

TWO-STEP FIXED-POINT PROXIMITY ALGORITHMS FOR MULTI-BLOCK SEPARABLE CONVEX PROBLEMS*

QIA LI [†], YUESHENG XU ^{†§}, AND NA ZHANG [‡]

Abstract. Multi-block separable convex problems recently received considerable attention. This class of optimization problems minimizes a separable convex objective function with linear constraints. The algorithmic challenges come from the fact that the classic alternating direction method of multipliers (ADMM) for the problem is not necessarily convergent. However, it is observed that ADMM 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 function. We then design a two-step fixed-point iterative scheme for solving these problems based on the characterization. We further prove convergence of the iterative scheme and show that it has $O(\frac{1}{k})$ 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 scheme and establish their global convergence. Numerical experiments for solving the sparse MRI problem demonstrate the numerical efficiency of the proposed 2SFPPA.

Key words. Multi-block separable convex problems, Fixed-point proximity algorithms, Two-step algorithms

AMS subject classifications. 90C25, 65K05

1. Introduction. We consider in this paper the convex minimization problem with linear constraints and a separable objective function in the form of the 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

$$(1.1) \quad \min \left\{ \sum_{i=1}^s f_i(x_i) : \sum_{i=1}^s A_i x_i = b, x_i \in \mathbb{R}^{n_i}, i = 1, 2, \dots, s \right\},$$

where $f_i : \mathbb{R}^{n_i} \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ 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 $i = 1, 2, \dots, s$ and $b \in \mathbb{R}^m$ is a given vector. Here, variable x is decomposed into s blocks, that is $x := (x_1, x_2, \dots, x_s)$.

Many problems arising from image processing and machine learning can be cast into the form of model (1.1). For example, the total-variation based image denoising model [26, 28], sparse representation based image restoration [3, 6, 7, 20], lasso regression [34] and support vector machines [9] are special cases of problem (1.1) with $s = 2$. In addition, we also refer to [21, 23, 29, 33] for some applications of model (1.1) with $s \geq 3$.

*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 and 11471013, and by the Natural Science Foundation of Guangdong Province under grants 2014A030310332 and 2014A030310414.

[‡]Department of Applied Mathematics, College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510640, P. R. China. (nzhsysu@gmail.com). Questions, comments, or corrections to this document may be directed to that email address.

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

[§]This author is also a professor emeritus of mathematics in Syracuse University, Syracuse, NY 13244, USA.

The alternating direction method of multipliers (ADMM) [13] was originally proposed for solving problem (1.1) with $s = 2$, and was recently widely used in the area of image processing [4, 14, 30, 35]. Since ADMM requires inner iterations to solve its subproblems of ADMM, its linearized version (LADMM) was proposed and was successfully used in applications [12]. As $s \geq 3$, one can directly extend the original ADMM (LADMM) to problem (1.1). Without an additional assumption, however, it was recently shown in [8] 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 [10, 17, 18, 22] 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 [11] 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 [15, 16] to make ADMM and LADMM converge. It was shown in [15, 16] 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 efficient algorithms.

As shown in [1, 19, 21, 24], the notion of proximity operators provides a useful tool for the algorithmic development due to its firmly nonexpansive property. ADMM was shown in [19] a special case of the proximity algorithms. Although the one-step fixed-point proximity algorithms proposed in [19] can be applied to model (1.1) directly, they do not utilize the separable property of the objective function, that is, the variable x_1, x_2, \dots, 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 (1.1), 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 [19], 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 [19] to develop convergent two-step fixed-point proximity algorithms.

This paper has the following contributions. First, we present a characterization of the solutions of problem (1.1) by fixed-points of a proximity related operator and develop a two-step fixed-point iterative scheme based on the fixed-point equation. Second, we prove convergence of the proposed iterative scheme by the notions of weakly firmly nonexpansive and Condition-M proposed in [19]. We prove that as long as the matrices involved in the scheme satisfy Condition-M, which can be easily verified, the iterative scheme converges and the sequence $\{(x_1^k, \dots, x_s^k) : k \in \mathbb{N}\}$ generated by the proposed algorithm converges to a solution of problem (1.1). Third, we analyze the convergence rate of the proposed iterative scheme. We prove that the scheme has $O(\frac{1}{k})$ ergodic convergence rate. In addition, the average of the sequence generated by the proposed scheme has $O(\frac{1}{k})$ convergence rate in the sense of the primal-dual gap. Fourth, several specific convergent algorithms are designed from the iterative scheme, including the two-step implicit and explicit fixed-point proximity algorithms as well as their variants. Furthermore, we apply the 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 necessarily convergent.

