Two-Step Fixed-Point Proximity Algorithms for Multi-block Separable Convex Problems

Qia Li · Yuesheng Xu · Na Zhang

Received: 6 February 2016 / Revised: 19 August 2016 / Accepted: 23 August 2016 / Published online: 29 August 2016
© Springer Science+Business Media New York 2016

Abstract Multi-block separable convex problems recently received considerable attention. Optimization problems of this type minimize separable convex objective functions with linear constraints. Challenges encountered in algorithmic development applying the classic alternating direction method of multipliers (ADMM) come from the fact that ADMM for the optimization problems of this type is not necessarily convergent. However, it is observed that ADMM applying to problems of this type outperforms numerically many of its variants with guaranteed theoretical convergence. The goal of this paper is to develop convergent and computationally efficient algorithms for solving multi-block separable convex problems. We first characterize the solutions of the optimization problems by proximity operators of the convex functions involved in their objective functions. We then design a class of two-step fixed-point iterative schemes for solving these problems based on the characterization. We further prove convergence of the iterative schemes and show that they have $O\left(\frac{1}{k}\right)$ of convergence rate in the ergodic sense and the sense of the partial primal-dual gap, where $k$ denotes the iteration number. Moreover, we derive specific two-step fixed-point proximity algorithms (2SFPPA) from the proposed iterative schemes and establish their global convergence. Numerical experiments for solving the sparse MRI problem demonstrate the numerical efficiency of the proposed 2SFPPA.

This research is supported in part by Guangdong Provincial Government of China through the “Computational Science Innovative Research Team” program, by the Natural Science Foundation of China under Grants 11501584, 11471013 and 91530117, by the US National Science Foundation under Grant DMS-1522332, and by the Natural Science Foundation of Guangdong Province under Grants 2014A030310332 and 2014A030310414.

Na Zhang
nzhsysu@gmail.com

1 Department of Applied Mathematics, College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, People’s Republic of China

2 School of Data and Computer Sciences and Guangdong Province Key Laboratory of Computational Science, Sun Yat-sen University, Guangzhou 510275, People’s Republic of China

3 Syracuse University, Syracuse, NY 13244, USA
Keywords Multi-block separable convex problems · Fixed-point proximity algorithms · Two-step algorithms

1 Introduction

We consider in this paper the convex minimization problem with linear constraints and a separable objective function in the form of a sum of several convex functions. For a positive integer \( d \), by \( \mathbb{R}^d \) we denote the usual \( d \)-dimensional Euclidean space. The minimization problem we consider in this paper has the form

\[
\min \left\{ \sum_{i=1}^s f_i(x_i) : \sum_{i=1}^s A_i x_i = b, \; x_i \in \mathbb{R}^{n_i}, \; i = 1, 2, \ldots, s \right\},
\]

where \( f_i : \mathbb{R}^{n_i} \to [0, +\infty) \) is a proper lower semicontinuous convex function, \( A_i \) is a given \( m \times n_i \) real matrix, \( n_i \) is the dimension of variable \( x_i \), for \( i = 1, 2, \ldots, s \) and \( b \in \mathbb{R}^m \) is a given vector. Here, variable \( x \) is decomposed into \( s \) blocks, that is \( x := (x_1, x_2, \ldots, x_s) \).

Many problems arising from image processing and machine learning can be cast into the form of model (1). For example, the total-variation based image denoising model [27,28], sparse representation based image restoration [3,7,8,21], lasso regression [35] and support vector machines [10] are special cases of problem (1) with \( s = 2 \). In addition, we refer to [22,24,29,34] for applications of model (1) with \( s \geq 3 \).

The alternating direction method of multipliers (ADMM) [14] was originally proposed for solving problem (1) with \( s = 2 \), and was recently widely used in the area of image processing [5,15,30,36]. Since ADMM requires inner iterations to solve its subproblems of ADMM, its linearized version (LADMM) without requiring inner iterations was proposed and was successfully used in applications [13]. As \( s \geq 3 \), one can directly extend the original ADMM (LADMM) to problem (1). Without an additional assumption, however, it was recently shown in [9] that the direct extension of ADMM to multi-block convex problems is not necessarily convergent, although it may work well in practice. Very recently, there were some investigations [4,11,18,23] on convergence of the extension of ADMM under some additional assumptions. Some researchers dedicated to modify ADMM or LADMM to make it convergent. For instance, the Jacobian-type ADMM was proposed in [12] for parallel computing, the semi-proximal ADMM proposed in [18,32] is for convex quadratic programming and conic programming, the Gaussian back substitution technique was proposed in [16,17] to make ADMM and LADMM converge. It was shown in [16,17] the attractiveness of the Gaussian back substitution technique for theoretical analysis on convergence of ADMM-type algorithms. However, the numerical results show that the correction step is time consuming and the ADMM (LADMM) with Gaussian back substitution may require more iterations than the direct extension of ADMM (LADMM) to achieve the same objective function value. Therefore, in this paper, we dedicate to establishing convergent and computationally efficient algorithms for solving the optimization problem (1).

As shown in [1,19,20,22,25], the notion of proximity operators provides a useful tool for the algorithmic development due to its firmly nonexpansive property. ADMM was shown in [20] a special case of the proximity algorithms. Although the one-step fixed-point proximity algorithms proposed in [20] can be applied to model (1) directly, they do not utilize the separable property of the objective function, that is, the variable \( x_1, x_2, \ldots, x_s \) are updated simultaneously. In contrast, ADMM takes advantage of the separability of the objective function and utilizes the block-wise Gauss-Seidel technique. Thus, in order to develop convergent
algorithms for problem (1), we propose to develop two-step fixed-point proximity algorithms. The term two-step means that when we update values of the next step, we not only use values of the current step but also those of the previous step. In one of our previous papers [20], we designed a multi-step iterative scheme, introduced the notions of weakly firmly nonexpansive operators and Condition-M (Semi-Condition-M), and presented the convergence results of the multi-step scheme with the help of the notions. In this paper, we will follow the idea of [20] to develop convergent two-step fixed-point proximity algorithms.

The contributions of this paper may be described as follows. First, we characterize solutions of problem (1) in terms of fixed-points of a proximity related operator and develop two-step fixed-point iterative schemes based on the fixed-point equation. Second, we prove convergence of the proposed iterative schemes by using the notions of weakly firmly nonexpansive maps and Condition-M originally introduced in [20]. We prove that as long as the matrices involved in the schemes satisfy Condition-M, which can be easily verified, the proposed iterative schemes converge and the sequences \( \{ (x^k_1, \ldots, x^k_s) : k \in \mathbb{N} \} \) generated by the proposed algorithms converge to a solution of problem (1). Third, we analyze the convergence rate of the proposed iterative schemes. We prove that the schemes have \( O \left( \frac{1}{k} \right) \) of ergodic convergence rate. In addition, the average of the sequences generated by the proposed schemes has \( O \left( \frac{1}{k} \right) \) of convergence rate in the sense of the primal-dual gap. Fourth, several specific convergent algorithms are designed from the iterative schemes, including the two-step implicit and explicit fixed-point proximity algorithms as well as their variants. Furthermore, we apply a proposed two-step fixed-point proximity algorithm to the sparse MRI reconstruction problem. Numerical results show that the proposed two-step fixed-point proximity algorithm performs as efficiently as the direct extension of LADMM, which is not guaranteed to converge.

We organize this paper in eight sections. In Sect. 2, we characterize solutions of problem (1) by fixed-points of a proximity related operator. Based on this characterization, we develop in Sect. 3 two-step iterative schemes and prove their convergence in Sect. 4. In Sect. 5, we analyze the convergence rate of the proposed iterative schemes. We design in Sect. 6 several specific algorithms from the iterative schemes and apply in Sect. 7 one of them to the sparse MRI reconstruction problem. We conclude this paper in Sect. 8.

2 A Characterization of Solutions of the Minimization Problem

In this section we present a characterization of solutions of model (1) in terms of a system of fixed-point equations via the proximity operators of the functions involved in the objective function. The system of fixed-point equations will serve as a basis for developing iterative schemes for solving the problem.

We now recall the notion of the proximity operator of a convex function. For \( x \) and \( y \) in \( \mathbb{R}^d \), we denote the standard 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 the standard \( \ell_2 \)-norm by \( \|x\|_2 := \langle x, x \rangle^{\frac{1}{2}} \). By \( \mathbb{S}_+^d \) we denote the set of symmetric positive definite matrices. For an \( H \in \mathbb{S}_+^d \), the \( H \)-weighted inner product is defined by \( \langle x, y \rangle_H := \langle x, Hy \rangle \) and the corresponding \( H \)-weighted \( \ell_2 \)-norm is defined by \( \|x\|_H := \langle x, x \rangle_H^{\frac{1}{2}} \). For a \( d \times \ell \) matrix \( A \), we define \( \|A\|_2 \) as the largest singular value of \( A \). By \( \Gamma_0(\mathbb{R}^d) \) we denote the class of all lower semicontinuous proper convex functions \( \varphi : \mathbb{R}^d \rightarrow \mathbb{R} \). For a function \( \varphi \in \Gamma_0(\mathbb{R}^d) \), the proximity operator of \( \varphi \) with respect to a given matrix \( H \in \mathbb{S}_+^d \), denoted by \( \text{prox}_{\varphi, H} \), is a mapping from \( \mathbb{R}^d \) to itself, defined for a given point \( x \in \mathbb{R}^d \) by
A characterization of the subdifferential of a function

\[ \text{dom} \varphi := \{ x \in \mathbb{R}^d : \varphi(x) \leq 0 \} \]

We remark here that the symmetric positive definite matrix \( H \) defines specific inner-product of the Hilbert space \( \mathbb{R}^d \) and if \( H = I \) we do not specify the matrix \( H \) for simplicity. As shown in [2], the proximity operator of a convex function is firmly nonexpansive and is contractive when the function is strongly convex.

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

\[ \varphi^*(y) := \sup \{ \langle x, y \rangle - \varphi(x) : x \in \mathbb{R}^d \} \]

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

\[ y \in \partial \varphi(x) \quad \text{if and only if} \quad x \in \partial \varphi^*(y). \tag{4} \]

The notion of the indicator function is also required. For a set \( S \subseteq \mathbb{R}^d \), the indicator function on \( S \), at point \( x \), is defined as

\[ \iota_S(x) := \begin{cases} 0, & \text{if } x \in S, \\ +\infty, & \text{else.} \end{cases} \]

Moreover, we denote by \( \text{cone}(S) \) the smallest cone in \( \mathbb{R}^d \) containing \( S \). Then the relative interior of \( S \) (see Definition 6.9 of [2]) is defined as

\[ \text{ri}(S) := \{ x \in S : \text{cone}(S - x) = \text{span}(S - x) \} \]

For simplicity, let \( n := \sum_{i=1}^d n_i \) and \( A := [A_1 \ A_2 \ldots \ A_s] \). Then, problem (1) can be rewritten as

\[ \min \left\{ \sum_{i=1}^s f_i(x_i) + \iota_C(Ax) : x_i \in \mathbb{R}^{n_i}, i \in \mathbb{N}_s \right\}, \tag{5} \]

where

\[ C := \{ b \}. \tag{6} \]

Now, we are ready to characterize the solutions of model (1) with the help of (3) and (4).
Theorem 1 Let $f_i \in \Gamma_0(\mathbb{R}^{n_i})$, $A_i$ an $m \times n_i$ matrix for $i \in \mathbb{N}_s$ and $b \in A(\text{ri}(\text{dom}(\sum_{i=1}^{s} f_i)))$. If $x := (x_1, x_2, \ldots, x_s) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \cdots \times \mathbb{R}^{n_s}$ is a solution of problem (1), then for any $\beta > 0$ and $\alpha_i > 0$, $i \in \mathbb{N}_s$, there exists a vector $y \in \mathbb{R}^m$ such that

$$x_i = \text{prox}_{\frac{\alpha_i}{\beta} A_i^T} \left( x_i - \frac{\alpha_i}{\beta} A_i^T y \right), \quad i \in \mathbb{N}_s,$$

$$y = \text{prox}_{\beta \ell_C^*} \left( y + \beta \sum_{i=1}^{s} A_i x_i \right).$$

Conversely, if there exist $\beta > 0$, $\alpha_i > 0$ for $i \in \mathbb{N}_s$, $x := (x_1, x_2, \ldots, x_s) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \cdots \times \mathbb{R}^{n_s}$ and $y \in \mathbb{R}^m$ satisfying Eqs. (7) and (8), then $x$ is a solution of problem (1).

