_2, s_2 \in (\mathcal{N}(G))^\perp$. By the assumption (29), we know that $As_1 = 0$ and $A^\top w_1 = 0$. Hence,

$$\langle w, As\rangle = \langle w_2, As_2\rangle. \tag{32}$$

Note that $G^{\dagger\frac{1}{2}}G^{\frac{1}{2}}w_2 = w_2$, $G^{\dagger\frac{1}{2}}G^{\frac{1}{2}}s_2 = s_2$, and $(G^\dagger)^{\frac{1}{2}}$ is symmetric. We further have that

$$\langle w_2, As_2\rangle = \langle G^{\frac{1}{2}}w_2, (G^\dagger)^{\frac{1}{2}}A(G^\dagger)^{\frac{1}{2}}G^{\frac{1}{2}}s_2\rangle,$$

which yields, due to $\mathcal{N}(G) = \mathcal{N}(G^{\frac{1}{2}})$, the following result

$$\langle w_2, As_2\rangle = \langle G^{\frac{1}{2}}w, (G^\dagger)^{\frac{1}{2}}A(G^\dagger)^{\frac{1}{2}}G^{\frac{1}{2}}s\rangle. \tag{33}$$

Thus, equation (30) follows from equations (32) and (33). Equation (31) is a direct consequence of (30). □

Applying Lemma 4.2 to the case $A = M_i$ and using the inclusion (28), we conclude that for any $w, s \in \mathbb{R}^d$

$$\langle w, M_i s\rangle = \langle H^{\frac{1}{2}}w, H^{\dagger\frac{1}{2}}M_i H^{\dagger\frac{1}{2}}H^{\frac{1}{2}}s\rangle, \quad \text{for} \quad 0 \le i \le \ell. \tag{34}$$

We next present a formula used frequently in convergence analysis. If $\{a^k : k \in \mathbb{N}\}$ is a sequence in $\mathbb{R}$, then for $1 \le s \le N-1$ and $1 \le i \le s$ there holds the inequality

$$\sum_{k=s}^{N-1}(a^{k+1} - a^{k-i+1}) = \sum_{j=1}^{i}(a^{N-j+1} - a^{s-j+1}). \tag{35}$$

Suppose that w is a fixed-point of T. From Definition 4.1, we get the following inequality

$$\langle w^{k+1} - w, M_0(w^{k+1} - w)\rangle \le \sum_{i\in\mathbb{N}_\ell}\langle w^{k+1} - w, M_i(w^{k-i+1} - w)\rangle, \quad \text{for all} \quad k \ge \ell. \tag{36}$$

For a set $\mathcal{M} := \{M_i : i \in \mathbb{N}_\ell\}$ of $d \times d$ matrices satisfying Condition-M and for a nonempty subset $\mathcal{W}$ of $\mathbb{R}^d$, we let

$$\widetilde{M} := (H^\dagger)^{\frac{1}{2}}M_\ell(H^\dagger)^{\frac{1}{2}}$$

and for a fixed $w \in \mathcal{W}$ let $e^k := w^k - w$ and $r_i^k := w^k - w^{k-i}$.

Lemma 4.3 *Suppose that a set $\mathcal{M}$ of $d \times d$ matrices satisfies Semi-Condition-M and $\mathcal{W}$ is a nonempty subset of $\mathbb{R}^d$. If $\mathbf{w} := \{w^k : k \in \mathbb{N}\}$ satisfies (36) for any $w \in \mathcal{W}$, then for any $\alpha > 0$, $\ell \ge 2$ and $\ell \le s \le N-1$,*

$$(1-\alpha)\sum_{i\in\mathbb{N}_{\ell-1}}(\ell-i)|e^{N-i+1}|_H^2 \le C_s - \left(1 - \alpha - \frac{\|\widetilde{M}\|_2^2}{\alpha}\right)\sum_{k=s}^{N-1}\sum_{i\in\mathbb{N}_{\ell-1}}|r_i^{k+1}|_H^2, \tag{37}$$

where $|\cdot|_H^2 := \langle\cdot, H\cdot\rangle$ and

$$C_s := \sum_{i\in\mathbb{N}_{\ell-1}} i|e^{s-\ell+i+1}|_H^2 + \frac{\|\widetilde{M}\|_2^2}{\alpha}\sum_{i\in\mathbb{N}_{\ell-1}}\sum_{j=1}^{\ell-i}|r_i^{s-j+1}|_H^2 - 2\sum_{i\in\mathbb{N}_{\ell-1}}\sum_{j=1}^{\ell-i}\langle e^{s-j+1}, M_\ell r_i^{s-j+1}\rangle \tag{38}$$

is independent of N.

Proof: We first consider the special case $\ell = 2$ and $M_2 = 0$. In this case, $\mathcal{M}$ satisfying Semi-Condition-M is equivalent to that $M_0 = M_1$ is in $\mathbb{S}^d$. Inequality (36) reduces to

$$\langle e^{k+1}, M_0 e^{k+1}\rangle \leq \langle e^{k+1}, M_0 e^k\rangle. \tag{39}$$

Since $M_0 \in \mathbb{S}^d$ and $r_1^{k+1} = e^{k+1} - e^k$, we get that

$$\langle e^{k+1}, M_0 e^k\rangle = \frac{1}{2}(|e^{k+1}|^2_{M_0} + |e^k|^2_{M_0} - |r^{k+1}|^2_{M_0}).$$

Thus (39) can be rewritten as

$$|e^{k+1}|^2_{M_0} \leq |e^k|^2_{M_0} - |r^{k+1}|^2_{M_0}.$$

Summing the above inequality from $k = s$ to $N-1$, we get (38) with $\ell = 2$ and $M_2 = 0$.

Now we prove the general case $\ell \geq 2$. Since the sequence $\mathbf{w}$ satisfies (36), inequality (36) can be rewritten in terms of e^k as

$$\langle e^{k+1}, M_0 e^{k+1}\rangle \leq \sum_{i\in\mathbb{N}_\ell} \langle e^{k+1}, M_i e^{k-i+1}\rangle.$$

By using (27), this inequality can be further expressed as

$$|e^{k+1}|^2_H \leq A_1 + A_2, \tag{40}$$

where

$$A_1 := \frac{1}{\ell-1}\sum_{i\in\mathbb{N}_{\ell-1}} \langle e^{k+1}, He^{k-i+1}\rangle,\ A_2 := \left\langle e^{k+1}, M_\ell\left(e^{k+1} - \frac{2}{\ell-1}\sum_{i\in\mathbb{N}_{\ell-1}} e^{k-i+1} + e^{k-\ell+1}\right)\right\rangle.$$

Since H is in $\mathbb{S}^d$ and $r_i^{k+1} = e^{k+1} - e^{k-i+1}$, we have that

$$A_1 = \frac{1}{2}|e^{k+1}|^2_H + \frac{1}{2(\ell-1)}\sum_{i\in\mathbb{N}_{\ell-1}} (|e^{k-i+1}|^2_H - |r_i^{k+1}|^2_H). \tag{41}$$

Using relations $r_i^{k+1} = e^{k+1} - e^{k-i+1}$ and $r_{\ell-i}^{k-i+1} = e^{k-i+1} - e^{k-\ell+1}$ and the identity

$$\sum_{i\in\mathbb{N}_{\ell-1}} r_{\ell-i}^{k-i+1} = \sum_{i\in\mathbb{N}_{\ell-1}} r_i^{k+i-\ell+1},$$

we obtain that

$$e^{k+1} - \frac{2}{\ell-1}\sum_{i\in\mathbb{N}_{\ell-1}} e^{k-i+1} + e^{k-\ell+1} = \frac{1}{\ell-1}\sum_{i\in\mathbb{N}_{\ell-1}}\left(r_i^{k+1} - r_i^{k+i-\ell+1}\right). \tag{42}$$

Using equation (42) and relation $r_{\ell-i}^{k+1} = e^{k+1} - e^{k+i-\ell+1}$, we obtain that

$$A_2 = \frac{1}{\ell-1}\sum_{i\in\mathbb{N}_{\ell-1}}\left(\langle e^{k+1}, M_\ell r_i^{k+1}\rangle - \langle e^{k+i-\ell+1}, M_\ell r_i^{k+i-\ell+1}\rangle - \langle r_{\ell-i}^{k+1}, M_\ell r_i^{k+i-\ell+1}\rangle\right). \tag{43}$$

We next estimate the last inner product in equation (43). Since the set $\mathcal{M}$ of matrices satisfies Condition-M, by Lemma 4.2 and applying the Cauchy-Schwartz inequality and the Arithmetic-Geometric Mean inequality, we get for any $\alpha > 0$ that

$$|\langle r_{\ell-i}^{k+1}, M_\ell r_i^{k+i-\ell+1}\rangle| \leq \frac{\alpha}{2}|r_{\ell-i}^{k+1}|_H^2 + \frac{\|\widetilde{M}\|_2^2}{2\alpha}|r_i^{k+i-\ell+1}|_H^2. \tag{44}$$

Now, by substituting equality (41) and equality (43) incorporated with inequality (44) into inequality (40), we observe that

$$J_{k,1} + J_{k,2} + J_{k,3} \geq 0, \tag{45}$$

where

$$\begin{aligned}
J_{k,1} &= -|e^{k+1}|_H^2 + \frac{1}{\ell-1}\sum_{i\in\mathbb{N}_{\ell-1}}|e^{k-i+1}|_H^2,\\
J_{k,2} &= -\frac{1-\alpha}{\ell-1}\sum_{i\in\mathbb{N}_{\ell-1}}|r_i^{k+1}|_H^2 + \frac{\|\widetilde{M}\|_2^2}{(\ell-1)\alpha}\sum_{i\in\mathbb{N}_{\ell-1}}|r_i^{k+i-\ell+1}|_H^2,\\
J_{k,3} &= \frac{2}{\ell-1}\sum_{i\in\mathbb{N}_{\ell-1}}\left(\langle e^{k+1}, M_\ell r_i^{k+1}\rangle - \langle e^{k+i-\ell+1}, M_\ell r_i^{k+i-\ell+1}\rangle\right).
\end{aligned}$$

For $\ell \leq s \leq N-1$, by summing inequality (45) for the index k running from s to $N-1$ and using the notation $J_1 := \sum_{k=s}^{N-1} J_{k,1}$, $J_2 := \sum_{k=s}^{N-1} J_{k,2}$, and $J_3 := \sum_{k=s}^{N-1} J_{k,3}$, we obtain that

$$J_1 + J_2 + J_3 \geq 0. \tag{46}$$

We next compute J_1, J_2 and J_3. By first employing equation (35) and then exchanging the order of the resulting summations, we have that

$$J_1 = -\frac{1}{\ell-1}\sum_{i\in\mathbb{N}_{\ell-1}}(\ell-i)|e^{N-i+1}|_H^2 + \frac{1}{\ell-1}\sum_{i\in\mathbb{N}_{\ell-1}} i|e^{s-\ell+i+1}|_H^2. \tag{47}$$

For J_2, we split

$$\frac{1-\alpha}{\ell-1} = \frac{1}{\ell-1}\left(1-\alpha-\frac{1}{\alpha}\|\tilde{M}\|_2^2\right) + \frac{1}{(\ell-1)\alpha}\|\tilde{M}\|_2^2$$

and use equation (35) to conclude that

$$J_2 = -\frac{\|\widetilde{M}\|_2^2}{(\ell-1)\alpha}\sum_{i\in\mathbb{N}_{\ell-1}}\sum_{j=1}^{\ell-i}(|r_i^{N-j+1}|_H^2 - |r_i^{s-j+1}|_H^2) - \frac{1-\alpha-\|\widetilde{M}\|_2^2/\alpha}{(\ell-1)}\sum_{k=s}^{N-1}\sum_{i\in\mathbb{N}_{\ell-1}}|r_i^{k+1}|_H^2. \tag{48}$$

Likewise, we get

$$J_3 = \frac{2}{\ell-1}\sum_{i\in\mathbb{N}_{\ell-1}}\sum_{j=1}^{\ell-i}\left(\langle e^{N-j+1}, M_\ell r_i^{N-j+1}\rangle - \langle e^{s-j+1}, M_\ell r_i^{s-j+1}\rangle\right) \tag{49}$$

Moreover, the first term of (49) can be bounded above by

$$\frac{\alpha}{\ell-1}\sum_{i\in\mathbb{N}_{\ell-1}}(\ell-i)|e^{N-i+1}|_H^2 + \frac{\|\widetilde{M}\|_2^2}{(\ell-1)\alpha}\sum_{i\in\mathbb{N}_{\ell-1}}\sum_{j=1}^{\ell-i}|r_i^{N-j+1}|_H^2. \tag{50}$$

Finally, substituting (47), (48) and (49) whose first term is replaced by (50) into inequality (46), we obtain (38). □

With the help of (37) in Lemma 4.3, we have the following result, which plays an important role in the convergence analysis of sequence $\mathbf{w}$.