We organize this paper in eight sections. In Section 2, we characterize the solutions of problem (1.1) by fixed-points of a proximity related operator. Based on this characterization, we develop in Section 3 a two-step iterative scheme and prove its convergence in Section 4. In Section 5, we analyze the convergence

rate of the proposed iterative scheme. We design several specific algorithms from the iterative scheme in Section 6 and apply in Section 7 one of them to the sparse MRI reconstruction problem. We conclude this paper in Section 8.

2. A Characterization of Solutions of the Minimization Problem. In this section we present a characterization of solutions of model (1.1) 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 $\mathbb{N}_d := \{1, 2, \dots, d\}$ and the standard ℓ_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 ℓ_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 $\varphi : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$. For a function $\varphi \in \Gamma_0(\mathbb{R}^d)$, the proximity operator of φ 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

$$(2.1) \quad \text{prox}_{\varphi, H}(x) := \operatorname{argmin} \left\{ \frac{1}{2} \|u - x\|_H^2 + \varphi(u) : u \in \mathbb{R}^d \right\}.$$

In particular, we use prox_φ for $\text{prox}_{\varphi, I}$.

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

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

We remark that if a function $\varphi \in \Gamma_0(\mathbb{R}^d)$ is Fréchet differentiable at a point $x \in \mathbb{R}^d$ then $\partial\varphi(x) = \{\nabla\varphi(x)\}$, where $\nabla\varphi(x)$ is the Fréchet gradient of φ . It is shown that for any $H \in \mathbb{S}_+^d$, $x \in \text{dom}(\varphi)$ and $y \in \mathbb{R}^d$,

$$(2.2) \quad Hy \in \partial\varphi(x) \quad \text{if and only if} \quad x = \text{prox}_{\varphi, H}(x + y).$$

For a discussion of this relation, see, e.g., [2, Proposition 16.34] or [24].

The proximity operator plays a crucial role in convex analysis and applications (see, e.g., [25, 27]). Recall that operator J is called firmly nonexpansive (resp., nonexpansive) with respect to a given matrix $H \in \mathbb{S}_+^d$ if for all $x, y \in \mathbb{R}^d$

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

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 φ in $\Gamma_0(\mathbb{R}^d)$ is that for $x \in \text{dom}(\varphi)$ and $y \in \text{dom}(\varphi^*)$

$$(2.3) \quad y \in \partial\varphi(x) \quad \text{if and only if} \quad x \in \partial\varphi^*(y).$$

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 the smallest cone in \mathbb{R}^d containing S by $\text{cone}(S)$. 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 $n := \sum_{i=1}^s n_i$ and $A := [A_1 \ A_2 \ \dots \ A_s]$. Then, problem (1.1) can be rewritten as

$$(2.4) \quad \min \left\{ \sum_{i=1}^s f_i(x_i) + \iota_C(Ax) : x_i \in \mathbb{R}^{n_i}, i \in \mathbb{N}_s \right\},$$

where

$$(2.5) \quad C := \{b\}.$$

Now, we are ready to characterize the solutions of model (1.1) with the help of (2.2) and (2.3).

THEOREM 2.1. *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 $x := (x_1, x_2, \dots, x_s) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \dots \times \mathbb{R}^{n_s}$ is a solution of problem (1.1), 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*

$$(2.6) \quad x_i = \text{prox}_{\frac{\alpha_i}{\beta} f_i} \left(x_i - \frac{\alpha_i}{\beta} A_i^\top y \right), \quad i \in \mathbb{N}_s,$$

$$(2.7) \quad y = \text{prox}_{\beta \iota_C^*} \left(y + \beta \sum_{i=1}^s A_i x_i \right).$$

Conversely, if there exist $\beta > 0$, $\alpha_i > 0$ for $i \in \mathbb{N}_s$, $x := (x_1, x_2, \dots, x_s) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \dots \times \mathbb{R}^{n_s}$ and $y \in \mathbb{R}^m$ satisfying equations (2.6) and (2.7), then x is a solution of problem (1.1).