Proof We prove this theorem by applying Fermat’s rule that a vector $x := (x_1, x_2, \ldots, x_s) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \cdots \times \mathbb{R}^{n_s}$ 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 $x := (x_1, x_2, \ldots, x_s) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \cdots \times \mathbb{R}^{n_s}$ be a solution of model (1). From Theorem 16.37 of [2], the chain rule of the subdifferential holds due to $b \in A(\text{ri}(\text{dom}(\sum_{i=1}^{m} f_i)))$. Then by Fermat’s rule, we obtain

$$0 \in \partial f_i(x_i) + A_i^T \partial \ell_C(Ax)$$

for $i \in \mathbb{N}_s$. Thus, there exists $y \in \mathbb{R}^m$ such that $y \in \partial \ell_C(Ax)$ and $-A_i^T y \in \partial f_i(x_i)$ for $i \in \mathbb{N}_s$. The last inclusion implies that for any $\alpha_i > 0$, $\beta > 0$, $-\frac{\alpha_i}{\beta} A_i^T y \in \partial(\frac{\alpha_i}{\beta} f_i)(x_i)$. Therefore, Eq. (7) follows from (3). By (4), from $y \in \partial \ell_C(Ax)$, we have that $Ax \in \partial \ell_C^*(y)$. Hence, for any $\beta > 0$, we obtain that $\beta Ax \in \partial(\beta \ell_C^*)(y)$, which by (3) is equivalent to Eq. (8).

Conversely, suppose that there exist $\alpha_i > 0$, $\beta > 0$, $y \in \mathbb{R}^m$ and $x_i \in \mathbb{R}^{n_i}$ for $i \in \mathbb{N}_s$ satisfying the system of fixed-point Eqs. (7) and (8). The relation (3) ensures that $y \in \partial \ell_C(Ax)$ and $-A_i^T y \in \partial f_i(x_i)$. Clearly, these inclusions together ensure that the relation (9) holds. That is, the zero vector is in the subdifferential of the objective function at $(x_1, \ldots, x_s)$. Again, by Fermat’s rule, $(x_1, \ldots, x_s)$ is a solution of model (1).

Theorem 1 characterizes a solution of problem (1) in terms of the system of fixed-point Eqs. (7) and (8). Through out this paper, for problem (1), we assume that $b \in A(\text{ri}(\text{dom}(\sum_{i=1}^{s} f_i)))$ and it has at least one solution. With these assumptions and by Theorem 1, we know that fixed-point Eqs. (7) and (8) have at least one solution for any $\alpha_i > 0$, $i \in \mathbb{N}_s$ and $\beta > 0$. This makes it possible for us to compute a solution of model (1) by developing fixed-point iterative schemes.

3 Two-Step Iterative Schemes

We develop in this section a class of two-step iterative schemes for solving optimization problem (1) by using the system of fixed-point Eqs. (7) and (8).

We begin with rewriting Eqs. (7) and (8) in a compact form. To this end, we first introduce an operator by integrating together the $s + 1$ proximity operators involved in Eqs. (7) and (8). Specifically, for given $f_i \in \Gamma_0(\mathbb{R}^{n_i})$, $\ell_C \in \Gamma_0(\mathbb{R}^m)$, $\alpha_i > 0$, $\beta > 0$, $i \in \mathbb{N}_s$, we define the operator $T := T_{(c, \beta)}(f_1, \ldots, f_s, \ell_C) : \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_s} \times \mathbb{R}^m \to \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_s} \times \mathbb{R}^m$ at a vector $v := (x_1, \ldots, x_s, y) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_s} \times \mathbb{R}^m$ as follows:

$$T(v) := \left( \text{prox}_{\frac{\alpha_1}{\beta} f_1}(x_1), \ldots, \text{prox}_{\frac{\alpha_s}{\beta} f_s}(x_s), \text{prox}_{\beta \ell_C^*}(y) \right).$$

Springer
Operator $T$ couples all the proximity operators  $	ext{prox}_{\frac{\mu_i}{T} f_i}$, $i \in \mathbb{N}_s$ and  $\text{prox}_{\frac{\mu_C}{C}}$. In the following lemma, we show that the operator $T$ is the proximity operator of a new convex function

$$
\Phi(v) := \sum_{i=1}^{s} f_i(x_i) + \iota_C^*(y)
$$

for $v := (x_1, \ldots, x_s, y)$ with respect to the matrix

$$
R := \text{diag} \left( \frac{\beta}{\alpha_1} 1_{n_1}, \ldots, \frac{\beta}{\alpha_s} 1_{n_s}, \frac{1}{\beta} 1_m \right),
$$

where $1_d$ (resp. $0_d$) is a $d$-dimensional vector with 1 (resp. 0) as its components for any $d \in \mathbb{N}$.

**Lemma 1** If operator $T$ is defined by (10), then $T$ is the proximity operator of the function $\Phi$ with respect to the matrix $R$, that is, $T = \text{prox}_{\Phi, R}$.

Here we omit the proof since one can complete it by referring to Lemma 3.1 of [20]. By Lemma 1, we know that the operator $T$ is firmly non-expansive with respect to the matrix $R$. Let

$$
P := \text{diag} \left( \frac{\beta}{\alpha_1} 1_{n_1}, \ldots, \frac{\beta}{\alpha_s} 1_{n_s} \right).
$$

With the help of the above notation, Eqs. (7) and (8) can be reformulated in a compact form

$$
v = (T \circ E)(v),
$$

where

$$
E := \begin{bmatrix} I & -P^{-1}A^T \\ \beta A & I \end{bmatrix}.
$$

Theorem 1 together with Eq. (14) indicates that finding a solution of problem (1) essentially amounts to computing a fixed-point of the operator $T \circ E$. As discussed at the end of Sect. 2, the operator $T \circ E$ has at least one fixed-point. We next focus on developing efficient iterative schemes for finding a fixed-point of the operator. As shown in [20], the matrix $E$ is not nonexpansive due to the fact that $\|E\|_2 > 1$. Therefore, a simple fixed-point iteration $v^{k+1} = (T \circ E)(v^k)$ for a given initial guess $v^0$, may not yield a convergent sequence $\{v^k : k \in \mathbb{N}\}$, where $\mathbb{N}$ is the set of all natural numbers.

Our idea is to split the expansive matrix $E$ into several terms, as in [20] and in [22]. Here, we split $E$ as

$$
E = (E - R^{-1}M_0) + R^{-1}M_1 + R^{-1}M_2,
$$

where $M_i \in \mathbb{R}^{(n+m) \times (n+m)}$ for $i = 0, 1, 2$ and $M_0 = M_1 + M_2$. Accordingly, Eq. (14) is equivalent to

$$
v = T((E - R^{-1}M_0)v + R^{-1}M_1v + R^{-1}M_2v).
$$

Thus, we propose the following two-step iterative scheme:

$$
v^{k+1} = T \left( (E - R^{-1}M_0)v^{k+1} + R^{-1}M_1v^{k} + R^{-1}M_2v^{k-1} \right).
$$

We point out here that although iterative scheme (17) is an implicit scheme for the whole vector $v$, it becomes explicit by choosing $M_0$ satisfying that $E - R^{-1}M_0$ is a strictly upper triangular or lower triangular matrix. Further, we assume that there exists a unique $v^{k+1}$
satisfying (17) for any \( v^k, v^{k-1} \in \mathbb{R}^{n+m} \) in the rest of this paper. We shall choose matrices \( M_0, M_1, M_2 \) in the next section so that iterative scheme (17) converges.

To close this section, we remark that when \( M_2 = 0 \) (in this case, \( M_0 = M_1 \)), the two-step iterative scheme (17) reduces to a one-step iterative scheme

\[
v^{k+1} = T \left((E - R^{-1}M_0)v^{k+1} + R^{-1}M_0v^k \right).
\]

(18)

Many efficient algorithms can be obtained from (18) by specifying the matrix \( M_0 \). The reader is referred to [20] for details.

### 4 Convergence Analysis of the Proposed Iterative Scheme

In this section, we study the convergence of iterative scheme (17). By applying the notion of weakly firmly nonexpansive operators and Condition-M, which were first introduced in [20], we prove that if the matrices \( M_0, M_1, M_2 \) satisfy Condition-M, then the sequence \( \{v^k : k \in \mathbb{N}\} \) generated from iterative scheme (17) converges to a solution of Eq. (14). Hence, the sequence \( \{x^k : k \in \mathbb{N}\} \) converges to a solution of model (1).

We begin with rewriting iterative scheme (17) in an explicit way. To this end, we introduce \( \mathcal{M} := \{M_0, M_1, M_2\} \). We also define \( T_{\mathcal{M}} : \mathbb{R}^{n+m} \times \mathbb{R}^{n+m} \to \mathbb{R}^{n+m} \), at \((u_1, u_2) \in \mathbb{R}^{n+m} \times \mathbb{R}^{n+m}\), as \( w = T_{\mathcal{M}}(u_1, u_2) \) with \( w \) satisfying

\[
w = T \left((E - R^{-1}M_0)w + R^{-1}M_1u_1 + R^{-1}M_2u_2 \right).
\]

(19)

The operator \( T_{\mathcal{M}} \) is well-defined if the corresponding set \( \mathcal{M} \) is carefully chosen. Here, the word “well-defined” means that there exists a unique \( w \in \mathbb{R}^{n+m} \) satisfying (17) for any \((u_1, u_2) \in \mathbb{R}^{n+m} \times \mathbb{R}^{n+m}\). With the help of \( \mathcal{M} \) and \( T_{\mathcal{M}}, (17) \) can be rewritten as

\[
v^{k+1} = T_{\mathcal{M}}\left(v^k, v^{k-1}\right).
\]

(20)

Now, we review the notion of weakly firmly nonexpansive operators and Condition-M, which were introduced in [20].

**Definition 1** (Weakly firmly nonexpansive) We say an operator \( T : \mathbb{R}^{2d} \to \mathbb{R}^{d} \) is weakly firmly nonexpansive with respect to \( \mathcal{M} \), if for any \((u_i, w_i, z_i) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \) satisfying \( z_i = T(u_i, w_i) \) for \( i = 1, 2 \), there holds

\[
\langle z_2 - z_1, M_0(z_2 - z_1) \rangle \leq \langle z_2 - z_1, M_1(u_2 - u_1) + M_2(w_2 - w_1) \rangle.
\]

Next we recall Condition-M.