Lemma 4.4 *Suppose that a set $\mathcal{M}$ of $d \times d$ matrices satisfies Semi-Condition-M and $\mathcal{W}$ is a nonempty subset of $\mathbb{R}^d$. If the sequence $\mathbf{w}$ satisfies (36) for any $w \in \mathcal{W}$, then the following statements hold:*

(i) The sequence $\{|w^k|_H : k \in \mathbb{N}\}$ is bounded.

(ii) $\lim_{k\to+\infty} |w^k - w^{k-i}|_H^2 = 0$ for $i = 1, 2, \ldots, \ell - 1$.

(iii) There exists a subsequence $\{H^{\frac{1}{2}} w^{k_j} : j \in \mathbb{N}\}$ that converges to a vector $t \in \mathbb{R}^d$ with a property of $H^{\frac{1}{2}}(H^\dagger)^{\frac{1}{2}} t = t$. Furthermore, if $(H^\dagger)^{\frac{1}{2}} t \in \mathcal{W}$ then the sequence $\{H^{\frac{1}{2}} w^k : k \in \mathbb{N}\}$ itself also converges to t.

Proof: We first prove Items (i) and (ii) together. From Lemma 4.3 we observe that inequality (37) holds for all $\alpha > 0$. In particular, we choose $\alpha := \frac{1}{2}$. It follows that $1 - \alpha > 0$ and $1 - \alpha - \frac{\|\tilde{M}\|_2^2}{\alpha} > 0$ due to $\|\widetilde{M}\|_2 < \frac{1}{2}$. We fix s in inequality (37) to conclude that the sequence $\{|e^k|_H : k \in \mathbb{N}\}$ is bounded and $\sum_{k=s}^{+\infty} |r_i^{k+1}|_H^2$ are finite for $i = 1, 2, \ldots, \ell - 1$. Thus, Items (i) and (ii) follow.

We next to prove Item (iii). By Item (i), there exists a convergent subsequence $\{H^{\frac{1}{2}} w^{k_j} : j \in \mathbb{N}\}$. Suppose that $\lim_{j\to\infty} H^{\frac{1}{2}} w^{k_j} = t$. It follows that

$$\lim_{j\to\infty} H^{\frac{1}{2}} H^{\dagger\frac{1}{2}} H^{\frac{1}{2}} w^{k_j} = H^{\frac{1}{2}} H^{\dagger\frac{1}{2}} t.$$

Since $HH^\dagger H = H$, we get $H^{\frac{1}{2}} H^{\dagger\frac{1}{2}} t = t$. Therefore,

$$\lim_{j\to+\infty} |w^{k_j} - H^{\dagger\frac{1}{2}} t|_H^2 = 0. \tag{51}$$

This further yields, by using Item (ii), that for $\tau = 0, 1, \ldots, \ell - 1$

$$\lim_{j\to+\infty} |w^{k_j+\tau} - H^{\dagger\frac{1}{2}} t|_H^2 = 0. \tag{52}$$

Furthermore, if $H^{\dagger\frac{1}{2}} t \in \mathcal{W}$, then by choosing $w := H^{\dagger\frac{1}{2}} t$ in expression $e^i = w^i - w$ and choosing $s := k_j + l - 1$ in (37), we get that

$$(1-\alpha) \sum_{i\in\mathbb{N}_{\ell-1}} (\ell - i)|e^{N-i+1}|_H^2 \quad \leq \quad A_3 + A_4,$$

where

$$A_3 = \sum_{i\in\mathbb{N}_{\ell-1}} i|e^{k_j+i}|_H^2 + \frac{\|\widetilde{M}\|_2^2}{\alpha} \sum_{i\in\mathbb{N}_{\ell-1}} \sum_{\tau=1}^{\ell-i} |r_i^{\ell-\tau+k_j}|_H^2,$$

and

$$A_4 = -2 \sum_{i\in\mathbb{N}_{\ell-1}} \sum_{\tau=1}^{\ell-i} \langle e^{\ell-\tau+k_j}, M_\ell r_i^{\ell-\tau+k_j} \rangle.$$

Applying a similar technique used in deriving (44) to A_4 leads to

$$(1-\alpha)\sum_{i\in\mathbb{N}_{\ell-1}}(\ell-i)|e^{N-i+1}|_H^2 \quad\leq\quad A_3+\sum_{i\in\mathbb{N}_{\ell-1}}\sum_{\tau=1}^{\ell-i}(\alpha|e^{\ell-\tau+k_j}|_H^2+\frac{\|\widetilde{M}\|_2^2}{\alpha}|r_i^{\ell-\tau+k_j}|_H^2).$$

Using equations (51) and (52) in the above inequality, we obtain that $\lim_{N\to+\infty}|w^N - H^{\dagger\frac{1}{2}}t|_H^2 = 0$, that is, the sequence $\{H^{\frac{1}{2}}w^k : k\in\mathbb{N}\}$ converges to t. □

We are now ready to establish the main result of this section.

Theorem 4.5 *Suppose that a set $\mathcal{M}$ of $d\times d$ matrices satisfies Semi-Condition-M, the operator $T:\mathbb{R}^{\ell d}\to 2^{\mathbb{R}^d}$ is weakly firmly nonexpansive with respect to $\mathcal{M}$, the set of fixed-points of T is nonempty and* $\mathrm{dom}(T)=\mathbb{R}^{\ell d}$. *If for given vectors w^i, $i=1,2,\ldots,\ell$, the sequence $\mathbf{w}$ is generated by the iterative scheme* (24)*, then the sequence $\{H^{\frac{1}{2}}w^k : k\in\mathbb{N}\}$ converges, where $H:=M_0+M_\ell$.*

Proof: Associated with the sequence $\mathbf{w}$, we define the set $\mathcal{W}$ of all vectors $w\in\mathbb{R}^{\ell d}$ satisfying inequality (36) for all $k\geq\ell$ with $\mathbf{w}$. Since the set $\mathcal{M}$ satisfies Semi-Condition-M and the set of fixed-points to the operator T is nonempty, by the definition of the weakly firmly nonexpansive operator and (24), all fixed-points belong to $\mathcal{W}$. Therefore, the set $\mathcal{W}$ is nonempty and the sequence $\{w^k : k\in\mathbb{N}\}$ satisfies (36). Hence, by Item (iii) of Lemma 4.4, there exists a subsequence $\{H^{\frac{1}{2}}w^{k_j} : j\in\mathbb{N}\}$ that converges, say, to a vector $\widetilde{w}\in\mathbb{R}^d$. By Item (ii) of Lemma 4.4, both subsequences $\{H^{\frac{1}{2}}w^{k_j+i} : j\in\mathbb{N}\}$ and $\{H^{\frac{1}{2}}w^{k_j-i} : j\in\mathbb{N}\}$, for $i=1,2,\ldots,\ell-1$, converge to $\widetilde{w}$.

Using again the definition of the weakly firmly nonexpansive operator, we have that

$$\langle w^{k+1}-w^{k_j+1}, M_0(w^{k+1}-w^{k_j+1})\rangle \leq \sum_{i\in\mathbb{N}_\ell}\langle w^{k+1}-w^{k_j+1}, M_i(w^{k-i+1}-w^{k_j-i+1})\rangle.$$

By inclusion relation (28), we use equation (34) for each inner product in the above inequality to yield that

$$\begin{aligned}&\langle H^{\frac{1}{2}}(w^{k+1}-w^{k_j+1}), H^{\dagger\frac{1}{2}}M_0H^{\dagger\frac{1}{2}}H^{\frac{1}{2}}(w^{k+1}-w^{k_j+1})\rangle\\ \leq\quad&\sum_{i\in\mathbb{N}_\ell}\langle H^{\frac{1}{2}}(w^{k+1}-w^{k_j+1}), H^{\dagger\frac{1}{2}}M_iH^{\dagger\frac{1}{2}}H^{\frac{1}{2}}(w^{k-i+1}-w^{k_j-i+1})\rangle.\end{aligned}$$

Letting $j\to+\infty$ in the above expression, noting that $H^{\frac{1}{2}}H^{\dagger\frac{1}{2}}\widetilde{w}=\widetilde{w}$ ensured by Item (iii) of Lemma 4.4, and then using equation (34), we conclude that

$$\langle w^{k+1}-H^{\dagger\frac{1}{2}}\widetilde{w}, M_0(w^{k+1}-H^{\dagger\frac{1}{2}}\widetilde{w})\rangle \leq \sum_{i\in\mathbb{N}_\ell}\langle w^{k+1}H^{\dagger\frac{1}{2}}\widetilde{w}, M_i(w^{k-i+1}-H^{\dagger\frac{1}{2}}\widetilde{w})\rangle.$$

That is, $H^{\dagger\frac{1}{2}}\widetilde{w}$ satisfies inequality (36). Hence, $H^{\dagger\frac{1}{2}}\widetilde{w}\in\mathcal{W}$. Consequently, convergence of the sequence $\{H^{\frac{1}{2}}w^k : k\in\mathbb{N}\}$ follows from Item (iii) of Lemma 4.4. □

Finally, we show that a fixed-point of T can be found by using iterative scheme (24) if T is continuous and $\mathcal{M}$ satisfies Condition-M.

Theorem 4.6 *Suppose that a set $\mathcal{M}$ of $d\times d$ matrices satisfies Semi-Condition-M, the operator $T:\mathbb{R}^{\ell d}\to 2^{\mathbb{R}^d}$ is weakly firmly nonexpansive with respect to $\mathcal{M}$, the set of fixed-points of T is nonempty and* $\mathrm{dom}(T)=\mathbb{R}^{\ell d}$. *If the set of fixed-points of T is nonempty and T is continuous, then the sequence $\mathbf{w}$ generated from the iterative scheme* (24)*, for given vectors w^i, $i=1,2,\ldots,\ell$, converges to a fixed-point of T.*

Proof: From Theorem 4.5 we know that the sequence $\{H^{\frac{1}{2}}w^k : k \in \mathbb{N}\}$ converges. It follows that $\mathbf{w}$ converges as well due to $H \in \mathbb{S}_+^d$. Since T is continuous, the limit of the sequence $\mathbf{w}$ is a fixed-point of the operator T due to (24). □

5 Convergence Analysis of the Proposed Iterative Scheme

We return to the convergence analysis of iterative scheme (13). The main result in this section is that if the matrices M_i, $i \in \mathbb{N}_\ell$, satisfy Condition-M, the sequence $\mathbf{v} := \{v^k : k \in \mathbb{N}\}$ generated from iterative scheme (13) converges to a solution of equation (11), hence, to a solution of model (1). We shall employ the general theory established in the last section to accomplish this.

We first present a lemma regarding the operator $\mathcal{T}$.