Proof. We prove this theorem by applying Fermat's rule that a vector $x := (x_1, x_2, \dots, x_s) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \dots \times \mathbb{R}^{n_s}$ is a solution of model (1.1) if and only if the zero vector is in the subdifferential of the objective function of model (1.1) evaluated at x .

Let $x := (x_1, x_2, \dots, x_s) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \dots \times \mathbb{R}^{n_s}$ be a solution of model (1.1). From Theorem 16.37 of [2], the chain rule of the subdifferential holds due to $b \in A(\text{ri}(\text{dom}(\sum_{i=1}^s f_i)))$. Then by Fermat's rule we obtain

$$(2.8) \quad 0 \in \partial f_i(x_i) + A_i^\top \partial \iota_C(Ax)$$

for $i \in \mathbb{N}_s$. Thus, there exists $y \in \mathbb{R}^m$ such that $y \in \partial \iota_C(Ax)$ and $-A_i^\top y \in \partial f_i(x_i)$ for $i \in \mathbb{N}_s$. The last inclusion implies that for any $\alpha_i > 0$, $\beta > 0$, $-\frac{\alpha_i}{\beta} A_i^\top y \in \partial(\frac{\alpha_i}{\beta} f_i)(x_i)$. Therefore, equation (2.6) follows from (2.2). By (2.3), from $y \in \partial \iota_C(Ax)$, we have that $Ax \in \partial \iota_C^*(y)$. Hence, for any $\beta > 0$, we obtain that $\beta Ax \in \partial(\beta \iota_C^*)(y)$, which by (2.2) is equivalent to equation (2.7).

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 equations (2.6) and (2.7). The relation (2.2) ensures that $y \in \partial \iota_C(Ax)$ and $-A_i^\top y \in \partial f_i(x_i)$. Clearly, these inclusions together ensure that the relation (2.8) holds. That is, the zero vector is in the subdifferential of the objective function at (x_1, \dots, x_s) . Again, by Fermat's rule, (x_1, \dots, x_s) is a solution of model (1.1). \square

Theorem 2.1 characterizes a solution of problem (1.1) in terms of the system of fixed-point equations (2.6) and (2.7). Through out this paper, for problem (1.1), 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 2.1, we know that fixed-point equations (2.6) and (2.7) 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 (1.1) by developing fixed-point iterative schemes.

3. A Two-step Iterative Scheme. We develop in this section a two-step iterative scheme for solving optimization problem (1.1) by using the system of fixed-point equations (2.6) and (2.7).

We begin with rewriting equations (2.6) and (2.7) in a compact form. To this end, we first introduce an operator by integrating together the $s + 1$ proximity operators involved in equations (2.6) and (2.7). Specifically, for given $f_i \in \Gamma_0(\mathbb{R}^{n_i})$, $\iota_C \in \Gamma_0(\mathbb{R}^m)$, $\alpha_i > 0$, $\beta > 0$, $i \in \mathbb{N}_s$, we define the operator $\mathcal{T} := T_{(\iota_C, \beta)}^{(f_1, \dots, f_s, \frac{\alpha_1}{\beta}, \dots, \frac{\alpha_s}{\beta})} : \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_s} \times \mathbb{R}^m \rightarrow \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_s} \times \mathbb{R}^m$ at a vector $v := (x_1, \dots, x_s, y) \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_s} \times \mathbb{R}^m$ as follows:

$$(3.1) \quad \mathcal{T}(v) := (\text{prox}_{\frac{\alpha_1}{\beta} f_1}(x_1), \dots, \text{prox}_{\frac{\alpha_s}{\beta} f_s}(x_s), \text{prox}_{\beta \iota_C^*}(y)).$$