**Definition 2** (Condition-M) We say a set \( \mathcal{M} := \{M_0, M_1, M_2\} \) of \( d \times d \) matrices satisfies Condition-M, if the following three hypotheses are satisfied:

(i) \( M_0 = M_1 + M_2 \),
(ii) \( H := M_0 + M_2 \) is in \( \mathbb{S}_{d}^+ \),
(iii) \( \|H^{-\frac{1}{2}}M_2H^{-\frac{1}{2}}\|_2 < \frac{1}{2} \).

We also need a property of weakly firmly nonexpansive operators established in [20].

**Theorem 2** Suppose that the operator \( T : \mathbb{R}^{2d} \to \mathbb{R}^{d} \) is weakly firmly nonexpansive with respect to \( \mathcal{M} := \{M_0, M_1, M_2\} \) with \( \text{dom}(T) = \mathbb{R}^{2d} \) and the set of fixed-points of \( T \) is nonempty. Let the sequence \( \{w^k : k \in \mathbb{N}\} \) be generated by \( w^{k+1} = T(w^k, w^{k-1}) \) for any given \( w^0, w^1 \in \mathbb{R}^{d} \). If \( \mathcal{M} \) satisfies Condition-M, then \( \{w^k : k \in \mathbb{N}\} \) converges. In addition, if \( T \) is continuous, then \( \{w^k : k \in \mathbb{N}\} \) converges to a fixed-point of \( T \).
According to the theorem above, in order to ensure convergence of iterative scheme (20), it suffices to prove \( T_M \) defined by (19) is weakly firmly nonexpansive and continuous. We show it in the next proposition. Before doing this, we define a skew-symmetric matrix \( S_A \) for an \( m \times n \) matrix \( A \) as

\[
S_A := \begin{bmatrix}
0 & -A^T \\
A & 0
\end{bmatrix}.
\]

(21)

Then, \( E = I + R^{-1}S_A \).

**Proposition 1** Let \( f_i \in \Gamma_0(\mathbb{R}^n) \), \( \alpha_i > 0 \) for \( i \in \mathbb{N} \), and \( \beta > 0 \). Let \( \mathcal{M} := \{M_0, M_1, M_2\} \) be a set of \((n + m) \times (n + m)\) matrices and \( T_M \) be defined by (19). If \( T_M \) is well-defined, then

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

(ii) \( T_M \) is continuous.

**Proof** We first prove Item (i). It follows from the definition of \( T_M \) that for any \((u, w, z) \in \mathbb{R}^{n+m} \times \mathbb{R}^{n+m} \times \mathbb{R}^{n+m}\) satisfying \( z_i = T_M(u_i, w_i) \), for \( i = 1, 2 \), there holds

\[
z_i = T((E - R^{-1}M_0)z_i + R^{-1}M_1u_i + R^{-1}M_2w_i).
\]

According to Lemma 1, \( T \) is firmly nonexpansive with respect to \( R \). Thus, we observe that

\[
\|z_2 - z_1\|^2_R \leq \langle z_2 - z_1, (RE - M_0)(z_2 - z_1) + M_1(u_2 - u_1) + M_2(w_2 - w_1) \rangle.
\]

Since \( RE = R + S_A \) and \( S_A \) is skew-symmetric, we have that

\[
\langle z_2 - z_1, M_0(z_2 - z_1) \rangle \leq \langle z_2 - z_1, M_1(u_2 - u_1) + M_2(w_2 - w_1) \rangle.
\]

From Definition 1, we get Item (i).

We next prove Item (ii). From the definition of \( T_M \), for any sequence \( \{(u^k, w^k, z^k) \in \mathbb{R}^{n+m} \times \mathbb{R}^{n+m} \times \mathbb{R}^{n+m} : k \in \mathbb{N}\} \) satisfying \( z^k = T_M(u^k, w^k) \) and converging to \((u, w, z)\), we have that

\[
z^k = T \left((E - M_0)z^k + R^{-1}M_1u^k + R^{-1}M_2w^k\right).
\]

This with the continuity of \( T \) implies that

\[
z = T \left((E - M_0)z + R^{-1}M_1u + R^{-1}M_2w\right).
\]

Thus, \( z = T_M(u, w) \), proving Item (ii).

We are now ready to prove convergence of the sequence generated from iterative scheme (17).

**Theorem 3** Let \( f_i \in \Gamma_0(\mathbb{R}^m) \), \( \alpha_i > 0 \) for \( i \in \mathbb{N} \), and \( \beta > 0 \). Let \( T \) and \( E \) be defined as (10) and (15) respectively, \( \mathcal{M} := \{M_0, M_1, M_2\} \) be a set of \((n + m) \times (n + m)\) matrices and \( T_M \) be defined by (19). Let \( \{v^k : k \in \mathbb{N}\} \) be generated by (17) for given points \( v^0, v_1 \). Suppose that \( T_M \) is well-defined. If \( \mathcal{M} \) satisfies Condition-M, then the sequence \( \{v^k : k \in \mathbb{N}\} \) converges to a fixed-point of \( T \circ E \), and \( \{x^k : k \in \mathbb{N}\} \) converges to a solution of problem (1).

**Proof** By the definition of \( T_M \), operators \( T_M \) and \( T \circ E \) share the same set of fixed-points. By Proposition 1, the operator \( T_M \) is weakly firmly non-expansive with respect to \( \mathcal{M} \) and continuous. Therefore, Theorem 2 ensures that the sequence \( \{v^k : k \in \mathbb{N}\} \) converges to a fixed-point of \( T_M \). By Proposition 1, the sequence \( \{x^k : k \in \mathbb{N}\} \) converges to a solution of problem (1).

□

Theorem 3 shows that convergence of iterative scheme (17) relies completely on whether the matrices set \( \mathcal{M} \) used in scheme (17) satisfies Condition-M. We will develop in Sect. 6 specific convergent algorithms by generating sets of \( \{M_0, M_1, M_2\} \) satisfying Condition-M.
5 Convergence Rate of the Proposed Two-Step Iterative Scheme

In this section, we study the convergence rate of the proposed fixed-point iterative scheme (17). We show that the proposed algorithm has $O\left(\frac{1}{k}\right)$ convergence rate in the ergodic sense and the sense of the partial primal-dual gap.

5.1 Ergodic $O\left(\frac{1}{k}\right)$ Rate

We first study the convergence rate of the proposed algorithm (17) in the ergodic sense. We prove in this subsection that the proposed iterative scheme (17) has $O\left(\frac{1}{k}\right)$ convergence in the ergodic sense. To this end, we first review a lemma presented in [31].

Lemma 2 If a sequence $\{a^k : k \in \mathbb{N}\}$ satisfies: $a^k \geq 0$ and $\sum_{i=1}^{+\infty} a^i < +\infty$, then

(i) $\frac{1}{k} \sum_{i=1}^{k} a^i = O\left(\frac{1}{k}\right)$,
(ii) $\min_{i \leq k} \{a^i\} = o\left(\frac{1}{k}\right)$.

The main results of this subsection are presented in the next theorem.

Theorem 4 Let $f_i \in \Gamma_0(\mathbb{R}^{n_i})$ and $A_i$ an $m \times n_i$ matrix for $i \in \mathbb{N}_s$. Let $\alpha_i > 0$ for $i \in \mathbb{N}_s$ and $\beta > 0$. Let $T$ and $E$ be defined as (10) and (15) respectively. Let the sequence $\{v^k : k \in \mathbb{N}\}$ be generated from (17) for any given $v^0, v^1 \in \mathbb{R}^{n+m}$. Suppose that $T_M$ is well-defined. If $M$ satisfies Condition-M, then

(i) the sequence $\{v^k : k \in \mathbb{N}\}$ has $O\left(\frac{1}{k}\right)$ convergence in the ergodic sense, that is

$$\frac{1}{k} \sum_{i=1}^{k} \|v^{i+1} - v^i\|_2^2 = O\left(\frac{1}{k}\right), \quad (22)$$

(ii) the running minimal of progress, $\min_{i \leq k} \{\|v^{i+1} - v^i\|_2\}$, has $o\left(\frac{1}{k}\right)$ convergence.

Proof By Lemma 2, we only need to prove

$$\sum_{i=1}^{+\infty} \|v^{i+1} - v^i\|_2^2 < +\infty. \quad (23)$$

By the definition of $T_M$, the sequence $\{v^k : k \in \mathbb{N}\}$ generated from (17) can also be generated by (20) for the same given $v^0, v^1$. Since $T_M$ is weakly firmly nonexpansive with respect to $M$ and $M$ satisfies Condition-M, by Lemma 4.4 of [20], we have for any $k \geq 3$ that

$$\|e^k\|_H^2 \leq 2\|e^2\|_H^2 + 2\|\tilde{M}\|_2^2 \|r^2\|_H^2 - 2\langle e^2, M_2 r^2 \rangle - \left(\frac{1}{2} - \frac{1}{2}\|\tilde{M}\|_2^2\right) \sum_{i=2}^{k-1} \|r^{i+1}\|_H^2, \quad (24)$$

where $e^i := v^i - v$ for a fixed-point of $T_M$, $r^i := v^i - v^{i-1}$ and $\tilde{M} := H^{-1/2} M_2 H^{-1/2}$. By (iii) of Condition-M, we have $\frac{1}{2} - \frac{1}{2}\|\tilde{M}\|_2^2 > 0$. Then (23) is obtained immediately from (24) and the fact that $H \in S^{n+m}_+$. \qed
5.2 Partial Primal-Dual Gap $O \left( \frac{1}{k} \right)$ Convergence Rate

In this subsection, we study the convergence rate of the proposed iterative algorithm (17) in the sense of the partial primal-dual gap. We prove that iterative scheme (17) has $O \left( \frac{1}{k} \right)$ convergence rate in the sense of the partial primal-dual gap.

We first introduce the notion of the partial primal-dual gap for convex problem (1). To this end, we review the primal-dual formulation of problem (1), that is

$$\min \left\{ \max \left\{ \sum_{i=1}^{s} f_i(x_i) - t_C^*(y) + \langle Ax, y \rangle : y \in \mathbb{R}^m \right\} : x_i \in \mathbb{R}^{n_i}, i \in \mathbb{N}_s \right\}. \quad (25)$$

One can refer to [2] for more details. For two bounded sets $B_1 \subseteq \mathbb{R}^n$ and $B_2 \subseteq \mathbb{R}^m$, the partial primal-dual gap for problem (1) at point $v := (x_1, \ldots, x_s, y) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_s} \times \mathbb{R}^m$ is defined as

$$G_{B_1 \times B_2}(v) := \max \left\{ \sum_{i=1}^{s} f_i(x_i) - t_C^*(y') + \langle Ax, y' \rangle : y' \in B_2 \right\} - \min \left\{ \sum_{i=1}^{s} f_i(x'_i) - t_C^*(y) + \langle Ax', y \rangle : x' \in B_1 \right\}. \quad (26)$$

We refer to [6] for more details on the partial primal-dual gap.