Lemma 5.1 *Suppose that $\varphi \in \Gamma_0(\mathbb{R}^n)$, $\psi \in \Gamma_0(\mathbb{R}^m)$, $P \in \mathbb{S}_+^n$ and $Q \in \mathbb{S}_+^m$. Let $\mathcal{T}$ and E be defined by* (8) *and* (10)*, respectively, and let $R := \mathrm{diag}(P, Q)$. If a pair $(w_i, a_i) \in \mathbb{R}^{m+n} \times \mathbb{R}^{m+n}$ satisfies the following equation*

$$w_i = \mathcal{T}(Ew_i + R^{-1}a_i), \quad i = 1, 2, \tag{53}$$

then

$$\langle w_2 - w_1, a_2 - a_1 \rangle \geq 0. \tag{54}$$

Proof: By Lemma 3.1 and (10), the firm non-expansiveness of the operator $\mathcal{T}$ with respect to R ensures that

$$\|w_2 - w_1\|_R^2 \leq \langle w_2 - w_1, R(w_2 - w_1) + S_B(w_2 - w_1) + (a_2 - a_1)\rangle,$$

where S_B is defined as (9). This implies that

$$0 \leq \langle w_2 - w_1, S_B(w_2 - w_1) + (a_2 - a_1)\rangle. \tag{55}$$

On the other hand, since S_B is skewed, we have that $\langle w_2 - w_1, S_B(w_2 - w_1)\rangle = 0$. Substituting this equation into (55) proves the desired result. □

We next verify that the operator $T_\mathcal{M}$ (associated with iterative scheme (13)) defined by (22) is continuous and weakly firmly nonexpansive with respect to $\mathcal{M}$.

Lemma 5.2 *Let $\varphi \in \Gamma_0(\mathbb{R}^n)$ and $\psi \in \Gamma_0(\mathbb{R}^m)$, $P \in \mathbb{S}_+^n$ and $Q \in \mathbb{S}_+^m$. Let $\mathcal{M} := \{M_i : i \in \mathbb{N}_\ell\}$ be a set of $(m+n) \times (m+n)$ matrices. Suppose for any $u \in \mathbb{R}^{\ell(m+n)}$ there is a $w \in \mathbb{R}^{m+n}$ such that* (21) *holds. If $T_\mathcal{M}$ is defined by* (22)*, then*

(i) $T_\mathcal{M}$ is weakly firmly nonexpansive with respect to $\mathcal{M}$,

(ii) $T_\mathcal{M}$ is continuous.

Proof: We first prove Item (i). it follows from the definition of $T_\mathcal{M}$ that for any pair $(u_i, w_i) \in \mathrm{gra}(T_\mathcal{M})$, for $i = 1, 2$, there holds

$$w_i = \mathcal{T}((E - R^{-1}M_0)w_i + R^{-1}Mu_i),$$

where $M = [M_1 \quad M_2 \quad \ldots \quad M_\ell]$. By Lemma 5.1, the following inequality holds for any $(u_i, w_i) \in \mathrm{gra}(T_\mathcal{M})$ for $i = 1, 2$

$$\langle w_2 - w_1, M_0(w_2 - w_1)\rangle \leq \langle w_2 - w_1, M(u_2 - u_1)\rangle.$$

From Definition 4.1, we get Item (i).

We next prove Item (ii). From the definition of $T_{\mathcal{M}}$, for any sequence $\{(u^k, w^k) \in \text{gra}(T_{\mathcal{M}}) : k \in \mathbb{N}\}$ converging to $(u, w) \in \mathbb{R}^{\ell(m+n)} \times \mathbb{R}^{m+n}$, we have that

$$w^k = \mathcal{T}(E_0 w^k + R^{-1} M u^k),$$

where $E_0 := E - R^{-1}M_0$. Let $\tilde{w} := \mathcal{T}(E_0 w + R^{-1}Mu)$. By Lemma 3.1 we get $\mathcal{T}$ is a firmly nonexpansive operator with respect to R. Thus, for any $k \in \mathbb{N}$

$$\|w^k - \tilde{w}\|_R^2 \leq \langle w^k - \tilde{w}, RE_0(w^k - w) + M(u^k - u)\rangle.$$

We get $\lim_{k\to+\infty} w^k = \tilde{w}$ by letting $k \to +\infty$. Therefore, $w = \tilde{w}$. That is, $(u, w) \in \text{gra}(T_{\mathcal{M}})$. Thus, we prove Item (ii). □

We are now ready to prove convergence of the sequence generated from the iterative scheme (13).

Theorem 5.3 *Let $\varphi \in \Gamma_0(\mathbb{R}^n)$, $\psi \in \Gamma_0(\mathbb{R}^m)$, $P \in \mathbb{S}_+^n$ and $Q \in \mathbb{S}_+^m$. Let $\mathcal{M} := \{M_i : i \in \mathbb{N}_\ell\}$ be a set of $(m+n)\times(m+n)$ matrices, and let $\mathcal{T}$ and E be defined as* (8) *and* (10) *respectively. Suppose for any $u \in \mathbb{R}^{\ell(m+n)}$ there is a $w \in \mathbb{R}^{m+n}$ such that* (21) *holds and suppose the set of fixed points to the operator $\mathcal{T} \circ E$ is nonempty. Suppose that for given points v^i, $i = 1, 2, \ldots, \ell$, the vectors v^k, for $k \geq \ell + 1$, are generated by the iterative scheme* (13)*. Then the following two statements hold:*

(i) If $\mathcal{M}$ satisfies Semi-Condition-M, then the sequence $\{H^{\frac{1}{2}} v^k : k \in \mathbb{N}\}$ converges, where $H = M_0 + M_\ell$.

(ii) If $\mathcal{M}$ satisfies Condition-M, then the sequence $\{x^k : k \in \mathbb{N}\}$ converges to a solution of problem (1)*.*

Proof: Note that for any $u \in \mathbb{R}^{\ell(m+n)}$ there is a $w \in \mathbb{R}^{m+n}$ such that (21) holds. We let $T_{\mathcal{M}}$ be defined by (22). Thus, the sequence $\{v_k : k \in \mathbb{N}\}$ satisfies equation (13) if and only if it satisfies the inclusion relation $v^{k+1} \in T_{\mathcal{M}}(v_\ell^k)$. Because the set of fixed-points to $\mathcal{T} \circ E$ is nonempty, the set of fixed-points to $T_{\mathcal{M}}$ is nonempty by Item (i) of Semi-Condition-M. By Lemma 5.2, the operator $T_{\mathcal{M}}$ is weakly firmly nonexpansive with respect to $\mathcal{M}$. Thus, Item (i) of this theorem follows from Theorem 4.5.

Now we prove Item (ii). From Lemma 5.2 we observe that the operator $T_{\mathcal{M}}$ is weakly firmly nonexpansive with respect to $\mathcal{M}$ and is continuous. Therefore, Theorem 4.6 ensures that the sequence $\{v^k : k \in \mathbb{N}\}$ converges to a fixed-point of $T_{\mathcal{M}}$. By Item (i) of Condition-M, it converges to a fixed-point of $\mathcal{T} \circ E$. From Proposition 2.1 we get Item (ii) of this theorem. □

Theorem 5.3 shows that convergence of the iterative scheme (13) relies completely on whether the matrices $\{M_i : i \in \mathbb{N}_\ell\}$ used in the scheme (13) satisfy Condition-M (or Semi-Condition-M). This condition can be easily verified.

To close this section, we point out that our work presented in this paper differs from the recent work on convergence analysis for the sum of two convex functions developed in [21] in several aspects. The algorithm in [21] is to solve the saddle-point problem of (1) while ours directly solves the primal problem (1). The main tools utilized in [21] are variational inequality, maximally monotone operator and contraction method while the main tools in this work are the proximity operator and its non-expansiveness. Most importantly, the algorithms studied in [21] essentially have a form of (14) which is a special case of our general framework. In such a special case, the matrix M_0 is required to be positive definite in [21] while it is only required to be positive semi-definite in our current work, which will be shown in the example of the alternating split Bregman iteration in Section 8. In the next section, we shall present new algorithms whose convergence can not be derived from the result in [21], but are direct consequences of Theorem 5.3.

6 Two-Step Algorithms

In this section, we introduce two classes of two-step algorithms by choosing two specific sets of $(m+n)\times(m+n)$ matrices $\{M_0, M_1, M_2\}$ which satisfy Condition-M.

When $\ell = 2$, the general iterative scheme (13) for model (1) reduces to

$$v^{k+1} = \mathcal{T}(E_0 v^{k+1} + R^{-1}(M_1 v^k + M_2 v^{k-1})). \tag{56}$$

We shall provide two specific sets of matrices $\{M_0, M_1, M_2\}$ that lead to convergence of the iterative scheme (56).

We present two technical lemmas which will be used later.

Lemma 6.1 *Let C be an $m \times n$ real matrix and let A be an $(m+n)\times(m+n)$ matrix defined by*

$$A := \begin{bmatrix} I & C^\top \\ C & I \end{bmatrix}.$$

Then the following statements hold:

(i) Matrix $A \in \mathbb{S}_+^{m+n}$ if and only if $\|C\|_2 < 1$. If $\|C\|_2 = 1$, $A \in \mathbb{S}^{m+n}$.

(ii) If $\|C\|_2 < 1$, then A is invertible and $\|A^{-1}\|_2 = (1 - \|C\|_2)^{-1}$.

Proof: Without loss of generality we assume that $m > n$ since the other case can be similarly handled.

We first prove (i). We write C as its singular value decomposition in the form

$$C = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T,$$

where U is an $m \times m$ orthogonal matrix, Σ is an $n \times n$ diagonal matrix with nonnegative diagonal entries (the singular values of C) $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$, 0 is the $(m-n)\times n$ zero matrix, and V is an $n \times n$ orthogonal matrix. With the singular value decomposition of C, we have the corresponding decomposition of A

$$A = \begin{bmatrix} V & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} I & \Sigma & 0 \\ \Sigma & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} V^T & 0 \\ 0 & U^T \end{bmatrix}.$$

Denote by S the middle block matrix of the right-hand side of the above equation. Clearly, S and A are similar. Hence, they have the same set of eigenvalues. By utilizing the special structure of the matrix S and with the help of a permutation matrix, the matrix S is similar to

$$\widetilde{S} := \begin{bmatrix} W_1 & & & & \\ & W_2 & & & \\ & & \ddots & & \\ & & & W_n & \\ & & & & I \end{bmatrix}, \tag{57}$$

where

$$W_i = \begin{bmatrix} 1 & \sigma_i \\ \sigma_i & 1 \end{bmatrix}, \quad \text{for} \quad i = 1, 2, \ldots, n,$$

with σ_i being the i-th singular value of C. The eigenvalues of W_i are $1+\sigma_i$ and $1-\sigma_i$. Consequently, the eigenvalues of $\widetilde{S}$ are 1 with multiplicity $(m-n)$ and $1 \pm \sigma_i$, for $i = 1, 2, \ldots, n$. Since all singular values σ_i are nonnegative, the smallest eigenvalue of the matrix $\widetilde{S}$ is $1 - \sigma_1$. Noting that σ_1 is the largest singular value of C, we observe that

$$1 - \sigma_1 = 1 - \|C\|_2.$$

Clearly, $\widetilde{S} \in \mathbb{S}_+^{m+n}$. That is, $A \in \mathbb{S}_+^{m+n}$, if and only if $\|C\|_2 < 1$. Furthermore, if $\|C\|_2 = 1$, the smallest eigenvalue of A is 0. Thus, $A \in \mathbb{S}^{m+n}$.

It remains to prove (ii). From the above discussion, we know that if $\|C\|_2 < 1$, then A is in $\mathbb{S}_+^{m+n}$. Therefore, in this case, A is invertible. Moreover, we see that

$$\|A^{-1}\|_2 = \|\widetilde{S}^{-1}\|_2.$$

As it is already known, the smallest eigenvalue of the matrix $\widetilde{S}$ is $1 - \|C\|_2$. Hence, $\|A^{-1}\|_2 = (1 - \|C\|_2)^{-1}$. □

A more general result of Lemma 6.1 is given below.