Operator \mathcal{T} couples all the proximity operators $\text{prox}_{\frac{\alpha_i}{\beta} f_i}$, $i \in \mathbb{N}_s$ and $\text{prox}_{\beta \iota_C^*}$. In the following lemma, we show that the operator \mathcal{T} is the proximity operator of a new convex function

$$(3.2) \quad \Phi(v) := \sum_{i=1}^s f_i(x_i) + \iota_C^*(y)$$

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

$$(3.3) \quad R := \text{diag} \left(\frac{\beta}{\alpha_1} \mathbf{1}_{n_1}, \dots, \frac{\beta}{\alpha_s} \mathbf{1}_{n_s}, \frac{1}{\beta} \mathbf{1}_m \right),$$

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

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

Here we omit the proof since one can complete it by referring to Lemma 3.1 of [19]. By Lemma 3.1, we know that the operator \mathcal{T} is firmly non-expansive with respect to the matrix R . Let

$$(3.4) \quad P := \text{diag} \left(\frac{\beta}{\alpha_1} \mathbf{1}_{n_1}, \dots, \frac{\beta}{\alpha_s} \mathbf{1}_{n_s} \right).$$

With the help of the above notation, equations (2.6) and (2.7) can be reformulated in a compact form

$$(3.5) \quad v = (\mathcal{T} \circ E)(v),$$

where

$$(3.6) \quad E := \begin{bmatrix} I & -P^{-1}A^\top \\ \beta A & I \end{bmatrix}.$$

Theorem 2.1 together with equation (3.5) indicates that finding a solution of problem (1.1) essentially amounts to computing a fixed-point of the operator $\mathcal{T} \circ E$. As discussed at the end of Section 2, the operator $\mathcal{T} \circ E$ 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 [19], the matrix E is not nonexpansive due to the fact that $\|E\|_2 > 1$. Therefore, a simple fixed-point iteration $v^{k+1} = (\mathcal{T} \circ E)(v^k)$ 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 [19] and in [21]. Here, we split E as

$$(3.7) \quad E = (E - R^{-1}M_0) + R^{-1}M_1 + R^{-1}M_2,$$

where $M_i \in \mathbb{R}^{(n+m) \times (n+m)}$ for $i = 0, 1, 2$ and $M_0 = M_1 + M_2$. Accordingly, equation (3.5) is equivalent to

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

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

$$(3.8) \quad v^{k+1} = \mathcal{T}((E - R^{-1}M_0)v^{k+1} + R^{-1}M_1v^k + R^{-1}M_2v^{k-1}).$$

We point out here that although iterative scheme (3.8) 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 (3.8) 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 (3.8) converges.

To close this section, we remark that when $M_2 = 0$ (in this case, $M_0 = M_1$), the two-step iterative scheme (3.8) reduces to a one-step iterative scheme

$$(3.9) \quad v^{k+1} = \mathcal{T}((E - R^{-1}M_0)v^{k+1} + R^{-1}M_0v^k).$$

Many efficient algorithms can be obtained from (3.9) by specifying the matrix M_0 . The reader is referred to [19] for details.

4. Convergence Analysis of the Proposed Iterative Scheme. In this section, we study the convergence of iterative scheme (3.8). By applying the notion of weakly firmly nonexpansive operators and Condition-M, which were first introduced in [19], 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 (3.8) converges to a solution of equation (3.5). Hence, the sequence $\{x^k : k \in \mathbb{N}\}$ converges to a solution of model (1.1).

We begin with rewriting iterative scheme (3.8) in an explicit way. To this end, we introduce $\mathcal{M} := \{M_0, M_1, M_2\}$. We also define $T_{\mathcal{M}} : \mathbb{R}^{n+m} \times \mathbb{R}^{n+m} \rightarrow \mathbb{R}^{n+m}$, 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

$$(4.1) \quad w = \mathcal{T}((E - R^{-1}M_0)w + R^{-1}M_1u_1 + R^{-1}M_2u_2).$$

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 (3.8) for any $(u_1, u_2) \in \mathbb{R}^{n+m} \times \mathbb{R}^{n+m}$. With the help of \mathcal{M} and $T_{\mathcal{M}}$, (3.8) can be rewritten as

$$(4.2) \quad v^{k+1} = T_{\mathcal{M}}(v^k, v^{k-1}).$$

Now, we recall the notion of weakly firmly nonexpansive operators and Condition-M, which were introduced in [19].