In order to analyze the convergence rate of iterative scheme (17), we define $G : \mathbb{R}^{n+m} \times \mathbb{R}^{n+m} \rightarrow \mathbb{R}$ by

$$G(v, v') := \Phi(v) - \Phi(v') + \langle v', S_A v \rangle, \quad (27)$$

where $\Phi$ and $S_A$ are defined as (11) and (21) respectively. For two bounded sets $B_1 \times B_2 \subseteq \mathbb{R}^{n+m}$, we define $G(v, v') := (x_1, \ldots, x_s, y) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_s} \times \mathbb{R}^m$ and $v := (x_1, \ldots, x_s, y) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_s} \times \mathbb{R}^m$, one can check that (27) is equivalent to

$$G(v, v') = \sum_{i=1}^{s} f_i(x_i) - t_C^*(y') + \langle Ax, y' \rangle - \left( \sum_{i=1}^{s} f_i(x'_i) - t_C^*(y) + \langle Ax', y \rangle \right).$$

Therefore, in order to analyze the partial primal-dual gap at point $v := (x, y) \in \mathbb{R}^{n} \times \mathbb{R}^{m}$, we only need to estimate the upper bound of $G(v, v')$ for $v' \in B_1 \times B_2$. The next lemma presents an important estimation of $G(k^{k+1}, v)$ for any $v \in \mathbb{R}^{n+m}$.

**Lemma 3** Let $f_i \in \Gamma_0(\mathbb{R}^{n_i}), A_i$ an $m \times n_i$ matrix, $\alpha_i > 0$ for $i \in \mathbb{N}_s$ and $\beta > 0$. Let $T$ and $E$ be defined as (10) and (15) respectively, $M := \{M_0, M_1, M_2\}$ be a set of $(n+m) \times (n+m)$ matrices. Let $\{v^k := (x^k, y^k) \in \mathbb{R}^n \times \mathbb{R}^m : k \in \mathbb{N}\}$ be generated from iterative scheme (17). For all $v \in \mathbb{R}^{n+m}$ there holds

$$G(v^{k+1}, v) \leq \langle M_1 v^k + M_2 v^{k-1} - M_0 v^k, v^{k+1} - v \rangle. \quad (28)$$

**Proof** From iterative scheme (17), Lemma 1 and 3, we have

$$R(E - R^{-1}M_0 - I)v^{k+1} + M_1 v^k + M_2 v^{k-1} \in \partial \Phi(v^{k+1}).$$

Due to $E = I + R^{-1}S_A$, we obtain that

$$-M_0 v^{k+1} + M_1 v^k + M_2 v^{k-1} + S_A v^{k+1} \in \partial \Phi(v^{k+1}).$$

By the definition of subdifferential and the convexity of $\Phi$, we have for any $v \in \mathbb{R}^{n+m}$ that

$$\Phi(v^{k+1}) + \langle M_1 v^k + M_2 v^{k-1} - M_0 v^k, v - v^{k+1} \rangle + \langle v - v^{k+1}, S_A v^{k+1} \rangle \leq \Phi(v).$$

Since $S_A$ is skew-symmetric, the above inequality is equivalent to

$$\Phi(v^{k+1}) + \langle v, S_A v^{k+1} \rangle - \Phi(v) \leq \langle M_1 v^k + M_2 v^{k-1} - M_0 v^k, v^{k+1} - v \rangle.$$
Then, we obtain (28) immediately by the definition of $G$. \hfill \Box

We next study the partial primal-dual gap at $\bar{v}_K := \frac{\sum_{k=2}^{K+1} v^k}{K}$.

**Lemma 4** Let $\{v^k : k \in \mathbb{N}\}$ be generated from iterative scheme (17). Under the same assumptions of Lemma 3, if $\mathcal{M}$ satisfies Condition-M, then for all $v \in \mathbb{R}^{n+m}$ there holds

$$G(\bar{v}_K, v) \leq \frac{3}{4} \|v^1 - v\|_{H}^{2} + \frac{1}{4} \|v^1 - v^0\|_{H}^{2},$$

where $H := M_0 + M_2$ and $\bar{v}_K := \frac{\sum_{k=2}^{K+1} v^k}{K}$.

**Proof** For simplicity, we define for $k \in \mathbb{N}$, $e^k := v^k - v$, $r^k := v^k - v^{k-1}$. By Lemma 3 and Item (i) of Condition-M, we have

$$G(v^{k+1}, v) \leq \left( M_1 e^k + M_2 e^{k-1}, e^{k+1} \right) - \left( M_0 e^{k+1}, e^{k+1} \right).$$

Using $H := M_0 + M_2$ and $M_0 = M_1 + M_2$, the above inequality implies that

$$G(v^{k+1}, v) \leq D_1 + D_2,$$

where $D_1 := -\|e^{k+1}\|_{H}^{2} + \langle e^{k+1}, H e^k \rangle$ and $D_2 := \langle e^{k+1}, M_2 e^{k-1} - 2M_2 e^k + M_2 e^{k-1} \rangle$. By the relationship $r^{k+1} = e^{k+1} - e^k$ and $\langle a, H b \rangle = \frac{1}{2}(\|a\|_{H}^{2} + \|b\|_{H}^{2} - \|a - b\|_{H}^{2})$ for $a, b \in \mathbb{R}^{n+m}$, we obtain that

$$D_1 = \frac{1}{2} \left( -\|e^{k+1}\|_{H}^{2} + \|e^{k}\|_{H}^{2} - \|r^{k+1}\|_{H}^{2} \right).$$

We also have

$$D_2 = \langle e^{k+1}, M_2 (r^{k+1} - r^k) \rangle = \langle e^{k+1}, M_2 e^{k+1} \rangle - \langle r^{k+1}, M_2 r^k \rangle - \langle e^k, M_2 r^k \rangle,$$

where the first equality is obtained by the relationship $r^k = e^k - e^{k-1}$ and the second equality holds due to $e^{k+1} = r^{k+1} + e^k$. Let $\bar{M} := (H^*)^{1/2} M_2 (H^*)^{1/2}$. Then it follows that for any $a > 0$,

$$|\langle r^{k+1}, M_2 r^k \rangle| \leq \frac{a}{2} \|r^{k+1}\|_{H}^{2} + \frac{\|\bar{M}\|_{2}^{2}}{2a} \|r^{k}\|_{H}^{2}.\tag{33}$$

Thus, by (30), (31), (32), and (33), we have

$$G(v^{k+1}, v) \leq \frac{1}{2} \left( -\|e^{k+1}\|_{H}^{2} + \|e^{k}\|_{H}^{2} \right) - \frac{1}{2} \left( 1 - a - \frac{\|\bar{M}\|_{2}^{2}}{a} \right) \|r^{k+1}\|_{H}^{2}$$

$$+ \frac{\|\bar{M}\|_{2}^{2}}{2a} \left( -\|r^{k+1}\|_{H}^{2} + \|r^{k}\|_{H}^{2} \right) + \langle e^{k+1}, M_2 e^{k+1} \rangle - \langle e^k, M_2 r^k \rangle.$$\tag{34}

Summing inequality (34) from $k = 1$ to $k = K$, we have

$$\sum_{k=1}^{K} G(v^{k+1}, v) \leq \frac{1}{2} \left( -\|e^{K+1}\|_{H}^{2} + \|e^{1}\|_{H}^{2} \right) - \frac{1}{2} \left( 1 - a - \frac{\|\bar{M}\|_{2}^{2}}{a} \right) \sum_{k=2}^{K+1} \|r^k\|_{H}^{2}$$

$$+ \frac{\|\bar{M}\|_{2}^{2}}{2a} \left( -\|r^{K+1}\|_{H}^{2} + \|r^{1}\|_{H}^{2} \right) + \langle e^{K+1}, M_2 e^{K+1} \rangle - \langle e^1, M_2 r^1 \rangle.$$\tag{35}
By applying
\[ |\langle e^k, M_2 r^k \rangle| \leq \frac{a}{2} \| e^k \|_H^2 + \frac{\| \tilde{M} \|_2^2}{2a} \| e^k \|_H^2 \]
for \( k = K + 1 \) and \( k = 1 \) to the last two terms of (35), we obtain that
\[
\sum_{k=1}^{K} G(v^{k+1}, v) \leq \frac{1}{2} \left( -1 + a \left( \| e^{K+1} \|_H^2 + (1 + a) \| e^1 \|_H^2 \right) - \frac{1}{2} \left( 1 - a - \frac{\| \tilde{M} \|_2^2}{a} \right) \sum_{k=2}^{K+1} \| r^k \|_H^2 + \frac{\| \tilde{M} \|_2^2}{a} \| r^1 \|_H^2. \right) \tag{36}
\]
Setting \( a = \frac{1}{2} \), it follows that \( 1 - a - \frac{\| \tilde{M} \|_2^2}{a} > 0 \) due to (iii) of Condition-M. This together with (36) yields
\[
\frac{1}{K} \sum_{k=1}^{K} G(v^{k+1}, v) \leq \frac{3}{4} \| e^1 \|_H^2 + \frac{1}{2} \| r^1 \|_H^2, \tag{37}
\]
Since \( G(\cdot, v) \) is convex, we conclude that \( G(\bar{v}_K, v) \leq \frac{1}{K} \sum_{k=1}^{K} G(v^{k+1}, v) \), which together with inequality (37) implies (29).

Now, we are ready to present the partial primal-gap convergence rate of the proposed algorithm (17) in the next theorem.

**Theorem 5** Let \( \{ v^k : k \in \mathbb{N} \} \) be generated from iterative scheme (17). Under the same assumptions of Lemma 3, if \( \mathcal{M} \) satisfies Condition-M, then iterative scheme (17) has \( O\left( \frac{1}{K} \right) \) convergence rate in the partial primal-dual gap sense, that is
\[
G_{B_1 \times B_2}(\bar{v}_K) = O\left( \frac{1}{K} \right),
\]
where \( B_1 \subseteq \mathbb{R}^n \) and \( B_2 \subseteq \mathbb{R}^m \) are bounded and \( \bar{v}_K := \frac{1}{K} \sum_{k=2}^{K+1} v^k \).

**Proof** This is a direct consequence of Lemma 4 and the boundedness of sets \( B_1 \) and \( B_2 \). \( \square \)

### 6 Specific Algorithms

In this section, we derive several specific two-step algorithms from the iterative scheme (17) by choosing specific sets of \( (n + m) \times (n + m) \) matrices \( \mathcal{M} := \{ M_0, M_1, M_2 \} \) which satisfy Condition-M.

#### 6.1 First-Order Primal-Dual Algorithms

In this subsection, we design a class of explicit one-step algorithms, which only utilize the vectors of the current step to update the vectors of the next step. In such case, \( M_2 = 0 \) and Condition-M reduces to \( M_0 = M_1 \) and \( M_0 \in S^{n+m}_+ \).