Lemma 6.2 *Let $E \in \mathbb{S}_+^n$, $F \in \mathbb{S}_+^m$, and C an $m \times n$ real matrix. Let*

$$A := \begin{bmatrix} E & C^\top \\ C & F \end{bmatrix}, \quad \text{and} \quad \widetilde{C} := F^{-\frac{1}{2}} C E^{-\frac{1}{2}}.$$

Then the following statements hold:

(i) matrix $A \in \mathbb{S}_+^{m+n}$ if and only if $\|\widetilde{C}\|_2 < 1$. If $\|\widetilde{C}\|_2 = 1$, $A \in \mathbb{S}^{m+n}$.

(ii) If $\|\widetilde{C}\|_2 < 1$, then A is invertible and

$$\|A^{-1}\|_2 \le \frac{\max\{\|E^{-1}\|_2, \|F^{-1}\|_2\}}{1 - \|\widetilde{C}\|_2}.$$

Proof: We first prove (i). Set

$$D := \begin{bmatrix} E^{\frac{1}{2}} & \\ & F^{\frac{1}{2}} \end{bmatrix} \quad \text{and} \quad \widetilde{A} := \begin{bmatrix} I & \widetilde{C}^\top \\ \widetilde{C} & I \end{bmatrix}.$$

Clearly, $D \in \mathbb{S}_+^{m+n}$. With the matrices D and $\widetilde{A}$, the matrix A can be decomposed as $A = D\widetilde{A}D$. Thus, matrices A and $\widetilde{A}$ have the same set of eigenvalues. Applying Lemma 6.1 to the matrix $\widetilde{A}$ yields Item (i) of this lemma.

Next, we proceed to prove Item (ii). Notice that

$$\|D^{-1}\|_2 = \max\{\|E^{-1/2}\|_2, \|F^{-1/2}\|_2\} \quad \text{and} \quad \|\widetilde{A}^{-1}\|_2 = (1 - \|\widetilde{C}\|_2)^{-1}.$$

We have that

$$\|A^{-1}\|_2 \le \|D^{-1}\|_2^2 \|\widetilde{A}^{-1}\|_2.$$

Combining this inequality with the two equations above completes the proof of (ii). □

Now we are ready to present the two-step algorithms. We begin with constructing sets $\mathcal{M} := \{M_0, M_1, M_2\}$ of matrices that satisfies Condition-M. For given $P \in \mathbb{S}^n_+$, $Q \in \mathbb{S}^m_+$, and for an $m \times n$ real matrix B and a real number θ, we let

$$M_0 := \begin{bmatrix} P & -B^\top \\ -\theta B & Q \end{bmatrix}, \quad M_1 := \begin{bmatrix} P & (\theta - 2)B^\top \\ -\theta B & Q \end{bmatrix}, \quad \text{and} \quad M_2 := \begin{bmatrix} 0 & (1-\theta)B^\top \\ 0 & 0 \end{bmatrix}. \tag{58}$$

Proposition 6.3 *Suppose that the set $\mathcal{M}$ is chosen as in* (58)*. If*

$$|\theta| \|Q^{-\frac{1}{2}} B P^{-\frac{1}{2}}\|_2 < 1 \tag{59}$$

and

$$\frac{\max\{\|P^{-1}\|_2, \|Q^{-1}\|_2\}}{1 - |\theta| \|Q^{-\frac{1}{2}} B P^{-\frac{1}{2}}\|_2} |1-\theta| \|B\|_2 < \frac{1}{2}, \tag{60}$$

then $\mathcal{M}$ satisfies Condition-M.

Proof: We see clearly that $M_0 = M_1 + M_2$, that is, Item (i) of Condition-M holds. Item (ii) of Condition-M is trivial. Define $H := M_0 + M_2$. Then

$$H = \begin{bmatrix} P & -\theta B^\top \\ -\theta B & Q \end{bmatrix}$$

which is symmetric. In light of (59) and Lemma 6.2, we have that $H \in \mathbb{S}^{m+n}_+$. Hence, Item (iii) of Condition-M holds. Since $H \in \mathbb{S}^{m+n}_+$, it yields that $\mathcal{N}(H) = \{0\}$. Consequently, Item (iv) of Condition-M holds.

Finally, we show that Item (v) of Condition-M is valid. Using the fact of $H \in \mathbb{S}^{m+n}_+$ again, we know that $H^\dagger = H^{-1}$. We observe that

$$M_2^\top M_2 = \begin{bmatrix} 0 & \\ & (1-\theta)^2 B B^\top \end{bmatrix}.$$

Hence, $\|M_2\|_2 = |1-\theta| \|B\|_2$. By using hypotheses (60) and the above results, we have that

$$\|(H^\dagger)^{\frac{1}{2}} M_2 (H^\dagger)^{\frac{1}{2}}\|_2 = \|H^{-\frac{1}{2}} M_2 H^{-\frac{1}{2}}\|_2 < \frac{1}{2}.$$

This proves the result. □

With the choice of matrices M_0, M_1 and M_2 given in (58), the iterative scheme (56) in terms of the vectors x^k and y^k becomes

$$\left\{ \begin{array}{lcl} x^{k+1} & = & \operatorname{prox}_{\varphi, P}(x^k - P^{-1} B^\top (y^k + (1-\theta)(y^k - y^{k-1}))), \\ y^{k+1} & = & \operatorname{prox}_{\psi^*, Q}(y^k + Q^{-1} B (x^{k+1} + \theta(x^{k+1} - x^k))). \end{array} \right. \tag{61}$$

By Proposition 3.3 together with the comments that follows, we point it out that the vectors x^k and y^k can be explicitly solved from (61). Convergence of this algorithm is presented below.

Theorem 6.4 *Suppose for any given point v^1, v^2, the vectors v^k, for $k > 2$, are generated from the iterative scheme* (56) *with the choice of the set $\{M_0, M_1, M_2\}$ of $(m+n) \times (m+n)$ matrices defined as in* (58)*. Suppose the set of fixed-points to the operator $\mathcal{T} \circ E$ is nonempty. If inequalities* (59) *and* (60) *are satisfied, the sequence $\{x^k : k \in \mathbb{N}\}$ converges to a solution of problem* (1)*, where $v^k = (x^k, y^k)$ with $x^k \in \mathbb{R}^n$ and $y^k \in \mathbb{R}^m$.*

Proof: This is a direct consequence of Proposition 6.3 and Theorem 5.3. □

We consider an alternative choice of the set $\mathcal{M}$ of matrices

$$M_0 = \begin{bmatrix} P & (1-\theta)B^\top \\ B & Q \end{bmatrix}, \quad M_1 = \begin{bmatrix} P & (1-\theta)B^\top \\ (1+\theta)B & Q \end{bmatrix}, \quad \text{and} \quad M_2 = \begin{bmatrix} 0 & 0 \\ -\theta B & 0 \end{bmatrix}. \tag{62}$$

Proposition 6.5 *If*

$$|1-\theta| \|Q^{-\frac{1}{2}} B P^{-\frac{1}{2}}\|_2 < 1 \tag{63}$$

and

$$\frac{\max\{\|P^{-1}\|_2, \|Q^{-1}\|_2\}}{1 - |1-\theta| \|Q^{-\frac{1}{2}} B P^{-\frac{1}{2}}\|_2} |\theta| \|B\|_2 < \frac{1}{2}, \tag{64}$$

then the set of matrices $\{M_0, M_1, M_2\}$ *satisfies Condition-M.*

Proof: Since the proof of this proposition is similar to that of Proposition 6.3, we omit it. □

With the choice of the matrices as defined in (62), the vector v^{k+1} in (56) can be explicitly solved by the following scheme in terms of the vectors x^k and y^k

$$\begin{cases} y^{k+1} = \mathrm{prox}_{\psi^*, Q}(y^k + Q^{-1}B(x^k + \theta(x^k - x^{k-1}))), \\ x^{k+1} = \mathrm{prox}_{\varphi, P}(x^k - P^{-1}B^\top(y^{k+1} + (1-\theta)(y^{k+1} - y^k))). \end{cases} \tag{65}$$

The next result regards the convergence of the iterative algorithm (56).

Theorem 6.6 *Suppose for any given point* v^1, v^2, *the vectors* v^k, *for* $k > 2$, *are generated from the iterative scheme* (56) *with the choice of the set* $\{M_0, M_1, M_2\}$ *of* $(m+n) \times (m+n)$ *matrices defined as in* (62). *Suppose the set of fixed-points to the operator* $\mathcal{T} \circ E$ *is nonempty. If inequalities* (63) *and* (64) *are satisfied, the sequence* $\{x^k : k \in \mathbb{N}\}$ *converges to a solution of problem* (1), *where* $v^k = (x^k, y^k)$ *with* $x^k \in \mathbb{R}^n$ *and* $y^k \in \mathbb{R}^m$.

Proof: This is a direct consequence of Proposition 6.5 and Theorem 5.3. □

We remark that Algorithms (61) and (65) actually generate the same sequence in the sense that for any sequence $\mathbf{v} := \{(x^k, y^k) : k \in \mathbb{N}\}$ generated from (61), we can find some initial point such that the sequence $\mathbf{v}' := \{(x^{k+1}, y^k) : k \in \mathbb{N}\}$ generated form (65) is the same as $\mathbf{v}$, and vise versa. Thus the convergence properties of Algorithms (61) and (65) are the same. The only difference is that (61) computes x^{k+1} first while (65) updates y^{k+1} first. However, since they are connected to two different $\mathcal{M} := \{M_0, M_1, M_2\}$, we obtain different convergence conditions derived from theorem 5.3. Therefore, if the parameters θ, P and Q satisfy (59) and (60) or satisfy (63) and (64), then the sequence $\{x^k : k \in \mathbb{N}\}$ generated from (61) or (65) converges to a solution of problem (1).

We devote the remaining part of this section to a consideration of using the Gauss-Seidel strategy in improving convergence speed of the two-step algorithms introduced earlier. The basic idea of the Gauss-Seidel strategy is that instead of making use of an updated vector until all of its components are updated, an updated component should be immediately used once it becomes available. The efficiency of the Gauss-Seidel strategy for implementing proximity operator based algorithms was demonstrated through numerical experiments in [25, 27].

Whether or not the Gauss-Seidel strategy would take effect depends upon the nature of an underlying algorithm. To explain this point, we look at the iterative scheme (61). The scheme can be viewed as a block-wise Gauss-Seidel iteration in the sense that once the vector x^{k+1} is updated from x^k, y^k, and y^{k-1}, it, together with y^k and x^k, is immediately used to generate the vector y^{k+1}. However, when we update the vector x^{k+1} (or y^{k+1}), updating x^{k+1} (or y^{k+1}) in the algorithm via the Jacobi iteration may be identical to updating it via the Gauss-Seidel iteration. To explain this, let us take $\varphi(\cdot) = \|\cdot - z\|_1$ or $\varphi(\cdot) = \frac{1}{2}\|\cdot - z\|_2^2$ as an example. We choose $P := \lambda I$. In such a case, we have that

$$\mathrm{prox}_{\|\cdot - z\|_1, \lambda I} = z + \mathrm{prox}_{\frac{1}{\lambda}\|\cdot\|_1}(\cdot - z) \quad \text{and} \quad \mathrm{prox}_{\frac{1}{2}\|\cdot - z\|_2^2, \lambda I} = \frac{\lambda}{1+\lambda} \cdot + \frac{1}{1+\lambda} z,$$

where $\mathrm{prox}_{\frac{1}{\lambda}\|\cdot\|_1}$ is the well-known soft-thresholding operator. We can see that both proximity operators act on the components of the associated vector separably. Therefore, an implementation of the component-wise Gauss-Seidel iteration for the vector x^{k+1} will not lead to a speedup of convergence of the algorithm. Similar situations may happen for updating the vector y^{k+1}. We can also encounter the same situation for the iterative scheme (65).