DEFINITION 4.1 (Weakly Firmly Nonexpansive). *We say an operator $T : \mathbb{R}^{2d} \rightarrow \mathbb{R}^d$ 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 describe the definition of Condition-M.

DEFINITION 4.2 (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 to review a property of weakly firmly nonexpansive operators established in [19].

THEOREM 4.3. *Suppose that the operator $T : \mathbb{R}^{2d} \rightarrow \mathbb{R}^d$ 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 .*

By the above theorem, in order to ensure convergence of iterative scheme (4.2), it suffices to prove $T_{\mathcal{M}}$ defined by (4.1) 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

$$(4.3) \quad S_A := \begin{bmatrix} 0 & -A^\top \\ A & 0 \end{bmatrix}.$$

Then, $E = I + R^{-1}S_A$.

PROPOSITION 4.4. *Let $f_i \in \Gamma_0(\mathbb{R}^{n_i})$, $\alpha_i > 0$ for $i \in \mathbb{N}_s$ and $\beta > 0$. Let $\mathcal{M} := \{M_0, M_1, M_2\}$ be a set of $(n+m) \times (n+m)$ matrices and $T_{\mathcal{M}}$ be defined by (4.1). If $T_{\mathcal{M}}$ is well-defined, 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 $(u_i, w_i, z_i) \in \mathbb{R}^{n+m} \times \mathbb{R}^{n+m} \times \mathbb{R}^{n+m}$ satisfying $z_i = T_{\mathcal{M}}(u_i, w_i)$, for $i = 1, 2$, there holds

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

According to Lemma 3.1, \mathcal{T} is firmly nonexpansive with respect to R . Thus, we observe that

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

Since $RE = R + S_A$ and S_A is skew-symmetric, we have

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

From Definition 4.1, we get Item (i).

We next prove Item (ii). From the definition of $T_{\mathcal{M}}$, for any sequence $\{(u^k, w^k, z^k) \in \mathbb{R}^{n+m} \times \mathbb{R}^{n+m} \times \mathbb{R}^{n+m} : k \in \mathbb{N}\}$ satisfying $z^k = T_{\mathcal{M}}(u^k, w^k)$ and converging to (u, w, z) , we have that $z^k = \mathcal{T}((E - M_0)z^k + R^{-1}M_1u^k + R^{-1}M_2w^k)$. This with the continuity of \mathcal{T} implies that $z = \mathcal{T}((E - M_0)z + R^{-1}M_1u + R^{-1}M_2w)$. Thus, $z = T_{\mathcal{M}}(u, w)$, proving Item (ii). \square

We are now ready to prove convergence of the sequence generated from iterative scheme (3.8).

THEOREM 4.5. *Let $f_i \in \Gamma_0(\mathbb{R}^{n_i})$, $\alpha_i > 0$ for $i \in \mathbb{N}_s$ and $\beta > 0$. Let \mathcal{T} and E be defined as (3.1) and (3.6) respectively, $\mathcal{M} := \{M_0, M_1, M_2\}$ be a set of $(n+m) \times (n+m)$ matrices and $T_{\mathcal{M}}$ be defined by (4.1). Let $\{v^k : k \in \mathbb{N}\}$ be generated by (3.8) for given points v^0, v_1 . Suppose that $T_{\mathcal{M}}$ is well-defined. If \mathcal{M} satisfies Condition-M, then the sequence $\{v^k : k \in \mathbb{N}\}$ converges to a fixed-point of $\mathcal{T} \circ E$, and $\{x^k : k \in \mathbb{N}\}$ converges to a solution of problem (1.1).*

Proof. By the definition of $T_{\mathcal{M}}$, operators $T_{\mathcal{M}}$ and $\mathcal{T} \circ E$ share the same set of fixed-points. By Proposition 4.4, the operator $T_{\mathcal{M}}$ is weakly firmly non-expansive with respect to \mathcal{M} and continuous. Therefore, Theorem 4.3 ensures that the sequence $\{v^k : k \in \mathbb{N}\}$ converges to a fixed-point of $T_{\mathcal{M}}$. By Proposition 2.1, the sequence $\{x^k : k \in \mathbb{N}\}$ converges to a solution of problem (1.1). \square

Theorem 4.5 shows that convergence of iterative scheme (3.8) relies completely on whether the matrices set \mathcal{M} used in scheme (3.8) satisfies Condition-M. We will develop in Section 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 (3.8). We show that the proposed algorithm has $O(\frac{1}{k})$ convergence rate in the ergodic sense and the sense of the partial primal-dual gap.