We begin with constructing \( M_0 \). If the matrix \( E - R^{-1} M_0 \) is strictly upper or lower triangular, then the resulting algorithms will be explicit. By (15), \( M_0 = M_1 \) can be chosen as \( Z_1 \) or \( Z_2 \) with
\[
Z_1 := \begin{bmatrix} P & -A^T \\ -A & 1/\bar{p} I \end{bmatrix}, \quad Z_2 := \begin{bmatrix} P A^T \\ A & 1/\bar{p} I \end{bmatrix},
\]
where \( P \) is defined by (13). By simple calculations, one can obtain that \( \iota_\nu^\mu(\cdot) = \langle \cdot, b \rangle \) and thus \( \text{prox}_{\iota_\nu^\mu}(y) = y - \beta b \) for \( y \in \mathbb{R}^m \). Then, iterative scheme (17) with respect to \( Z_1 \) and \( Z_2 \) become, respectively,

\[
\begin{align*}
    x_i^{k+1} &= \text{prox}_{\nu_i^\mu f_i} \left( x_i^k - \frac{\alpha_i}{\beta} A_i^T y^k \right), \quad i \in \mathbb{N}_s, \\
    y^{k+1} &= y^k + \beta \left( \sum_{i=1}^s A_i \left( 2x_i^{k+1} - x_i^k \right) - b \right),
\end{align*}
\]

(38)

and

\[
\begin{align*}
    y^{k+1} &= y^k + \beta \left( \sum_{i=1}^s A_i x_i^k - b \right), \\
    x_i^{k+1} &= \text{prox}_{\nu_i^\mu f_i} \left( x_i^k - \frac{\alpha_i}{\beta} A_i^T (2y^{k+1} - y^k) \right), \quad i \in \mathbb{N}_s.
\end{align*}
\]

(39)

We note that, algorithms (38) and (39) are actually special cases of the one-step first-order primal-dual algorithm [6,13,20], which solves the following optimization problem

\[
\min \{ f(x) + g(Ax) : x \in \mathbb{R}^n \}
\]

(40)

with \( f \in \Gamma_0(\mathbb{R}^n), g \in \Gamma_0(\mathbb{R}^m) \) and \( A \) an \( m \times n \) matrix. Here, if we set \( x := (x_1, \ldots, x_s) \), \( g := \iota_C \) and \( f : \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_s} \to \mathbb{R} \) defined at \( x \) as \( f(x) := \sum_{i=1}^s f_i(x_i) \), then problem (40) is exactly the optimization problem (1). Clearly, algorithms (38) and (39) are special cases of the one-step first-order primal-dual algorithm [6,20] by the fact that \( \text{prox}_{f, \mu}(x) = (\text{prox}_{\nu_1^\mu f_1}(x_1), \ldots, \text{prox}_{\nu_s^\mu f_s}(x_s)) \). The corresponding convergence results are presented in the following theorem.

**Theorem 6** Let \( f_i \in \Gamma_0(\mathbb{R}^{n_i}), A_i \) an \( m \times n \) matrix, \( \alpha_i > 0 \) for \( i \in \mathbb{N}_s \) and \( \beta > 0 \). Let the sequence \( \{ (x_1^k, \ldots, x_s^k, y^k) : k \in \mathbb{N} \} \) generated from (38) or (39) for any \( (x_1^0, \ldots, x_s^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m \). If \( \|AQ\|_2 < 1 \), where \( Q := \text{diag}(\sqrt{\alpha_1}1_{n_1}, \ldots, \sqrt{\alpha_s}1_{n_s}) \), then \( \{ (x_1^k, \ldots, x_s^k, y^k) : k \in \mathbb{N} \} \) converges and the sequence \( \{ x^k : k \in \mathbb{N} \} \) converges to a solution of problem (1).

We omit the proof since it can be obtained immediately by applying Lemma 6.2 in [20] and Theorem 3.

To close this subsection, we remark that both algorithms (38) and (39) do not take advantage of the separability of function \( f \) and vector \( x \). More precisely, the information of \( x_j^{k+1} \) for \( j = 1, \ldots, i - 1 \) is not used when we update \( x_i^{k+1} \). We dedicate the next two subsections to developing new algorithms which make use of the block-wise Gauss-Seidel technique to update blocks \( x_1, \ldots, x_2, y \).

### 6.2 Convergent Implicit Two-Step Proximity Algorithms

In this subsection, we propose a two-step implicit fixed-point proximity algorithm from iterative scheme (17). We begin with constructing the set of matrices \( M := \{ M_0, M_1, M_2 \} \) by setting
To this end, we introduce the augmented Lagrangian function for (1). The PADMM for (1) reads as iterative scheme (41), we obtain that

\[
M_0 := \begin{bmatrix}
\frac{\beta}{\alpha_1} I & -\beta A_1^\top A_2 & -\beta A_1^\top A_3 & \cdots & -\beta A_1^\top A_s & 0 \\
0 & \frac{\beta}{\alpha_2} I & -\beta A_2^\top A_3 & \cdots & -\beta A_2^\top A_s & 0 \\
0 & 0 & \frac{\beta}{\alpha_3} I & \cdots & -\beta A_3^\top A_s & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \frac{\beta}{\alpha_s} I & 0 \\
0 & 0 & 0 & \cdots & 0 & \frac{1}{\beta} I
\end{bmatrix},
\]

(41)

\[
M_1 := \begin{bmatrix}
\frac{\beta}{\alpha_1} I & -2\beta A_1^\top A_2 & -2\beta A_1^\top A_3 & \cdots & -2\beta A_1^\top A_s & 0 \\
0 & \frac{\beta}{\alpha_2} I & -2\beta A_2^\top A_3 & \cdots & -2\beta A_2^\top A_s & 0 \\
0 & 0 & \frac{\beta}{\alpha_3} I & \cdots & -2\beta A_3^\top A_s & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \frac{\beta}{\alpha_s} I & 0 \\
0 & 0 & 0 & \cdots & 0 & \frac{1}{\beta} I
\end{bmatrix},
\]

(42)

\[
M_2 := \begin{bmatrix}
0 & \beta A_1^\top A_2 & \beta A_1^\top A_3 & \cdots & \beta A_1^\top A_s & 0 \\
0 & 0 & \beta A_2^\top A_3 & \cdots & \beta A_2^\top A_s & 0 \\
0 & 0 & 0 & \cdots & \beta A_3^\top A_s & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0
\end{bmatrix},
\]

(43)

With this choice of matrices \(M_0, M_1, M_2\), noting that \(\text{prox}_{\beta \ell_1}(y) = y - \beta b\), iterative scheme (17) leads to

\[
y^{k+1} = y^k + \beta \left( \sum_{i=1}^{s} A_i x_i^{k+1} - b \right).
\]

We then replace \(y^{k+1}\) by \(y^k + \beta (\sum_{i=1}^{s} A_i x_i^{k+1} - b)\) as we update \(x_i^{k+1}\) for \(i \in \mathbb{N}_s\) in iterative scheme (17), we obtain that

\[
\begin{align*}
x_j^{k+1} &= \text{prox}_{\frac{\beta}{\alpha_j} f_j} \left( x_j^k - \alpha_j A_j^\top \left( \sum_{i=1}^{\ell} A_i x_i^{k+1} + \sum_{i=j+1}^{\ell} A_i \left( 2x_i^k - x_i^{k-1} \right) - b \right) - \frac{\alpha_j}{\beta} A_j^\top y^k \right), j \in \mathbb{N}_s, \\
y^{k+1} &= y^k + \beta \left( \sum_{i=1}^{s} A_i x_i^{k+1} - b \right).
\end{align*}
\]

(44)

We point out the connections of the proposed algorithm (44) with the proximal ADMM (PADMM). To this end, we introduce the augmented Lagrangian function for (1)

\[
\mathcal{L}(x_1, \ldots, x_s, y) := \sum_{i=1}^{s} f_i(x_i) + \frac{\beta}{2} \left\| \sum_{i=1}^{s} A_i x_i - b \right\|_2^2 + \left\langle y, \sum_{i=1}^{s} A_i x_i - b \right\rangle.
\]

(45)

The PADMM for (1) reads as

\[
\begin{align*}
x_j^{k+1} &= \arg \min \left\{ \mathcal{L}(x_1^{k+1}, \ldots, x_j^{k+1}, x_j, x_{j+1}^{k+1}, \ldots, x_s^{k+1}, y^k) + \frac{\beta}{2\alpha_j} \|x_j - x_j^k\|_2^2 : x_j \in \mathbb{R}^n \right\}, j \in \mathbb{N}_s, \\
y^{k+1} &= y^k + \beta \left( \sum_{i=1}^{s} A_i x_i^{k+1} - b \right).
\end{align*}
\]

(46)
On the other hand, by the definition of proximity operator (2), the proposed algorithm (44) can be equivalently rewritten as
\[
\begin{align*}
    x_j^{k+1} &= \arg\min \left\{ L \left( x_1^{k+1}, \ldots, x_j^{k+1}, x_j, \tilde{x}_{j+1}^k, \ldots, \tilde{x}_n^k, y^k \right) + \frac{\beta}{2\alpha_i} \| x_j - x_j^k \|^2 : x_j \in \mathbb{R}^{n_j} \right\}, \\
    y^{k+1} &= y^k + \beta \left( \sum_{i=1}^s A_i x_i^{k+1} - b \right), \\
    \tilde{x}_j^{k+1} &= 2x_j^{k+1} - x_j^k, \\
    j &\in \mathbb{N}_s.
\end{align*}
\]
(47)

We can observe that our proposed algorithm (44) reduces to the PADMM if we set \( x_j^{k+1} = x_j^{k+1} \) for \( j \in \mathbb{N}_s \) in (47). As shown in [9], convergence of ADMM directly applied to problem (1) with \( s \geq 3 \) is not guaranteed. Also, it was shown in [18] that PADMM may not converge unless extra assumptions on \( f_i \) for \( i \in \mathbb{N}_s \) are added. However, algorithm (47) is ensured to converge without extra assumptions on \( f_i \) for \( i \in \mathbb{N}_s \). We next establish the convergence result of algorithm (44).

**Proposition 2** Let \( M_0, M_1, M_2 \) be defined as (41), (42), and (43). Let \( \tilde{M}_2 := \frac{1}{\beta} M_2 \). If
\[
0 < \alpha_i < \frac{1}{2\| M_2 \|^2}, \quad \text{for } i \in \mathbb{N}_s \quad \text{and} \quad \beta > 0,
\]
(48)
then the set \( \{ M_0, M_1, M_2 \} \) satisfies Condition-M.

**Proof** Clearly, we see that \( M_0 = M_1 + M_2 \), that is, Item (i) of Condition-M holds. Define \( H := M_0 + M_2 \). Then \( H = \text{diag}(\frac{\beta}{\alpha_1} 1_{n_1}, \ldots, \frac{\beta}{\alpha_s} 1_{n_s}, 1/\beta 1_m) \) is diagonal and symmetric. Item (ii) of Condition-M is trivial due to \( \alpha_i > 0 \) for \( i \in \mathbb{N}_s \) and \( \beta > 0 \). We then prove the validity of Item (iii) of Condition-M. Since the last \( m \) columns and rows of \( M_2 \) are all zeros, we have that
\[
\| H^{-1/2} M_2 H^{-1/2} \|_2 = \| \tilde{H}^{-1/2} \tilde{M}_2 \tilde{H}^{-1/2} \|_2,
\]
where \( \tilde{H} := \text{diag}(\frac{\beta}{\alpha_1} 1_{n_1}, \ldots, \frac{\beta}{\alpha_s} 1_{n_s}, 0_m) \). By using hypothesis (48), we find that
\[
\| \tilde{H}^{-1/2} \tilde{M}_2 \tilde{H}^{-1/2} \|_2 \leq \max \{ \alpha_i : i \in \mathbb{N}_s \} \| \tilde{M}_2 \|_2 < \frac{1}{2},
\]
which leads to Item (iii) of Condition-M. \( \square \)

The convergence results of algorithm (44) is presented below.

**Theorem 7** Let \( f_i \in L_0(\mathbb{R}^{n_i}), A_i \) an \( m \times n_i \) matrix, \( \alpha_i > 0 \) for \( i \in \mathbb{N}_s \) and \( \beta > 0 \). Let the sequence \( \{ (x^k, y^k) : k \in \mathbb{N} \} \) be generated from the algorithm (47) for any \((x^0, y^0), (x^1, y^1) \in \mathbb{R}^n \times \mathbb{R}^m \). Let \( M_2 \) be defined as (43) and \( \tilde{M}_2 = \frac{1}{\beta} M_2 \). If the condition (48) is satisfied, then the sequence \( \{ x^k : k \in \mathbb{N} \} \) converges to a solution of problem (1).