Our goal is to reformulate (61) and (65) so that the Gauss-Seidel strategy could take effect for the resulting algorithms in the case when both P and Q are diagonal positive definite matrices and both functions φ and ψ are separable. Under these hypotheses, we observe that both proximity operator $\mathrm{prox}_{\varphi,P}$ and $\mathrm{prox}_{\psi,Q}$ are component-wise operators in the sense that the operators perform component by component for the vector on which the operators act. To see this, from Lemma 2.2 we have that

$$\mathrm{prox}_{\varphi,P} = P^{-\frac{1}{2}} \circ \mathrm{prox}_{\varphi \circ P^{-\frac{1}{2}}} \circ P^{\frac{1}{2}}.$$

Note that $\psi^* \circ Q^{-\frac{1}{2}} = (\psi \circ Q^{\frac{1}{2}})^*$. Using the Moreau decomposition $\mathrm{prox}_{(\psi \circ Q^{\frac{1}{2}})^*} = I - \mathrm{prox}_{\psi \circ Q^{\frac{1}{2}}}$ and Lemma 2.2 again, we obtain that

$$\mathrm{prox}_{\psi^*,Q} = Q^{-\frac{1}{2}} \circ (I - \mathrm{prox}_{\psi \circ Q^{\frac{1}{2}}}) \circ Q^{\frac{1}{2}}.$$

Hence, both $\mathrm{prox}_{\varphi,P}$ and $\mathrm{prox}_{\psi^*,Q}$ are component-wise operators. Because of this, for the Gauss-Seidel strategy to take effect, we should properly rewrite the vectors associated with $\mathrm{prox}_{\varphi,P}$ and $\mathrm{prox}_{\psi,Q}$ in (61) and (65).

The iterative scheme (61) may be reformulated in an equivalent form

$$\left\{ \begin{array}{rcl} x^{k+1} & = & \mathrm{prox}_{\varphi,P}((I - \frac{1-\theta}{\xi} P^{-1} B^\top B) x^k - P^{-1} B^\top (y^k - (1-\theta) d^k)), \\ b^{k+1} & = & -\frac{1}{\eta} B^\top y^k - x^{k+1} + x^k, \\ y^{k+1} & = & \mathrm{prox}_{\psi^*,Q}((I - \frac{\theta}{\eta} Q^{-1} B B^\top) y^k + Q^{-1} B (x^{k+1} - \theta b^{k+1})), \\ d^{k+1} & = & \frac{1}{\xi} B x^{k+1} - y^{k+1} + y^k, \end{array} \right. \tag{66}$$

where ξ and η are two real numbers. Once $B^\top B$ and $BB^\top$ are not diagonal matrices which is an usual case, an update from the iterative scheme (66) with the Gauss-Seidel strategy takes effect. A similar reformulation for algorithm (65) may be performed to make the Gauss-Seidel strategy take effect. A detailed implementation of Gauss-Seidel iteration for proximity operator based algorithms can be found in [27].

7 Relations with Existing Algorithms

In this section, we identify four recently developed algorithms as special cases of Algorithms (61) and (65).

7.1 The Fixed-Point Algorithm for the ROF Model

The Rudin-Osher-Fatemi (ROF) L2-TV denoising model [38] is a special case of (1) with $\varphi := \frac{1}{2}\|\cdot - z\|_2^2$, B being the first-order difference matrix, and ψ being a scaled ℓ_1-norm (for anisotropic total variation) or a certain linear combination of the norm $\|\cdot\|_2$ in $\mathbb{R}^2$. Here z is a noisy image in $\mathbb{R}^n$.

The unique solution x of the ROF model can be computed, via the proximity operator based fixed-point algorithm introduced in [27]. That is,

$$x = z - B^\top y, \tag{67}$$

where for any fixed positive number β, y is a solution of

$$y = \frac{1}{\beta}(I - \mathrm{prox}_{\beta\psi})(Bz + (\beta I - BB^\top)y). \tag{68}$$

In other words, finding the solution to the ROF model amounts to seeking a solution of equation (68). It was shown in [27] that if $\beta > \|B\|_2^2/2$, then for any initial vector y^1, the sequence $\{y^k : k \in \mathbb{N}\}$ generated by the iteration

$$y^{k+1} = \frac{1}{\beta}(I - \mathrm{prox}_{\beta\psi})(Bz + (\beta I - BB^\top)y^k) \tag{69}$$

converges to the solution of (68).

We next show how the above iterative scheme the ROF model can be rewritten in a form of (61). To this end, we need the identities

$$I - \mathrm{prox}_{\beta\psi} = \beta \mathrm{prox}_{\frac{1}{\beta}\psi^*} \circ \left(\frac{1}{\beta}I\right) \quad \text{and} \quad \mathrm{prox}_{\frac{1}{\beta}\psi^*} = \mathrm{prox}_{\psi^*,\beta I}. \tag{70}$$

These identities can be directly verified by the definition of the proximity operator and the conjugate function. Note that the first identity in (70) is the well-known Moreau decomposition [2, 29]. Using equation (70), we rewrite equation (69) as

$$y^{k+1} = \mathrm{prox}_{\psi^*,\beta I}(\frac{1}{\beta}Bz + (I - \frac{1}{\beta}BB^\top)y^k).$$

Noting that the proximity operator of $\varphi := \frac{1}{2}\|\cdot - z\|_2^2$ has the form

$$\mathrm{prox}_{\varphi,I}(\cdot) = \mathrm{prox}_{\varphi}(\cdot) = \frac{z + \cdot}{2},$$

the iterative scheme (69) is algebraically equivalent to

$$\begin{cases} x^{k+1} = \mathrm{prox}_{\varphi,I}(x^k - B^\top y^k), \\ y^{k+1} = \mathrm{prox}_{\psi^*,\beta I}(y^k + \frac{1}{\beta}B(2x^{k+1} - x^k)). \end{cases} \tag{71}$$

Clearly, iterative scheme (71) is a special case of algorithm (61) with

$$P = I, \quad Q = \beta I \quad \text{and} \quad \theta = 1. \tag{72}$$

We next present a theorem regarding convergence of algorithm (71).

Theorem 7.1 *Let $\{v^k : k \in \mathbb{N}\}$ with $v^k = (x^k, y^k) \in \mathbb{R}^n \times \mathbb{R}^m$ be a sequence generated from (71) for the ROF denoising model. If $\beta > \|B\|_2^2$ then the sequence $\{x^k : k \in \mathbb{N}\}$ converges to the solution of the model.*

Proof: Since (71) is a special case of (61) with P, Q and θ defined by (72), it follows from Theorem 6.4. □

Note that two algebraically equivalent algorithms may not be algorithmically equivalent. Although two algorithms (69) and (71) are algebraically equivalent, from the numerical implementation viewpoint, they are different. Scheme (69) allows us to explore the Gauss-Seidel method for accelerating the rate of convergence of the algorithm, which was verified numerically in [27], but algorithm (71) does not. Algorithmically, the scheme (69) is preferable for the ROF denoising model. Moreover, the range $\beta > \|B\|_2^2/2$ guaranteed by the convergence result in [27] for the scheme (69) to converge is wider than the range $\beta > \|B\|_2^2$ guaranteed by Theorem 7.1 for the scheme (71) to converge. This is because that in [27] the strong convexity of *varphi* was used to obtain a stronger result and Theorem 7.1 is obtained by using Theorem 6.4 which requires only $\varphi \in \Gamma_0$, without the strong convexity assumption on φ.

7.2 The Proximity Algorithm for the L1-TV Denoising Model

The L1-TV denoising model [12, 32] is another special case of (1) with $\varphi := \|\cdot - z\|_1$, B being the first-order difference matrix, and ψ being a scaled ℓ_1-norm (for anisotropic total variation) or a certain linear combination of the norm $\|\cdot\|_2$ in $\mathbb{R}^2$. Here z is a noisy image in $\mathbb{R}^n$. Unlike the ROF denoising model, solutions of the L1-TV denoising model are not unique. The model has attracted considerable attention due to its various attractive mathematical properties [12, 32].

A proximity operator based algorithm for the L1-TV denoising model proposed in [25] has been verified numerically to be efficient in removing impulsive noise. This algorithm can be described as follows:

$$\begin{cases} x^{k+1} = \mathrm{prox}_{\frac{1}{\lambda}\varphi}((I - \frac{1}{\lambda\beta}B^\top B)x^k - \frac{1}{\lambda\beta}B^\top(b^k - d^k)), \\ d^{k+1} = \mathrm{prox}_{\beta\psi}(b^k + Bx^{k+1}), \\ b^{k+1} = Bx^{k+1} + b^k - d^{k+1}. \end{cases} \tag{73}$$

From the last two equations of (73), we have that

$$b^{k+1} = (I - \mathrm{prox}_{\beta\psi})(b^k + Bx^{k+1}). \tag{74}$$

That is, the order in computing d^{k+1} and b^{k+1} is commutable. Therefore, the iterative scheme generates vectors of the sequences $\{x^k : k \in \mathbb{N}\}$, $\{b^k : k \in \mathbb{N}\}$, $\{d^k : k \in \mathbb{N}\}$, alternatively, in the order

$$\cdots \Rightarrow x^{k+1} \Rightarrow b^{k+1} \Rightarrow d^{k+1} \Rightarrow x^{k+2} \Rightarrow b^{k+2} \Rightarrow d^{k+3} \Rightarrow \cdots$$

By using the first identity of (70) and equation (74), and setting $y^k := \frac{1}{\beta}b^k$, we may rewrite iterative scheme (73) as

$$\begin{cases} y^{k+1} = \mathrm{prox}_{\frac{1}{\beta}\psi^*}(y^k + \frac{1}{\beta}Bx^k), \\ d^{k+1} = Bx^k + \beta y^k - \beta y^{k+1}, \\ x^{k+1} = \mathrm{prox}_{\frac{1}{\lambda}\varphi}((I - \frac{1}{\lambda\beta}B^\top B)x^k - \frac{1}{\lambda\beta}B^\top(\beta y^{k+1} - d^{k+1})). \end{cases} \tag{75}$$

Eliminating the variable d^{k+1} in scheme (75) and using the second identity of (70), we rewrite (73) in its algebraically equivalent form

$$\begin{cases} y^{k+1} = \operatorname{prox}_{\psi^*,\beta I}(y^k + \frac{1}{\beta}Bx^k), \\ x^{k+1} = \operatorname{prox}_{\varphi,\lambda I}(x^k - \frac{1}{\lambda}B^\top(2y^{k+1} - y^k)). \end{cases} \tag{76}$$

This is a special case of (65) with

$$P = \lambda I, \quad Q = \beta I, \quad \text{and} \quad \theta = 0. \tag{77}$$

In the next theorem we present the convergence result of algorithm (76) for the L1-TV denoising model.

Theorem 7.2 *Let $\{x^k : k \in \mathbb{N}\}$ be a sequence generated from (76) for the L1-TV denoising model. If $\lambda\beta > \|B\|_2^2$, then the sequence $\{x^k : k \in \mathbb{N}\}$ converges to a solution of the model.*

Proof: Since algorithm (76) is a special case of (65) with P, Q and θ given by (77), the result of this theorem follows directly from Theorem 6.6. □

A convergence result for algorithm (73) having the same range for the parameter β as in the above theorem was given in [25].

7.3 The First-Order Primal-Dual Algorithm

A first-order primal-dual algorithm for problem (1) was introduced in [9] recently based on the saddle point formulation of the model. It can be described as

$$\begin{cases} x^{k+1} = \operatorname{prox}_{\varphi,\lambda I}(x^k - \frac{1}{\lambda}B^\top y^k), \\ y^{k+1} = \operatorname{prox}_{\psi^*,\beta I}(y^k + \frac{1}{\beta}B(2x^{k+1} - x^k)), \end{cases} \tag{78}$$

where λ, β are two positive numbers. It was proved in [9] that if $\lambda\beta \geq \|B\|_2^2$ then the sequence $\{x^k : k \in \mathbb{N}\}$ generated from algorithm (78) converges to a solution of model (1). Algorithm (78) may be identified as a special case of Algorithm (61) with

$$P = \lambda I, \quad Q = \beta I, \quad \text{and} \quad \theta = 1. \tag{79}$$

The following convergence result for Algorithm (78) is a direct consequence of Theorem 6.4.