5.1. Ergodic $O(\frac{1}{k})$ Rate. We first study the convergence rate of the proposed algorithm (3.8) in the ergodic sense. We prove in this subsection that the proposed iterative scheme (3.8) has $O(\frac{1}{k})$ convergence in the ergodic sense. To this end, we first review a lemma presented in [31].

LEMMA 5.1. *If a sequence $\{a^k : k \in \mathbb{N}\}$ satisfies: $a^k \geq 0$ and $\sum_{i=1}^{+\infty} a^i < +\infty$, then*

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

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

THEOREM 5.2. *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 \mathcal{T} and E be defined as (3.1) and (3.6) respectively. Let the sequence $\{v^k : k \in \mathbb{N}\}$ be generated from (3.8) for any given $v^0, v^1 \in \mathbb{R}^{n+m}$. Suppose that $T_{\mathcal{M}}$ is well-defined. If \mathcal{M} satisfies Condition-M, then*

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

$$(5.1) \quad \frac{1}{k} \sum_{i=1}^k \|v^{i+1} - v^i\|_2^2 = O(\frac{1}{k}),$$

- (ii) *the running minimal of progress, $\min_{i \leq k} \{\|v^{i+1} - v^i\|_2^2\}$, has $o(\frac{1}{k})$ convergence.*

Proof. By Lemma 5.1, we only need to prove

$$(5.2) \quad \sum_{i=1}^{+\infty} \|v^{i+1} - v^i\|_2^2 < +\infty.$$

By the definition of $T_{\mathcal{M}}$, the sequence $\{v^k : k \in \mathbb{N}\}$ generated from (3.8) can also be generated by (4.2) for the same given v^0, v^1 . Since $T_{\mathcal{M}}$ is weakly firmly nonexpansive with respect to \mathcal{M} and \mathcal{M} satisfies Condition-M, by Lemma 4.4 of [19], we have for any $k \geq 3$ that

$$(5.3) \quad \|e^k\|_H^2 \leq 2\|e^2\|_H^2 + 2\|\widetilde{M}\|_2^2 \|r^2\|_H^2 - 2\langle e^2, M_2 r^2 \rangle - (\frac{1}{2} - 2\|\widetilde{M}\|_2^2) \sum_{i=2}^{k-1} \|r^{i+1}\|_H^2,$$

where $e^i := v^i - v$ for v a fixed-point of $T_{\mathcal{M}}$, $r^i := v^i - v^{i-1}$ and $\widetilde{M} := H^{-1/2} M_2 H^{-1/2}$. By (iii) of Condition-M, we have $\frac{1}{2} - 2\|\widetilde{M}\|_2^2 > 0$. Then (5.2) is obtained immediately from (5.3) and the fact that $H \in \mathbb{S}_+^{n+m}$. \square

5.2. Partial Primal-dual Gap $O(\frac{1}{k})$ Convergence Rate . In this subsection, we study the convergence rate of the proposed iterative algorithm (3.8) in the sense of the partial primal-dual gap. We prove that iterative scheme (3.8) has $O(\frac{1}{k})$ 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 (1.1). To this end, we review the primal-dual formulation of problem (1.1), that is

$$(5.4) \quad \min \{ \max \{ \sum_{i=1}^s f_i(x_i) - \iota_C^*(y) + \langle Ax, y \rangle : y \in \mathbb{R}^m \} : x_i \in \mathbb{R}^{n_i}, i \in \mathbb{N}_s \}.$$

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 (1.1) at point $v := (x_1, \dots, x_s, y) \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_s} \times \mathbb{R}^m$ is defined as

$$(5.5) \quad \mathcal{G}_{B_1 \times B_2}(v) := \max \{ \sum_{i=1}^s f_i(x_i) - \iota_C^*(y') + \langle Ax, y' \rangle : y' \in B_2 \} - \min \{ \sum_{i=1}^s f_i(x'_i) - \iota_C^*(y) + \langle Ax', y \rangle : x' \in B_1 \}.$$

We refer to [5] for more details on the partial primal-dual gap.