**Proof** By Proposition 2 and Theorem 3, it suffices to prove \( T_M \) is well-defined when \( M_0, M_1, M_2 \) are defined by (41), (42), and (43). In this case, if \( w = T_M(u, v) \), where \( w := (w_1, \ldots, w_s, w_y), u := (u_1, \ldots, u_s, u_y), v := (v_1, \ldots, v_s, v_y) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_s} \times \mathbb{R}^m \), then \( w_y = u_y + \sum_{j=1}^s A_j w_j \) and each \( w_i \) for \( i \in \mathbb{N}_s \) can be calculated by
\[
w_i = \arg\min \left\{ f_i(x_i) + \frac{\beta}{2} \sum_{j=1}^{i-1} A_j w_j + A_i x_i + \sum_{j=i+1}^s A_j (2u_j - v_j) - b \right\}^2_2 + \{ u_y, A_i x_i \}
\]
\[
+ \frac{\beta}{2\alpha_i} \| x_i - u_i \|_2^2 : x_i \in \mathbb{R}^{n_i} \right\}.
\]
Since the objective function of the above optimization problem is strongly convex, \( T_M \) is well-defined.

To end this subsection, we point out that compared with algorithms (38) and (39), algorithm (44) takes advantage of the separable structure of variable \( x \) and applies the block-wise Gauss-Seidel technique to blocks \( x_1, \ldots, x_s, y \). We also note that solving the subproblems involved in (44) may require inner iterations. In practice, it will affect the computational efficiency of the algorithm (44). In the next subsection, we develop an explicit two-step algorithm. As long as the proximity operators of \( f_i \) for \( i \in \mathbb{N}_s \) have closed form solutions, the algorithm can be implemented efficiently.

### 6.3 Convergent Explicit Two-Step Proximity Algorithms

In this subsection, we propose a class of explicit algorithms, which apply the block-wise Gauss-Seidel technique to blocks \( x_1, \ldots, x_s, y \).

We begin with specifying the set of matrices \( M \). We set

\[
M_0 := \begin{bmatrix}
\frac{\beta}{\alpha_1} I - \beta A_1^T A_1 & -\beta A_1^T A_2 & \cdots & -\beta A_1^T A_s & 0 \\
0 & \frac{\beta}{\alpha_2} I - \beta A_2^T A_2 & \cdots & -\beta A_2^T A_s & 0 \\
0 & 0 & \frac{\beta}{\alpha_3} I - \beta A_3^T A_3 & \cdots & -\beta A_3^T A_s & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & \frac{\beta}{\alpha_s} I - \beta A_s^T A_s & 0 \\
0 & 0 & 0 & \ldots & 0 & \frac{1}{\beta} I
\end{bmatrix}, \quad (49)
\]

and let \( M_2 \) be defined as in (43). We can obtain an implicit algorithm by directly substituting (49), (50), and (43) into the iterative scheme (17). As the same as the algorithm (44), it implies

\[
y^{k+1} = y^k + \beta \left( \sum_{i=1}^s A_i x_i^{k+1} - b \right).
\]

As in Sect. 6.2, we replace \( y^{k+1} \) by \( y^k + \beta (\sum_{i=1}^s A_i x_i^{k+1} - b) \) when we update \( x_i^{k+1} \) for \( i \in \mathbb{N}_s \). This leads to the following explicit algorithm

\[
(2SFPPA) \begin{cases}
x_j^{k+1} = \text{prox}_{\frac{\alpha}{\beta} f_j} \left( x_j^k - \alpha_j A_j^T \left( \sum_{i=1}^{j-1} A_i x_i^{k+1} + A_j x_j^k \right) + \sum_{i=j+1}^s A_i \left( 2x_i^k - x_i^{k-1} \right) - \frac{\alpha_j}{\beta} A_j^T y^k \right), \quad j \in \mathbb{N}_s, \\
y^{k+1} = y^k + \beta \left( \sum_{i=1}^s A_i x_i^{k+1} - b \right).
\end{cases} \quad (51)
\]

We point out here the relationship between the proposed algorithm (51) and the LADMM. To this end, we first review the exact extension of LADMM to problem (1). For \( j \in \mathbb{N}_s \), let
is guaranteed. Next we present the convergence results of the algorithm (51). As mentioned in [17], the direct extension of LADMM to the multi-block problem (1) is not necessarily convergent. Nevertheless, the convergence of the proposed algorithm (51) is guaranteed. Next we present the convergence results of the algorithm (51).

**Proposition 3** Let $M_0$, $M_1$ and $M_2$ be defined as in (49), (50), and (43). Let $\tilde{M}_2 := \frac{1}{\beta} M_2$. If

\[ 0 < \alpha_i < \frac{1}{\|A_i\|_2^2 + 2\|M_2\|_2}, \quad \text{for } i \in \mathbb{N}_s \quad \text{and} \quad \beta > 0, \tag{54} \]

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

**Proof** It is clear that Item (i) of Condition-M is satisfied. Define $H := M_0 + M_2$. Then $H = \text{diag}(\frac{\beta}{\alpha_1} I - \beta A_1^T A_1, \ldots, \frac{\beta}{\alpha_s} I - \beta A_s^T A_s, 1/\beta I)$ is symmetric. In light of (54), we have that $H \in \mathbb{S}^{n_s + m}$. Hence, Item (ii) of Condition-M holds. We next show Item (iii) of Condition-M. Similar to the proof of Proposition 2,

\[ \|H^{-1/2} M_2 H^{-1/2}\|_2 = \|\tilde{H}^{-1/2} M_2 \tilde{H}^{-1/2}\|_2, \]

where $\tilde{H} := \text{diag}(\frac{\beta}{\alpha_1} I - \beta A_1^T A_1, \ldots, \frac{\beta}{\alpha_s} I - \beta A_s^T A_s, 0)$. Hypothesis (54) leads to

\[ \|\tilde{H}^{-1/2} M_2 \tilde{H}^{-1/2}\|_2 \leq \max \left\{ \frac{1}{\alpha_i} - \|A_i\|_2^2 : i \in \mathbb{N}_s \right\} \|M_2\|_2 < \frac{1}{2}. \]

This completes the proof. \hfill \Box

The following theorem regards the convergence of algorithm (51).

**Theorem 8** Let $f_i \in \Gamma_0(\mathbb{R}^{n_i})$, $A_i$ an $m \times n_i$ matrix, $\alpha_i > 0$ for $i \in \mathbb{N}_s$ and $\beta > 0$. Let the sequence \{ $(x^k, y^k) : k \in \mathbb{N}$ \} be generated from algorithm (51) for any $(x^0, y^0), (x^1, y^1) \in \mathbb{R}^n \times \mathbb{R}^m$. Let $M_2$ be defined as (43) and $\tilde{M}_2 := \frac{1}{\beta} M_2$. If condition (54) is satisfied, then the sequence \{ $x^k : k \in \mathbb{N}$ \} converges to a solution of problem (1).

**Proof** By Theorem 3 and Proposition 3, we only need to prove $T_M$ is well-defined. In this case, from algorithm (51), it is obvious that $T_M$ can be computed explicitly. Therefore $T_M$ is well-defined. \hfill \Box
To close this subsection, we remark that when the proximity operators of $f_i$ for $i \in \mathbb{N}_s$ have closed form solutions, the two-step algorithm (51) may be more efficient than the two-step algorithm (44). This is because the two-step algorithm (44) may require inner iterations to solve the subproblems involved, while each step of algorithm (51) can be implemented efficiently by making use of the closed form.

6.4 Variants of Algorithms (44) and (51)

There is a wide variety of the choices of $\{M_0, M_1, M_2\}$ satisfying condition-M, including those of algorithms (44) and (51). In this subsection, we present other choices of $\{M_0, M_1, M_2\}$ satisfying condition-M. With these choices the two step iterative scheme (17) reduces to a class of new algorithms, which can be viewed as variants of algorithms (44) and (51).

**Modifications of diagonal blocks:** The diagonal blocks of $M_1$ and $M_2$ can be chosen in other ways. We only present two examples in the following. For instance, the diagonal entries of $M_1$ in (42) can be chosen as $((\theta + 1) \frac{\beta}{\alpha_1} \mathbf{1}_{n_1}, \ldots, (\theta + 1) \frac{\beta}{\alpha_s} \mathbf{1}_{n_s}, \frac{1}{\beta} \mathbf{1}_m)$ with $\theta \in [0, 1)$ and correspondingly, the diagonal entries of $M_2$ in (43) should be $(-\theta \frac{\beta}{\alpha_1} \mathbf{1}_{n_1}, \ldots, -\theta \frac{\beta}{\alpha_s} \mathbf{1}_{n_s}, \mathbf{0}_m)$. With such a choice of $\{M_0, M_1, M_2\}$, iterative scheme (17) reduces to a variant of algorithm (44)

$$
\begin{align*}
    x_{j}^{k+1} &= \text{prox}_{\frac{\beta}{\alpha_j} f_j} (x_j^k + \theta \left(x_j^k - x_{j}^{k-1}\right) - \alpha_j A_j^\top \left(\sum_{i=1}^{j} A_i x_i^{k+1}\right) + \sum_{i=j+1}^{s} A_i \left(2x_i^k - x_i^{k-1}\right) - b) \frac{\alpha_j}{\beta} A_j^\top y^k), \quad j \in \mathbb{N}_s, \\
    y_{k}^{+1} &= y^k + \beta \left(\sum_{i=1}^{s} A_i x_i^{k+1} - b\right).
\end{align*}
$$

As a second example, the diagonal blocks of $M_1$ in (50) can be chosen as $(\frac{\beta}{\alpha_1} I - 2\beta A_1^\top A_1, \ldots, \frac{\beta}{\alpha_s} I - 2\beta A_s^\top A_s, \frac{1}{\beta} I)$. Accordingly, the diagonal blocks of $M_2$ in (43) should be $(\beta A_1^\top A_1, \ldots, \beta A_s^\top A_s, \mathbf{0})$ to make $M_0 = M_1 + M_2$. These matrices lead to a variant of algorithm (51)

$$
\begin{align*}
    x_{j}^{k+1} &= \text{prox}_{\frac{\beta}{\alpha_j} f_j} (x_j^k - \alpha_j A_j^\top \left(\sum_{i=1}^{j-1} A_i x_i^{k+1}\right) + \sum_{i=j}^{s} A_i \left(2x_i^k - x_i^{k-1}\right) - b) \frac{\alpha_j}{\beta} A_j^\top y^k), \quad j \in \mathbb{N}_s, \\
    y_{k}^{+1} &= y^k + \beta \left(\sum_{i=1}^{s} A_i x_i^{k+1} - b\right).
\end{align*}
$$