Theorem 7.3 *Let $\{v^k : k \in \mathbb{N}\}$ with $v^k = (x^k, y^k) \in \mathbb{R}^n \times \mathbb{R}^m$ be a sequence generated from (78) for model (1). If $\lambda\beta > \|B\|_2^2$, then the sequence $\{x^k : k \in \mathbb{N}\}$ converges to a solution of the model.*

Theorem 7.3 was obtained in [9] from a different approach.

7.4 The Preconditioned First-Order Primal-Dual Algorithm

The first-order primal-dual algorithm (for model (1)) discussed in the previous subsection was recently extended in [34] to a *preconditioned* first-order primal-dual algorithm. Employing the notation used in this paper, the algorithm can be described by replacing the diagonal matrices λI

and βI in the first-order primal-dual algorithm with symmetric positive definite matrices P and Q, respectively. More precisely, it has the form

$$\begin{cases} x^{k+1} = \mathrm{prox}_{\varphi,P}(x^k - P^{-1}B^\top y^k), \\ y^{k+1} = \mathrm{prox}_{\psi^*,Q}(y^k + Q^{-1}B(2x^{k+1} - x^k)). \end{cases} \tag{80}$$

Clearly, algorithm (80) is a special case of algorithm (61) with $\theta = 1$.

The following convergence result for Algorithm (80) is a direct consequence of Theorem 6.4.

Theorem 7.4 *Let $\{v^k : k \in \mathbb{N}\}$ with $v^k = (x^k, y^k) \in \mathbb{R}^n \times \mathbb{R}^m$ be a sequence generated from (80) for model (1). If P and Q are two matrices in $\mathbb{S}_+^{m+n}$ such that $\|Q^{-\frac{1}{2}}BP^{-\frac{1}{2}}\|_2 < 1$ then the sequence $\{x^k : k \in \mathbb{N}\}$ converges to a solution of the model.*

Theorem 7.4 was obtained in [34] from a different approach.

8 An Improved Alternating Split Bregman Iteration

We develop in this section an improved alternating split Bregman iteration method from the general algorithm (14) and consider its convergence property. For this purpose, we first identify the standard alternating split Bregman iteration (ASBI) introduced in [20] as a special case of (14) with two specific matrices, noting that ASBI has been successfully applied in image processing [6, 26, 33, 43]. We then show from our general setting that the dual variable sequence generated from ASBI converges but the primal variable sequence converges *only* with respect to a symmetric semi-positive definite matrix. Convergence of the primal variable sequence generated from ASBI is not guaranteed. Motivated from this analysis, we propose an improved ASBI aiming at convergence of the resulting primal variable sequence.

We first identify ASBI as a special case of the general algorithm (14). To this end, we reformulation model (1) as an equivalent constrained optimization problem

$$\min\{\varphi(x) + \psi(d) : x \in \mathbb{R}^n,\ d \in \mathbb{R}^m,\ Bx - d = 0\}. \tag{81}$$

The split Bregman iteration for solving this constrained problem is as follows:

$$\begin{cases} d^{k+1} &= \arg\min\{\psi(d) + \frac{\nu}{2}\|d - Bx^k - b^k\|_2^2 : d \in \mathbb{R}^m\}, \\ b^{k+1} &= b^k + Bx^k - d^{k+1}, \\ x^{k+1} &= \arg\min\{\varphi(x) + \frac{\nu}{2}\|d^{k+1} - Bx - b^{k+1}\|_2^2 : x \in \mathbb{R}^n\}. \end{cases} \tag{82}$$

By the first-order optimality condition in convex analysis, ASBI (82) is equivalent to

$$\begin{cases} 0 \in \partial\psi(d^{k+1}) + \nu(d^{k+1} - Bx^k - b^k), \\ d^{k+1} = Bx^k + b^k - b^{k+1}, \\ 0 \in \partial\varphi(x^{k+1}) + \nu B^\top(Bx^{k+1} + b^{k+1} - d^{k+1}). \end{cases} \tag{83}$$

Using formula (3) of the conjugate function, the first inclusion of (83) can be written as

$$\nu d^{k+1} \in \partial(\nu\psi^*)(\nu(Bx^k + b^k - d^{k+1})). \tag{84}$$

Set $y^k := \nu b^k$. Clearly, we have that $\nu d^{k+1} + \nu(Bx^k + b^k - d^{k+1}) = \nu Bx^k + y^k$. By multiplying the second equation in expression (83) with ν, we have that $\nu(Bx^k + b^k - d^{k+1}) = y^{k+1}$. Hence,

$$\nu B^\top(Bx^{k+1} + b^{k+1} - d^{k+1}) \quad = \quad \nu B^\top B(x^{k+1} - x^k) + B^\top(2y^{k+1} - y^k). \tag{85}$$

Applying equation (2) to inclusion (84) and the second inclusion in (83) together with equation (85) yields a reformulation of (82) as

$$\begin{cases} y^{k+1} = \operatorname{prox}_{\nu\psi^*}(\nu B x^k + y^k), \\ x^{k+1} = \operatorname{prox}_{\varphi}(x^{k+1} - \nu B^\top B(x^{k+1} - x^k) - B^\top(2y^{k+1} - y^k)). \end{cases} \tag{86}$$

Scheme (86) is in a form of (14) with $v^k := (x^k, y^k) \in \mathbb{R}^n \times \mathbb{R}^m$ and with the choices

$$P := I, \quad Q := \frac{1}{\nu} I, \quad \text{and} \quad M_0 = M_1 := \begin{bmatrix} \nu B^\top B & B^\top \\ B & \frac{1}{\nu} I \end{bmatrix}. \tag{87}$$

Noticing that $M_0 = \frac{1}{\nu} \begin{bmatrix} \nu B & I \end{bmatrix}^\top \begin{bmatrix} \nu B & I \end{bmatrix}$, M_0 is in $\mathbb{S}^{m+n}$, but not in $\mathbb{S}_+^{m+n}$. It can be verified that when $\ell = 2$ and $M_2 = 0$, the set $\mathcal{M} := \{M_0, M_1, M_2\}$ of matrices satisfies Condition-M (resp. Semi-Condition-M) if and only if $M_0 = M_1$ and $M_0 \in \mathbb{S}_+^{m+n}$ (resp. $M_0 \in \mathbb{S}^{m+n}$). By Theorem 5.3, the sequence $\{M_0^{\frac{1}{2}} v^k : k \in \mathbb{N}\}$ converges.

Although convergence of the primal variable sequence $\mathbf{x} := \{x^k : k \in \mathbb{N}\}$ to a solution of problem (1) is not guaranteed by this theorem, it is proved in [39] that the dual variable sequence $\mathbf{y} := \{y^k : k \in \mathbb{N}\}$ converges to a solution of the dual problem of (1) by reformulating ASBI as the Douglas-Rachford splitting method. In the next theorem, we obtain the same convergence result from the general convergence theorem of this paper, without using the Douglas-Rachford splitting approach.

Theorem 8.1 *Suppose that problem* (1) *has a solution. Let* $\mathbf{v} := \{(x^k, y^k) : k \in \mathbb{N}\}$ *be generated from* (86)*. Then the following results hold:*

(i) The sequence $B\mathbf{x} := \{Bx^k : k \in \mathbb{N}\}$ *converges.*

(ii) The sequence $\mathbf{y}$ *converges to a solution of the dual problem*

$$\min\{\psi^*(y) + \varphi^*(-B^\top y) : y \in \mathbb{R}^m\} \tag{88}$$

of model (1)*.*

Proof: This proof consists of two parts. The first part is to show that both sequences $\mathbf{y}$ and $B\mathbf{x}$ converge and the second part is to prove that the limit of sequence $\mathbf{y}$ is a solution of model (88).

Since ASBI (86) is identified as a form of the iterative scheme (14) with P, Q, M_0, M_1 defined by (87) and $M_0 \in \mathbb{S}^{m+n}$, by Theorem 5.3 the sequence $M_0^{\frac{1}{2}}\mathbf{v}$ converges, and so does the sequence $M_0\mathbf{v}$. We take the last m rows of M_0 to form a new matrix L, that is, $L = \begin{bmatrix} B & \frac{1}{\nu} I \end{bmatrix}$. Then, the sequence $L\mathbf{v}$ also converges.

By Proposition 2.1, the existence of solutions to problem (1) ensures that there exists a vector $v_\star = (x_\star, y_\star) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying

$$x_\star = \operatorname{prox}_{\varphi}(x_\star - B^\top y_\star) \quad \text{and} \quad y_\star = \operatorname{prox}_{\nu\psi^*}(\nu B x_\star + y_\star). \tag{89}$$

From the first equation in (86) and the second equation in (89), the non-expansiveness of the proximity operator leads to $\|y^{k+1} - y_\star\|_2 \le \nu \|L(v^k - v_\star)\|_2$. This inequality with the convergence of sequence $L\mathbf{v}$ implies the boundedness of the sequence $\mathbf{y}$. Furthermore, since $Lv^k = Bx^k + \frac{1}{\nu} y^k$, it follows that sequence $B\mathbf{x}$ is also bounded. Therefore, there exist two subsequences $\{Bx^{k_j} : j \in \mathbb{N}\}$ and $\{y^{k_j} : j \in \mathbb{N}\}$ that converge to $\widetilde{z} \in \mathbb{R}^m$ and $\widetilde{y} \in \mathbb{R}^m$, respectively. This, in turn, implies that

$$\lim_{k \to +\infty} Lv^k = \widetilde{z} + \frac{1}{\nu} \widetilde{y}. \tag{90}$$

Choosing $k := k_j - 1$ in the first equation of (86), we get by using (90) that as $j \to \infty$,

$$\widetilde{y} = \operatorname{prox}_{\nu\psi^*}(\nu\widetilde{z} + \widetilde{y}). \tag{91}$$

From the first equation in (86) and equation (91), the non-expansiveness of the proximity operator leads to

$$\|y^{k+1} - \widetilde{y}\|_2 \leq \nu \left\| Lv^k - (\widetilde{z} + \frac{1}{\nu}\widetilde{y}) \right\|_2 .$$

From (90), we have that

$$\lim_{k\to+\infty} y^k = \widetilde{y} \quad \text{and} \quad \lim_{k\to+\infty} Bx^k = \lim_{k\to+\infty} \left(Lv^k - \frac{1}{\nu} y^k \right) = \widetilde{z}. \tag{92}$$

This completes the first part of the proof and we get Item (i).

Next we prove the second part. Applying equation (2) to the second equation of (86) and employing relation (3) give

$$x^{k+1} \in \partial\varphi^*(-B^\top(\nu Bx^{k+1} - \nu Bx^k + 2y^{k+1} - y^k)).$$

Multiplying the both sides of the above inclusion relation by the matrix B and then applying the chain rule of subdifferential to its right-hand side of the resulting relation give that

$$-Bx^{k+1} \in \partial(\varphi^* \circ (-B^\top))(\nu Bx^{k+1} - \nu Bx^k + 2y^{k+1} - y^k).$$

In the above inclusion letting $k \to \infty$ and using (92), we have that $-\widetilde{z} \in \partial(\varphi^* \circ (-B^\top))(\widetilde{y})$. This together with equation (91) (which is equivalent to $\widetilde{z} \in \partial\psi^*(\widetilde{y})$) yields $0 \in \partial(\varphi^* \circ (-B^\top))(\widetilde{y}) + \partial\psi^*(\widetilde{y})$. That is, $\widetilde{y}$ is a solution of problem (88). □

For ASBI, we can only show that the sequence $B\mathbf{x}$ converges and the sequence $\mathbf{y}$ converges to a solution of the associated dual problem. Theorem 8.1 cannot guarantee the convergence of the primal variable sequence $\mathbf{x}$ to a solution of problem (1). We next use Condition-M as a guide to remedy ASBI so that the sequence $\mathbf{x}$ generated from the improved ASBI can converge to a solution of problem (1).