In order to analyze the convergence rate of iterative scheme (3.8), we define $G : \mathbb{R}^{n+m} \times \mathbb{R}^{n+m} \rightarrow \overline{\mathbb{R}}$ by

$$(5.6) \quad G(v, v') := \Phi(v) - \Phi(v') + \langle v', S_A v \rangle,$$

where Φ and S_A are defined as (3.2) and (4.3) respectively. For $(v, v') \in \mathbb{R}^{n+m} \times \mathbb{R}^{n+m}$, where $v' := (x'_1, \dots, x'_s, y') \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_s} \times \mathbb{R}^m$ and $v := (x_1, \dots, x_s, y) \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_s} \times \mathbb{R}^m$, one can check that (5.6) is equivalent to

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

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 $G(v^{k+1}, v)$ for any $v \in \mathbb{R}^{n+m}$.

LEMMA 5.3. *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 \mathcal{T} and E be defined as (3.1) and (3.6) respectively, $\mathcal{M} := \{M_0, M_1, M_2\}$ 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 (3.8). For all $v \in \mathbb{R}^{n+m}$ there holds*

$$(5.7) \quad G(v^{k+1}, v) \leq \langle M_1 v^k + M_2 v^{k-1} - M_0 v^{k+1}, v^{k+1} - v \rangle.$$

Proof. From iterative scheme (3.8), Lemma 3.1 and (2.2), 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 Φ , we have for any $v \in \mathbb{R}^{n+m}$ that

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

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

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

Then, we obtain (5.7) immediately by the definition of G . \square

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

LEMMA 5.4. *Let $\{v^k : k \in \mathbb{N}\}$ be generated from iterative scheme (3.8). Under the same assumptions of Lemma 5.3, if \mathcal{M} satisfies Condition-M, then for all $v \in \mathbb{R}^{n+m}$ there holds*

$$(5.8) \quad G(\bar{v}_K, v) \leq \frac{\frac{3}{4}\|v^1 - v\|_H^2 + \frac{1}{2}\|v^1 - v^0\|_H^2}{K},$$

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 5.3 and Item (i) of Condition-M, we have

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

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

$$(5.9) \quad G(v^{k+1}, v) \leq D_1 + D_2,$$

where $D_1 := -\|e^{k+1}\|_H^2 + \langle e^{k+1}, He^k \rangle$ 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 $\langle a, Hb \rangle = \frac{1}{2}(\|a\|_H^2 + \|b\|_H^2 - \|a - b\|_H^2)$ for $a, b \in \mathbb{R}^{n+m}$, we obtain that

$$(5.10) \quad D_1 = \frac{1}{2}(-\|e^{k+1}\|_H^2 + \|e^k\|_H^2 - \|r^{k+1}\|_H^2).$$

We also have

$$(5.11) \quad \begin{aligned} D_2 &= \langle e^{k+1}, M_2(r^{k+1} - r^k) \rangle \\ &= \langle e^{k+1}, M_2 r^{k+1} \rangle - \langle r^{k+1}, M_2 r^k \rangle - \langle e^k, M_2 r^k \rangle, \end{aligned}$$

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 $\widetilde{M} := (H^\dagger)^{1/2} M_2 (H^\dagger)^{1/2}$. Then it follows that for any $a > 0$,

$$(5.12) \quad |\langle r^{k+1}, M_2 r^k \rangle| \leq \frac{a}{2} \|r^{k+1}\|_H^2 + \frac{\|\widetilde{M}\|_2^2}{2a} \|r^k\|_H^2.$$