**Modifications of nondiagonal blocks:** We change the $(i, j)$-th block of $M_0$ (defined by (41) or (49)) for $i > j$ from 0 to $\theta \beta A_i^\top A_j$ and keep other blocks of $M_0$ unchanged. In order to make $M_0 + M_2$ symmetric, the matrix $M_2$ should be chosen as $\theta + 1$ multiplying the original matrix $M_2$ defined in (43). Accordingly, the matrix $M_1$ can be determined by $M_1 = M_0 - M_2$. Then we can derive the following two algorithms from iterative scheme (17)

$$
\begin{align*}
    x_{j}^{k+1} &= \text{prox}_{\frac{\beta}{\alpha_j} f_j} (x_j^k - \alpha_j A_j^\top \left(\sum_{i=1}^{j-1} A_i x_i^{k+1}\right) + \theta \left(x_j^{k+1} - x_j^k\right)) + A_j x_j^{k+1} + \sum_{i=j+1}^{s} A_i \left((2 + \theta)x_i^k - (\theta + 1)x_i^{k-1}\right) - b) \frac{\alpha_j}{\beta} A_j^\top y^k), \quad j \in \mathbb{N}_s, \\
    y_{k}^{+1} &= y^k + \beta \left(\sum_{i=1}^{s} A_i x_i^{k+1} - b\right),
\end{align*}
$$
Hybrids of both algorithms: Both algorithms (44) and (51) share the same matrix $M_2$. Matrices $M_0$ for algorithms (44) and (51) are almost the same except the diagonal blocks. Let $S_1 \subseteq \mathbb{N}_s$ and $S_2 = \mathbb{N}_s \setminus S_1$. Suppose the subproblems of (47) for $x_i^{k+1}$, $i \in S_1$ can be solved efficiently. We also assume inner iterations are required to solve the subproblems of (47) for $x_i^{k+1}$, $i \in S_2$. We set the $i$-th diagonal block of $M_0$ to be $\frac{\beta}{\alpha_i} I$ for $i \in S_1$ and to be $\frac{\beta}{\alpha_i} I - \beta A_i^T A_i$ for $i \in S_2$. The nondiagonal blocks of $M_0$ are chosen to be the same as in (41) and (49). We further choose the matrix $M_2$ as in (43). Accordingly, the matrix $M_1$ is determined by $M_1 = M_0 - M_2$. Then we obtain the following hybrid algorithm

$$
\begin{align*}
x_j^{k+1} & = \prox_{\frac{\alpha_j}{\beta} f_j} (x_j^k - \alpha_j A_j^T (\sum_{i=1}^j A_i x_i^{k+1} \\
& + \sum_{i=j+1}^s A_i (2x_i^k - x_i^{k-1}) - b) - \frac{\alpha_j}{\beta} A_j^T y^k), \quad j \in S_1, \\
x_j^{k+1} & = \prox_{\frac{\alpha_j}{\beta} f_j} (x_j^k - \alpha_j A_j^T (\sum_{i=1}^{j-1} A_i x_i^{k+1} + A_j x_j^k \\
& + \sum_{i=j+1}^s A_i (2x_i^k - x_i^{k-1}) - b) - \frac{\alpha_j}{\beta} A_j^T y^k), \quad j \in S_2, \\
y^{k+1} & = y^k + \beta \left( \sum_{i=1}^s A_i x_i^{k+1} - b \right). 
\end{align*}
$$

We point out here that convergence of the above algorithms is guaranteed. One can obtain the convergence results by verifying that the corresponding set of matrices $\mathcal{M} := \{M_0, M_1, M_2\}$ satisfies Condition-M and $T_{\mathcal{M}}$ is well-defined. We omit the details here since the proofs are similar to those of algorithms (44) and (51).

### 7 Numerical Experiments

In this section, we demonstrate the efficiency of the proposed two-step fixed-point proximity algorithms by applying 2SFPPA to the sparse Magnetic Resonance Imaging (MRI) reconstruction problem [24]. We shall compare performance of the proposed 2SFPPA with that of other LADMM-type algorithms.

#### 7.1 Sparse MRI problem

For convenience of exposition, we assume that an image considered has a size of $d_1 \times d_2$. The image is treated as a vector in $\mathbb{R}^{d_1 d_2}$ in such a way its $(i, j)$-th pixel corresponds to the $(i + (j - 1)d_2)$-th component of the vector in $\mathbb{R}^{d_1 d_2}$. We set $d := d_1 d_2$. Let $K \in \mathbb{R}^{p \times d}$ ($p < d$) be a partial Fourier transform matrix and $b \in \mathbb{R}^p$ represent the observed data. Then the general form of the sparse MRI reconstruction model can be written as

$$
\min \{ F(u) : u \in \mathbb{R}^d, Ku = b \},
$$

where $F(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is a sparse-promoting function. It is well-known that a superior reconstructed image can be obtained when $F(\cdot)$ is chosen to be a hybrid of the total variation and the $\ell_1$-norm of the Haar wavelet transform. Denote by $W \in \mathbb{R}^{q \times d}$ the Haar wavelet transform matrix and define the $q \times q$ diagonal matrix $A := \text{diag}(\lambda_1, \ldots, \lambda_q)$ with $\lambda_i \geq \cdots \geq \lambda_q$. Springer
0, \ i \in \mathbb{N}_q. We turn to considering the following specific sparse MRI problem
\[
\min \{ \mu \| u \|_{TV} + \| A W u \|_1 : u \in \mathbb{R}^d, Ku = b \},
\]  
(57)
where \( \mu > 0 \) trades the total variation with sparsity of the wavelet coefficients \( Wu \).

In order to apply the proposed algorithms, we need to reformulate problem (57). First, we rewrite \( \| \cdot \|_{TV} \) as a function composed with a linear mapping. To this end, we recall the \( r \times r \) difference matrix \( D_r := \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ \vdots & \vdots \\ -1 & 1 \end{bmatrix} \).
(58)
Through the matrix Kronecker product \( \otimes \), we define the \( 2d \times d \) matrix \( B \) by
\[
B := \begin{bmatrix} I_{d_2} \otimes D_{d_1} \\ D_{d_2} \otimes I_{d_1} \end{bmatrix}.
\]  
(59)
Moreover, we define function \( \psi : \mathbb{R}^{2d} \to \mathbb{R} \) at \( y \in \mathbb{R}^{2d} \) as
\[
\psi(y) := \sum_{i=1}^{d} \| [y_i, y_{d+i}]^\top \|_2.
\]  
(60)
With the definition of matrix \( B \) (59) and the convex function \( \psi \) (60), the (isotropic) total variation of an image \( x \) can be represented by
\[
\| x \|_{TV} = \psi(Bx).
\]  
(61)
Moreover, we define \( \varphi : \mathbb{R}^q \to \mathbb{R} \) at \( y \in \mathbb{R}^q \) as \( \varphi(y) := \| Ay \|_1 \). Then with help of the formula (61), function \( \varphi \) and the indicator function \( \iota_{\{b\}} \), problem (57) can be equivalently reformulated as
\[
\min \{ \mu \psi(Bu) + \varphi(Wu) + \iota_{\{b\}}(Ku) : u \in \mathbb{R}^d \}.
\]  
(62)
Recall that the dual problem of (62) has the form
\[
\min \{ (\mu \psi)^*(x_1) + \varphi^*(x_2) + \iota_{\{b\}}^*(x_3) : B^\top x_1 + W^\top x_2 + K^\top x_3 = 0, \\
x_1 \in \mathbb{R}^{2d}, x_2 \in \mathbb{R}^q, x_3 \in \mathbb{R}^p \}.
\]  
(63)
By the definition of the Fenchel conjugate function, one can easily check that the Fenchel conjugate functions in (63) have the form
\[
(\mu \psi)^* = \iota_{S_1}, \quad \varphi^* = \iota_{S_2}, \quad \text{and} \quad \iota_{\{b\}}^*(\cdot) = \langle b, \cdot \rangle,
\]
where the sets \( S_1 \subseteq \mathbb{R}^{2d} \) and \( S_2 \subseteq \mathbb{R}^q \) are defined as
\[
S_1 := \{ \| [y_i, y_{d+i}] \|_2 \leq \mu, \forall i \in \mathbb{N}_d : y \in \mathbb{R}^{2d} \}
\]
and
\[
S_2 := \{ |y_j| \leq \lambda_j, \forall j \in \mathbb{N}_q : y \in \mathbb{R}^q \}.
\]
Therefore, we obtain the following minimization problem
\[
\min \{ \iota_{S_1}(x_1) + \iota_{S_2}(x_2) + \langle b, x_3 \rangle : B^\top x_1 + W^\top x_2 + K^\top x_3 = 0, x_1 \in \mathbb{R}^{2d}, x_2 \in \mathbb{R}^q, x_3 \in \mathbb{R}^p \}.
\]  
(64)
Obviously, problem (64) is a special case of the multi-block problem (1) with the block number \( s = 3 \). Thus we can directly apply 2SFPPA to solving problem (64). In particular, all the proximity operators of the convex functions involved in (64) have closed forms. More precisely, the proximity operators \( \text{prox}_{\frac{\alpha_1}{\rho}S_1} \) and \( \text{prox}_{\frac{\alpha_2}{\rho}S_2} \) are exactly the projection operator onto the sets \( S_1 \) and \( S_2 \) respectively. The proximity operator \( \text{prox}_{\frac{\alpha_3}{\rho}L} \) is just the shift operator. We describe the 2SFPPA for the sparse MRI model in Algorithm 1.

### Algorithm 1 (2SFPPA for the sparse MRI)

1: Given: observed data \( b \) in \( \mathbb{R}^p; \mu > 0, \Lambda \geq 0, \alpha_1, \alpha_2, \alpha_3 > 0 \) and \( \beta > 0 \)
2: Initialization: \( x_1^0 = K^\top b, x_2^0 = x_2^{-1} = 0, x_3^0 = x_3^{-1} = 0, \gamma_0 = 0. \)
3: repeat
4: \hspace{0.5cm} Step 1: \( x_1^{k+1} \leftarrow \text{Proj}_{S_1}(x_1^k - \alpha_1 B(B^\top x_1^k + W^\top (2x_2^k - x_2^{k-1}) + K^\top (2x_3^k - x_3^{k-1}) + \frac{1}{\rho} y^k)) \)
5: \hspace{0.5cm} Step 2: \( x_2^{k+1} \leftarrow \text{Proj}_{S_2}(x_2^k - \alpha_3 W(B^\top x_1^{k+1} + W^\top x_2^k + K^\top (2x_3^k - x_3^{k-1}) + \frac{1}{\rho} y^k)) \)
6: \hspace{0.5cm} Step 3: \( x_3^{k+1} \leftarrow \text{Proj}_{S_3}(x_3^k - \alpha_3 K(B^\top x_1^{k+1} + W^\top x_2^{k+1} + K^\top x_3^k + \frac{1}{\rho} y^k) - \frac{\alpha_3}{\rho} b) \)
7: \hspace{0.5cm} Step 4: \( y^{k+1} \leftarrow y^k + \beta(B^\top x_1^{k+1} + W^\top x_2^{k+1} + K^\top x_3^{k+1}) \)
8: until “convergence”
9: Write the output of \( -y^k \) from the above loop as \( u^\infty \).

### 7.2 Numerical Results

In this subsection, we compare numerical results of the proposed 2SFPPA with those of the Jacobi-type LADMM (JADMM) (39), the LADMM and LADMM with Gaussian back substitution (LADMMG) for the sparse MRI problem. All the experiments are conducted in Matlab 7.6 (R2008a) installed on a laptop with Intel Core i5 CPU at 2.5GHz, 8G RAM running Windows 7.