We examine the matrix M_0 in (87). This matrix is singular no matter whether the matrix B is singular or not. To ensure the sequence $\mathbf{x}$ converging to a solution of problem (1), based on Theorem 5.3, it suffices to replace the matrix M_0 in ASBI by its perturbation in $\mathbb{S}_+^{m+n}$. This motivates us to choose

$$M_0 = \begin{bmatrix} \nu\theta B^\top B + \alpha(1-\theta)I & B^\top \\ B & \frac{1}{\nu}I \end{bmatrix}, \tag{93}$$

where α is a positive number and θ is number number between 0 and 1. This modification in the choice of matrix M_0 corresponds to changing the third equation in ASBI for updating x^{k+1} in (82) by the following

$$x^{k+1} = \arg\min \left\{ \varphi(x) + \frac{\nu\theta}{2}\|d^{k+1} - Bx - b^{k+1}\|_2^2 + \frac{\alpha(1-\theta)}{2}\|x - x^k\|_2^2 : x \in \mathbb{R}^n \right\}.$$

Adding the quadrature term $\frac{\alpha}{2}\|x - x^k\|_2^2$ to the cost function of the optimization problem actually imposes an additional condition that the update x^{k+1} should not deviate from the current x^k too much. With this new M_0 in (93), the improved ASBI is

$$\begin{cases} y^{k+1} = \operatorname{prox}_{\nu\psi^*}(\nu Bx^k + y^k), \\ x^{k+1} = \operatorname{prox}_{\varphi}(x^{k+1} - (\nu\theta B^\top B + \alpha(1-\theta)I)(x^{k+1} - x^k) - B^\top(2y^{k+1} - y^k)). \end{cases} \tag{94}$$

Note that when $\theta = 1$, the improved ASBI reduces to ASBI.

We prove below the main result of this section that when $0 \le \theta < 1$, the primal variable sequence $\mathbf{x}$ generated from the improved ASBI converges to a solution of problem (1).

Theorem 8.2 *Suppose that problem* (1) *has a solution and* $\theta \in [0, 1)$. *If the parameters* α *and* ν *are positive and satisfy the condition*

$$\frac{\alpha}{\nu} > \|B\|_2^2, \tag{95}$$

then the sequence $\mathbf{x}$ *generated from* (94) *converges to a solution of problem* (1).

Proof: We prove this theorem by employing Theorem 5.3. Since the iterative scheme (94) can be viewed as a special case of (14) with $P = I$, $Q = \frac{1}{\nu}I$, M_0 and M_1 given by (93), if we can show that M_0 is in $\mathbb{S}_+^{m+n}$, then, Theorem 5.3 ensures that the sequence $\mathbf{x}$ generated from (94) converges to a solution of problem (1). It suffices to show $M_0 \in \mathbb{S}_+^{m+n}$. By Item 1 of Lemma 6.2 it remains to prove that if α and ν satisfy inequality (95) and $\theta \in [0, 1)$, then

$$\sqrt{\nu}\|B(\nu\theta B^\top B + \alpha(1-\theta)I)^{-\frac{1}{2}}\|_2 < 1. \tag{96}$$

We next proceed the proof of (96). We first assume that $n < m$. The singular value decomposition of B gives a factorization of the form

$$B^\top = V \begin{bmatrix}\Sigma & 0\end{bmatrix} U^\top,$$

where U is the $m \times m$ orthogonal matrix, the matrix Σ is an $n \times n$ diagonal matrix with nonnegative numbers (the singular values of B) $\|B\|_2 = \sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_n$ on the diagonal, 0 at the right-hand side of the decomposition is the $n \times (m-n)$ zero matrix, and V is the $n \times n$ orthogonal matrix. This yields that

$$B(\nu\theta B^\top B + \alpha(1-\theta)I)^{-\frac{1}{2}} = U \begin{bmatrix}\Sigma(\nu\theta\Sigma^2 + \alpha(1-\theta)I)^{-\frac{1}{2}} \\ 0\end{bmatrix} V^\top.$$

This ensures that the singular values of $B(\nu\theta B^\top B + \alpha(1-\theta)I)^{-\frac{1}{2}}$ are given by

$$\frac{\sigma_i}{(\nu\theta\sigma_i^2 + \alpha(1-\theta))^{1/2}} \quad \text{for} \quad i = 1, 2, \ldots, n$$

with

$$\frac{\sigma_1}{(\nu\theta\sigma_1^2 + \alpha(1-\theta))^{1/2}} = \frac{\|B\|_2}{(\nu\theta\|B\|_2^2 + \alpha(1-\theta))^{1/2}}$$

as its largest one. We then conclude that

$$\sqrt{\nu}\|B(\nu\theta B^\top B + \alpha(1-\theta)I)^{-\frac{1}{2}}\|_2 = \frac{\sqrt{\nu}\|B\|_2}{\sqrt{\nu\theta\|B\|_2^2 + \alpha(1-\theta)}}$$

is strictly less than 1 if $\theta \in [0, 1)$ and positive numbers α and ν satisfy (95). That is, condition (96) holds for the case $n < m$. The case $n \ge m$ may be handled similarly and the proof is complete. □

We remark that when $\theta = 1$ which corresponds to ASBI, Theorem 8.2 fails to hold.

9 Numerical Examples

In this section, we take the L1-TV denoising model as an example to demonstrate the computational performance of the two-step algorithms. We shall mainly compare numerical results of algorithm (66) with those of the Chambolle and Pock algorithm (Algorithm CP) [9]. All the numerical experiments are performed under Windows 7 and MATLAB R2008a running on a PC equipped with an Intel Core 2 Quad CPU at 2.12 GHz and 2G RAM memory(using a C code).

It is helpful to determine the parameters ξ and η in algorithm (66) when the specific form of the L1-TV image denoising model is presented. To this end, we begin with giving explicit expressions of ψ and B according to the definition of the total-variation [38]. The concrete expressions of ψ and B depend on how the two dimensional images are vectorized. For convenience of exposition, we assume that an image considered in this paper has a size of $p \times q$. The image is treated as a vector in $\mathbb{R}^{pq}$ in such a way that the ij-th pixel of the image corresponds to the $(i + (j-1)p)$-th component of the vector in $\mathbb{R}^{pq}$. We set $n := pq$ and $m := 2n$. To define the $m \times n$ matrix B, we recall the $d \times d$ difference matrix D_d by

$$D_d := \begin{bmatrix} 0 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}. \tag{97}$$

Now, through the matrix Kronecker product $\otimes$, we define the $m \times n$ matrix B by

$$B := \begin{bmatrix} I_q \otimes D_p \\ D_q \otimes I_p \end{bmatrix}. \tag{98}$$

The convex function $\psi : \mathbb{R}^m \to \mathbb{R}$ is defined at $y \in \mathbb{R}^m$ as

$$\psi(y) := \sum_{i=1}^{n} \left\| [y_i, y_{n+i}]^\top \right\|_2, \tag{99}$$

with the corresponding $\psi(Bx)$ being referred to as (isotropic) total-variation of x. With these notation, the L1-TV denoising model is to minimize the cost function

$$E(\cdot) = \lambda \| \cdot - z\|_1 + \psi(B\cdot),$$

where λ is a regularized parameter and z is the noisy image to be denoised.

We choose

$$P := \alpha I, \quad Q := \beta I, \quad \alpha := 0.1\lambda, \quad \text{and} \quad \beta := \frac{4}{\alpha}.$$

the parameters for algorithm (66). Next, we propose a way to select the parameters ξ and η. Note that in the first step of the proposed algorithm, x^k is multiplied by the matrix $I - \frac{1-\theta}{\alpha\xi} B^\top B$ from the left. As a result, if we choose θ, α, and ξ, such that $\frac{1-\theta}{\alpha\xi} \le \frac{2}{\|B\|^2}$, then the spectral radius of the matrix $I - \frac{1-\theta}{\alpha} B^\top B$ is less than one. Similarly, if we choose θ, β, and η such that $\frac{\theta}{\beta\eta} \le \frac{2}{\|B\|^2}$, then the spectral radius of the matrix $I - \frac{\theta}{\beta\eta} BB^\top$, which multiplies y^k from the left in the third step of our algorithm, is less than one. Since the matrix B in the L1-TV model is the first order difference matrix, the L2 norm of B is presented in the next lemma.

Lemma 9.1 *If B is the matrix defined by* (98) *for $p, q \in \mathbb{N}$, then*

$$\|B\|_2^2 = 4 \sin^2 \frac{(p^2-1)\pi}{2p^2} + 4 \sin^2 \frac{(q^2-1)\pi}{2q^2}.$$

Proof: The key idea in this proof is to write

$$B^\top B = I_q \otimes D_p^\top D_p + D_q^\top D_q \otimes I_p$$

and to present the eigenvalues of $B^\top B$ in terms of the sum of the eigenvalues of $D_p^\top D_p$ and $D_q^\top D_q$, where D_d is the matrix defined by (97).

Since $D_p^\top D_p$ and $D_q^\top D_q$ are symmetric, there exist orthonormal matrices F_p and F_q such that

$$D_p^\top D_p = F_p \Lambda_p F_p^\top \quad \text{and} \quad D_q^\top D_q = F_q \Lambda_q F_q^\top,$$

where Λ_p and Λ_q are diagonal matrices whose diagonal entries are eigenvalues of $D_p^\top D_p$ and $D_q^\top D_q$ respectively. Therefore, we have that

$$(F_q^\top \otimes F_p^\top) B^\top B (F_q \otimes F_p) = I_q \otimes \Lambda_p + \Lambda_q \otimes I_p.$$

Since $F_q^\top \otimes F_p^\top$ is orthonormal, the eigenvalues of $B^\top B$ are the sums of the eigenvalues of $D_p^\top D_p$ and $D_q^\top D_q$. Hence, $\|B\|_2^2 = \|D_p\|_2^2 + \|D_q\|_2^2$. According to [27], the largest eigenvalue of $D_d^\top D_d$ is $4\sin^2 \frac{(d^2-1)\pi}{2d^2}$. This completes the proof of the lemma. □

Lemma 9.1 ensures that $\|B\|_2^2 \leq 8$. Accordingly, we choose ξ and η as

$$\xi := \begin{cases} \frac{4}{\alpha}, & \theta = 0, \\ \frac{4(1-\theta)}{\alpha}, & 0 < \theta < 1, \\ \text{arbitary}, & \theta = 1. \end{cases} \quad \text{and} \quad \eta := \begin{cases} \text{arbitary}, & \theta = 0, \\ \frac{8\theta}{\beta}, & 0 < \theta < 1, \\ \frac{4}{\beta}, & \theta = 1. \end{cases}$$

To evaluate the performance of the tested algorithms, we first compute the minimal value of E. Therefore, we iterate Algorithm CP 100000 times to obtain a "true" solution $\tilde{x}$ of the L1-TV model and estimate the minimal value of E by $E(\tilde{x})$. All algorithms tested in the experiments are carried out until the stopping criteria

$$\frac{|E(x^k) - E(\tilde{x})|}{|E(\tilde{x})|} < 10^{-4}$$

is satisfied. The quality of a denoised image x^∞ is evaluated in terms of the peak signal-to-noise ratio (PSNR) defined by

$$\text{PSNR} = 10 \log_{10} \frac{255^2 n}{\|x^\infty - x\|^2} (dB),$$

where x is the original image with a total of n pixels. Each PSNR-value reported in this section is the average over five runs.

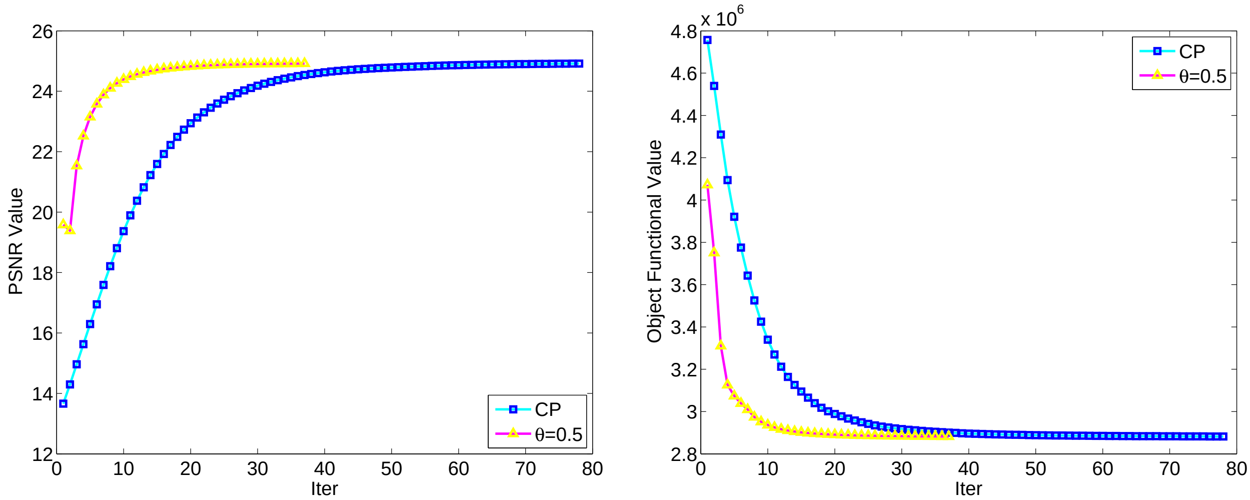
Three test images are displayed in Figure 9.1. They are the images "Cameraman" and "Lena" of size 256×256 and the image "Plane" of size 320×400. We explore how the algorithms perform when the level of impulsive noise added to the test images changes. The numerical results obtained from these algorithms with noise level 10%, 20%, 30%, and 40% are reported in Table I. We find that when the parameter $\theta \leq 0.5$, the proposed algorithm uses much less CPU-time than Algorithm CP does. Furthermore, for the image "Cameraman" corrupted with impulsive noise at level 30%, the changes of the PSNR values of x^k and the objective function values at x^k as k increases are illustrated in Figure 9.2 for Algorithm CP and the proposed algorithm. It can be seen from the plots that when $\theta \leq 0.5$, the proposed algorithm converges significantly faster than Algorithm CP.

(a) (b) (c)

Figure 9.1: Original images. (a) "Cameraman"; (b) "Lena"; and (c) "Airplane".

Table I: The summary of the restoration results

		Cameraman		Lena		Airplane	
Level	Method	PSNR (dB)	Time (seconds)	PSNR (dB)	Time (seconds)	PSNR (dB)	Time (seconds)
10%	CP	35.41	11.09	34.19	15.24	38.35	11.59
	$\theta = 0.0$	35.74	4.64	35.07	4.97	38.37	6.77
	$\theta = 0.1$	35.72	4.74	35.04	5.08	38.37	6.22
	$\theta = 0.5$	35.64	4.99	34.82	5.58	38.36	6.15
	$\theta = 0.9$	35.58	5.46	34.66	6.27	38.36	6.21
	$\theta = 1.0$	35.33	11.19	34.16	17.13	38.35	6.41
20%	CP	32.96	15.02	31.36	15.31	36.18	22.54
	$\theta = 0.0$	33.02	8.24	31.67	6.90	36.18	12.72
	$\theta = 0.1$	33.02	8.08	31.66	6.99	36.18	11.71
	$\theta = 0.5$	32.99	8.21	31.55	7.48	36.18	11.83
	$\theta = 0.9$	32.97	8.46	31.48	8.00	36.18	12.03
	$\theta = 1.0$	32.90	10.48	31.29	15.83	36.19	11.84
30%	CP	31.06	19.7	30.16	21.17	34.35	24.20
	$\theta = 0.0$	31.07	10.78	30.18	11.94	34.35	13.37
	$\theta = 0.1$	31.07	10.37	30.18	11.52	34.35	12.35
	$\theta = 0.5$	31.06	10.74	30.17	11.83	34.3	12.61
	$\theta = 0.9$	31.05	11.11	30.16	12.02	34.35	12.94
	$\theta = 1.0$	31.02	12.45	30.09	14.85	34.35	12.75
40%	CP	29.21	22.59	27.96	30.17	32.49	27.38
	$\theta = 0.0$	29.22	12.35	27.96	16.92	32.49	14.51
	$\theta = 0.1$	29.22	11.74	27.96	16.01	32.49	14.33
	$\theta = 0.5$	29.21	12.18	27.96	16.30	32.49	14.00
	$\theta = 0.9$	29.21	12.57	27.96	16.88	32.49	15.93
	$\theta = 1.0$	29.18	13.91	27.93	18.68	32.49	14.60



(a) (b)

Figure 9.2: The plots of (a) PSNR values versus the number of iterations and (b) the values of the objective function of the L1/TV model versus the number of iterations for the image "Cameraman" corrupted with 30% impulsive noise.

10 Conclusions

In this paper, we study an optimization problem, which appears in image processing, whose objective function is the sum of two convex functions with one composed by an affine transformation. We propose a class of multi-step proximity algorithms for the problem based on the characterization of its solution. Many existing algorithms can be identified as special cases of the proposed algorithms. The notion of weakly firmly non-expansive operators is introduced as a mathematical tool for analyzing convergence of the algorithms. We provide an unified approach for convergence analysis of the new algorithms and as well as existing algorithms. Convergence analysis of the algorithms can be carried out by verifying conditions on the matrices associated with the algorithms. In particular, we show that the well-known alternating split Bregman iteration method is a special case of the proposed algorithms and we introduce a parameter to the iteration improve its convergence property. The proposed algorithm is tested numerically for the L1-TV denoising model. Numerical results show that the proposed algorithm outperforms an existing algorithm of Chambolle and Pock.

References

[1] G. Aubert and L. Vese, *A variational method in image recovery*, SIAM Journal on Numerical Analysis, 34 (1997), pp. 1948–1979.

[2] H. L. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, AMS Books in Mathematics, Springer, New York, 2011.

[3] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.

[4] J.-F. Cai, R. Chan, L. Shen, and Z. Shen, *Simultaneously inpainting in image and transformed domains*, Numerische Mathematik, 112 (2009), pp. 509–533.

[5] J.-F. Cai, R. H. Chan, L. Shen, and Z. Shen, *Convergence analysis of tight framelet approach for missing data recovery*, Advances in Computational Mathematics, 31 (2009), pp. 87–113.

[6] J.-F. Cai, S. Osher, and Z. Shen, *Linearized Bregman iteration for frame based image deblurring*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 226–252.

[7] ——, *Split Bregman methods and frame based image restoration*, Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal, 2 (2009), pp. 337–369.

[8] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 89–97.

[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, 40 (2011), pp. 120–145.

[10] R. Chan, T. Chan, L. Shen, and Z. Shen, *Wavelet algorithms for high-resolution image reconstruction*, SIAM Journal on Scientific Computing, 24 (2003), pp. 1408–1432.

[11] R. Chan, S. D. Riemenschneider, L. Shen, and Z. Shen, *Tight frame: The efficient way for high-resolution image reconstruction*, Applied and Computational Harmonic Analysis, 17 (2004), pp. 91–115.

[12] T. Chan and S. Esedoglu, *Aspects of total variation regularized l^1 function approximation*, SIAM Journal on Applied Mathematics, 65 (2005), pp. 1817–1837.

[13] T. Chan, G. H. Golub, and P. Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM Journal on Scientific Computing, 20 (1999), pp. 1964–1977.

[14] C. Clason, B. Jin, and K. Kunisch, *A duality-based splitting method for L1-TV image restoration with automatic regularization parameter choice*, SIAM Journal on Scientific Computing, 32 (2010), pp. 1484–1505.

[15] P. Combettes and V. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal, 4 (2005), pp. 1168–1200.

[16] J. Darbon and M. Sigelle, *Image restoration with discrete constrained total variation part i: Fast and exact minimization*, Journal of Mathematical Imaging and Vision, 26 (2006), pp. 261–276.

[17] I. Daubechies, M. Defrise, and C. D. Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications on Pure and Applied Mathematics, 57 (2004), pp. 1413–1541.

[18] Y. Dong, M. Hintermuller, and M. Neri, *A primal-dual method for l^1-TV image denoising*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 1168–1189.

[19] M. Figueiredo and R. D. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Transactions on Image Processing, 12 (2003), pp. 906–916.

[20] T. Goldstein and S. Osher, *The split Bregman method for ℓ^1 regularization problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 323–343.

[21] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective*, SIAM Journal on Imaging Sciences, 5 (2012), pp. 119–149.

[22] J. Hiriart-Urruty, C. Lemarechal, *Convex Analysis and Minimization Algorithms: Part 1: Fundamentals, Volume 1*, Springer, New York, NY, 1996.

[23] M. Hintermuller and G. Stadler, *An feasible primal-dual algorithm for TV-based inf-convolution-type image restoration*, SIAM Journal on Scientific Computing, 28 (2006), pp. 1–23.

[24] A. Krol, S. Li, L. Shen, and Y. Xu, *Preconditioned alternating projection algorithms for maximum* a Posteriori *ECT reconstruction*, Inverse Problems, under revision, (2012).

[25] Q. Li, C. A. Micchelli, L. Shen, and Y. Xu, *A proximity algorithm accelerated by Gauss-Seidel iterations for* L1/TV *denoising models*, Inverse Problems, accepted, 28(2012), p.095003.

[26] Q. Li, L. Shen, and L. Yang, *Split-Bregman iteration for framelet based image inpainting*, Applied and Computational Harmonic Analysis, 32 (2012), pp. 145–154.

[27] C. A. Micchelli, L. Shen, and Y. Xu, *Proximity algorithms for image models: Denoising*, Inverse Problems, 27 (2011), p. 045009(30pp).

[28] C. A. Micchelli, L. Shen, Y. Xu, and X. Zeng, *Proximity algorithms for image models II: L1/TV denosing*, Advances in Computational Mathematics, online version available, (2011).

[29] J.-J. Moreau, *Fonctions convexes duales et points proximaux dans un espace hilbertien*, C.R. Acad. Sci. Paris Sér. A Math., 255 (1962), pp. 1897–2899.

[30] Y. Nesterov, *Smooth minimization of non-smooth functions*, Mathematical Programming, 103 (2005), pp. 127–152.

[31] M. Nikolova, *Local strong homogeneity of a regularized estimator*, SIAM Journal of Applied Mathematics, 61 (2000), pp. 633–658.

[32] M. Nikolova, *A variational approach to remove outliers and impulse noise*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 99–120.

[33] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal, 4 (2005), pp. 460–489.

[34] T. Pock and A. Chambolle, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, 2011.

[35] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[36] ——, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization, 14 (1976), pp. 877–898.

[37] L. Rudin and S. Osher, *Total variation based image restoration with free local constraints*, in IEEE International Conference on Image Processing, 1994, pp. 31–35.

[38] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.

[39] S. Setzer, *Split Bregman Algorithm, Douglas-Rachford Splitting and Frame Shrinkage*, In: Morken, K., Lysaker, M., Lie, K.-A., Tai, X.-C.(eds.) Scale Space and Variational Methods in Computer Vision, LNCS, vol. 5567, pp. 464-476, Springer, Berlin (2009).

[40] C. Vogel and M. Oman, *Iterative methods for total variation denoising*, SIAM Journal on Scientific Computing, 17 (1996), pp. 227–238.

[41] ——, *Fast, robust total variation-based reconstruction of noisy, blurring images*, IEEE Transactions on Image Processing, 7 (1998), pp. 813–824.

[42] J. Yang, Y. Zhang, and W. Yin, *An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise*, SIAM Journal on Scientific Computing, 31 (2009), pp. 2842–2865.

[43] X. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, Journal of Scientific Computing, 46 (2010), pp. 20–46.