In the experiment, we select the \( 256 \times 256 \) “Shepp-Logan” phantom as the test image, see Fig.1a. The observed data \( b \) is obtained by sampling the discrete Fourier transform of the phantom along 17 pseudo-radial lines, as shown in Fig.1b. The Haar wavelet transform \( W \in \mathbb{R}^{p \times d} \) is chosen to be non-decimated and thus we have that \( p = 4d \). We assume that the upper \( d \times d \) sub-matrix of \( W \) is formed by the low-pass filter while the remaining \( 3d \times d \) sub-matrix is formed by the high-pass filters. Accordingly, we set the diagonal entries of the diagonal matrix \( \Lambda \) as follows

\[
\lambda_i = \begin{cases} 
0, & i \in \mathbb{N}_d, \\
\frac{1}{2}, & i \in \mathbb{N}_p \backslash \mathbb{N}_d.
\end{cases}
\]

We further take the regularization parameters \( \mu = 3 \) throughout the test. We measure the computational efficiency of the compared algorithms by two criteria. One criterion is the relative error between values of the objective function at each iteration and the optimal function value of problem (62). We remark that the indicator function \( \iota(b) \) is involved in the objective function and the iterates \( u^k = -y^k \) may not always satisfy \( Ku^k = b \). Therefore, for fair numerical comparisons we compute the following relative error

\[
\epsilon_1^k := \left( F(u^k) + \tau \| Ku^k - b \|_2 - F^* \right) / F^*
\]
where $\tau > 0$ is a penalty parameter and $F^*$ denotes the optimal function value. In practice, we set $\tau = 1000$ and run the LADMM for 5000 iterations to obtain an approximation of $F^*$. The other one is that the relative error between two successive iterates

$$
\varepsilon^k_2 := \frac{\|y^k - y^{k-1}\|_2}{\|y^k\|_2}.
$$

The quality of the reconstructed image is evaluated in terms of the peak signal-to-noise ratio (PSNR) defined by

$$
\text{PSNR} = 10\log_{10} \frac{255\sqrt{d}}{\|u^\infty - u^*\|_2} \, \text{(dB)},
$$

where $u^*$ is the original image vector and $u^\infty$ is the recovered image vector.

For the JLADMM, we set

$$
\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{8} \quad \text{and} \quad \beta = 1.
$$

For the LADMMG, LADMM, and 2SFPPA, we set

$$
\alpha_1 = \frac{1}{8}, \alpha_2 = \frac{0.999999}{\|W\|_2^2}, \alpha_3 = \frac{0.999999}{\|K\|_2^2}, \quad \text{and} \quad \beta = 1.
$$

Besides, as suggested in [17], the parameter $\theta$ involved in LADMMG is set to be 1. With such a choice of parameters, all the four algorithms achieve their best performance in terms of the convergence speed.

Tables 1 and 2 summarize the numbers of iterations, PSNR values and CPU times when the three algorithms achieve the given accuracy. We observe that the proposed 2SFPPA performs slightly better than LADMM and much better than JLADMM and LADMMG in terms of computational time. The LADMMG costs much more CPU time than both of LADMM and 2SFPPA, due to its Gaussian back substitution step which ensures its convergence. The evolution of the objective function values and PSNR values with respect to the CPU time and the number of iterations are shown in Fig. 2. The sequence of function values from 2SFPPA decreases faster to the minimum value than those from JLADMM and LADMMG. Similarly, the sequence of PSNR values from 2SFPPA grows faster to the maximum value than those from JLADMM and LADMMG. We conclude that overall 2SFPPA performs as efficiently as LADMM and much better than JLADMM and LADMMG.
Table 1  Performance comparison for the sparse MRI

|       | $\epsilon = 10^{-4}$ | $\epsilon = 10^{-5}$ | $\epsilon = 10^{-6}$ |
|-------|----------------------|----------------------|----------------------|
| JLADMM| (3410, 65.68, 178.79)| (−, −, −)           | (−, −, −)           |
| LADMMG| (1237, 63.71, 119.59)| (3667, 69.43, 363.04)| (−, −, −)           |
| LADMM | (1140, 63.63, 64.46) | (3452, 69.29, 201.41)| (4778, 70.88, 279.21)|
| 2SFPPA | (1026, 63.61, 60.12) | (3175, 69.18, 184.92)| (4455, 70.78, 259.99)|

For a given error tolerance $\epsilon$, the first column in the bracket represents the first iteration number $k$ such that $\epsilon^k_1 < \epsilon$, the second column and the third column in the bracket show the corresponding PSNR and CPU time.

Table 2  Performance comparison for the sparse MRI

|       | $\epsilon = 5 \times 10^{-5}$ | $\epsilon = 5 \times 10^{-6}$ | $\epsilon = 5 \times 10^{-7}$ |
|-------|--------------------------------|--------------------------------|--------------------------------|
| JLADMM| (646, 55.15, 33.15)          | (1388, 59.97, 71.79)          | (3416, 65.69, 179.15)          |
| LADMMG| (473, 57.68, 46.57)          | (928, 62.05, 89.70)           | (2451, 67.37, 242.11)          |
| LADMM | (468, 58.18, 26.76)          | (920, 62.44, 51.77)           | (2366, 67.42, 137.29)          |
| 2SFPPA | (438, 58.26, 26.22)          | (909, 62.98, 53.50)           | (2305, 67.66, 133.94)          |

For a given error tolerance $\epsilon$, the first column in the bracket represents the first iteration number $k$ such that $\epsilon^k_2 < \epsilon$, the second column and the third column in the bracket show the corresponding PSNR and CPU time.

Fig. 2  a PSNR versus computational time, b objective function value versus computational time, c PSNR versus number of iterations, d objective function value versus number of iterations
8 Conclusions

We develop in this paper two-step fixed-point iterative schemes for solving the multi-block separable convex optimization problem, which minimizes the sum of several convex functions with linear constraints. We prove convergence of the iterative schemes and establish their linear convergence rate in the ergodic sense and in the sense of the partial primal-dual gap. Based on the iterative schemes, we propose a class of convergent two-step algorithms for the multi-block separable convex optimization problem. Convergence analysis for the specific algorithms can be carried out by verifying conditions on the matrices used to construct the algorithms. We demonstrate the accuracy and computational efficiency of the proposed two-step algorithms by applying them to the sparse MRI problems. Numerical results show that the proposed algorithms perform as efficiently as LADMM (which is not guaranteed to converge) and outperform the JLADMM and LADMMG. We conclude that fixed-point formulation via proximity operators is a convenient approach for developing convergent and efficient algorithms for solving non-smooth convex optimization problems.

References

1. Attouch, H., Briceno-Arias, L.M., Combettes, P.L.: A parallel splitting method for coupled monotone inclusions. SIAM J. Control Optim. 48, 3246–3270 (2010)
2. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, AMS Books in Mathematics. Springer, New York (2011)
3. Cai, J., Chan, R., Shen, Z.: A framelet-based image inpainting algorithm. Appl. Comput. Harmonic Anal. 24, 131–149 (2007)
4. Cai, X., Han, D., Yuan, X.: On the convergence of the direct extension of ADMM for three-block separable convex minimization models with one strongly convex function. Comput. Optim. Appl. (2016). doi: 10.1007/s10589-016-9860-y
5. Cai, J., Osher, S., Shen, Z.: Linearized Bregman iteration for frame based image deblurring. SIAM J. Imaging Sci. 2, 226–252 (2009)
6. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. 40, 120–145 (2011)
7. Chan, R., Chan, T., Shen, L., Shen, Z.: Wavelet algorithms for high-resolution image reconstruction. SIAM J. Sci. Comput. 24, 1408–1432 (2003)
8. Chan, R., Riemenschneider, S.D., Shen, L., Shen, Z.: Tight frame: the efficient way for high-resolution image reconstruction. Appl. Comput. Harmonic Anal. 17, 91–115 (2004)
9. Chen, C., He, B., Ye, Y., Yuan, X.: The direct extension of admm for multi-block convex minimization problems is not necessarily convergent. Math. Program. 155, 57–79 (2016)
10. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. 20, 273–297 (1995)
11. Davis, D., Yin, W.: A three-operator splitting scheme and its optimization applications. UCLA CAM Report 15–13
12. Deng, W., Lai, M.-J., Peng, Z., Yin, W.: Parallel multi-block admm with o(1/k) convergence. UCLA CAM 13–64 (2014)
13. Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. 3, 1015–1046 (2010)
14. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. comput. math. appl. 2(1), 17–40. Comput. Math. Appl. 2, 17–40 (1976)
15. Goldstein, T., Osher, S.: The split Bregman method for l1 regularization problems. SIAM J. Imaging Sci. 2, 323–343 (2009)
16. He, B., Tao, M., Yuan, X.: Alternating direction method with gaussian back substitution for separable convex programming. SIAM J. Optim. 22, 313–340 (2012)
17. He, B., Yuan, X.: Linearized alternating direction method of multipliers with gaussian back substitution for separable convex programming. Numer. Algebra Control Optim. 22, 247–260 (2013)
18. Li, M., Sun, D., Toh, K.-C.: A convergent 3-block semi-proximal admm for convex minimization problems with one strongly convex block. Asia Pacific J. Oper. Res. 32, 1550024 (2015)
19. Li, Q., Micchelli, C.A., Shen, L., Xu, Y.: A proximity algorithm accelerated by Gauss-Seidel iterations for L1/TV denoising models. Inverse Probl. 28, 095003 (2012)
20. Li, Q., Shen, L., Yueheng, X., Zhang, N.: Multi-step fixed-point proximity algorithms for solving a class of convex optimization problems arising from image processing. Adv. Comput. Math. 41, 387–422 (2015)
21. Li, Q., Shen, L., Yang, L.: Split-bregman iteration for framelet based image inpainting. Appl. Comput. Harmonic Anal. 32, 145–154 (2012)
22. Li, Q., Zhang, N.: Fast proximity-gradient algorithms for structured convex optimization problems. Appl. Comput. Harmonic Anal. 41, 491–517 (2016)
23. Lin, T., Ma, S., Zhang, S.: On the convergence rate of multi-block admm. J. Oper. Res. Soc. China 3, 251–274 (2015)
24. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn. Reson. Med. 58, 1182–1195 (2007)
25. Micchelli, C.A., Shen, L., Xu, Y.: Proximity algorithms for image models: denoising. Inverse Probl. 27, 045009 (2011)
26. Moreau, J.J.: Fonctions convexes duales et points proximaux dans un espace hilbertien. C.R. Acad. Sci. Paris Sér. A Math. 255, 1897–2899 (1962)
27. Nikolova, M.: Local strong homogeneity of a regularized estimator. SIAM J. Appl. Math. 61, 633–658 (2000)
28. Rudin, L.I., Osher, S.: Total variation based image restoration with free local constraints, In: IEEE International Conference on Image Processing, pp. 31–35 (1994)
29. Ruszczynski, A.: Parallel decomposition of multistage stochastic programming problems. Math. Program. 58, 201–228 (1993)
30. Sawatzky, A., Qi, X., Schirra, C.O., Anastasio, M.A.: Proximal ADMM for multi-channel image reconstruction in spectral X-ray CT. IEEE Trans. Med. Imaging 33, 1657–1668 (2014)
31. Shi, W., Ling, W., Wu, W., Yin, W.: Extra: an exact first-order algorithm for decentralized consensus optimization. SIAM J. Optim. 25, 944–966 (2015)
32. Sun, D., Toh, K.C., Yang, L.: A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type of constraints. SIAM J. Optim. 25, 882–915 (2014)
33. Tyrrell Rockafellar, R.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. 14, 877–898 (1976)
34. Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K.: Sparsity and smoothness via the fused lasso. J. R. Stat. Soc. 67, 91–108 (2005)
35. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. 58, 267–288 (1996)
36. Wen, Z., Goldfarb, D., Yin, W.: Alternating direction augmented lagrangian methods for semidefinite programming. Math. Program. Comput. 2, 203–230 (2010)