Thus, by (5.9), (5.10), (5.11) and (5.12), we have

$$(5.13) \quad \begin{aligned} G(v^{k+1}, v) &\leq \frac{1}{2}(-\|e^{k+1}\|_H^2 + \|e^k\|_H^2) - \frac{1}{2}\left(1 - a - \frac{\|\widetilde{M}\|_2^2}{a}\right) \|r^{k+1}\|_H^2 \\ &\quad + \frac{\|\widetilde{M}\|_2^2}{2a} (-\|r^{k+1}\|_H^2 + \|r^k\|_H^2) + \langle e^{k+1}, M_2 r^{k+1} \rangle - \langle e^k, M_2 r^k \rangle. \end{aligned}$$

Summing inequality (5.13) from $k = 1$ to $k = K$, we have

$$(5.14) \quad \begin{aligned} \sum_{k=1}^K G(v^{k+1}, v) &\leq \frac{1}{2}(-\|e^{K+1}\|_H^2 + \|e^1\|_H^2) - \frac{1}{2}\left(1 - a - \frac{\|\widetilde{M}\|_2^2}{a}\right) \sum_{k=2}^{K+1} \|r^k\|_H^2 \\ &\quad + \frac{\|\widetilde{M}\|_2^2}{2a} (-\|r^{K+1}\|_H^2 + \|r^1\|_H^2) + \langle e^{K+1}, M_2 r^{K+1} \rangle - \langle e^1, M_2 r^1 \rangle. \end{aligned}$$

By applying

$$|\langle e^k, M_2 r^k \rangle| \leq \frac{a}{2} \|e^k\|_H^2 + \frac{\|\widetilde{M}\|_2^2}{2a} \|r^k\|_H^2$$

for $k = K + 1$ and $k = 1$ to the last two terms of (5.14), we obtain that

$$(5.15) \quad \begin{aligned} \sum_{k=1}^K G(v^{k+1}, v) &\leq \frac{1}{2}(-(1-a)\|e^{K+1}\|_H^2 + (1+a)\|e^1\|_H^2) \\ &\quad - \frac{1}{2}\left(1 - a - \frac{\|\widetilde{M}\|_2^2}{a}\right) \sum_{k=2}^{K+1} \|r^k\|_H^2 + \frac{\|\widetilde{M}\|_2^2}{a} \|r^1\|_H^2. \end{aligned}$$

Setting $a = \frac{1}{2}$, it follows that $1 - a - \frac{\|\widetilde{M}\|_2^2}{a} > 0$ due to (iii) of Condition-M. This together with (5.15) yields

$$(5.16) \quad \frac{1}{K} \sum_{k=1}^K G(v^{k+1}, v) \leq \frac{\frac{3}{4}\|e^1\|_H^2 + \frac{1}{2}\|r^1\|_H^2}{K}.$$

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 (5.16) implies (5.8). \square

Now, we are ready to present the partial primal-gap convergence rate of the proposed algorithm (3.8) in the next theorem.

THEOREM 5.5. *Let $\{v^k : k \in \mathbb{N}\}$ be generated from iterative scheme (3.8). Under the same assumptions of Lemma 5.3, if \mathcal{M} satisfies Condition-M, then iterative scheme (3.8) has $O(\frac{1}{K})$ convergence rate in the partial primal-dual gap sense, that is*

$$\mathcal{G}_{B_1 \times B_2}(\bar{v}_K) = O\left(\frac{1}{K}\right),$$

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 5.4 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 (3.8) 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 $M_0 \in \mathbb{S}_+^{n+m}$.

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 (3.6), $M_0 = M_1$ can be chosen as Z_1 or Z_2 with

$$Z_1 := \begin{bmatrix} P & -A^\top \\ -A & \frac{1}{\beta}I \end{bmatrix}, Z_2 = \begin{bmatrix} P & A^\top \\ A & \frac{1}{\beta}I \end{bmatrix},$$

where P is defined by (3.4). By simple calculations, one can obtain that $\iota_C^*(\cdot) = \langle \cdot, b \rangle$ and thus $\text{prox}_{\beta \iota_C^*}(y) = y - \beta b$ for $y \in \mathbb{R}^m$. Then, iterative scheme (3.8) with respect to Z_1 and Z_2 become, respectively,

$$(6.1) \quad \begin{cases} x_i^{k+1} = \text{prox}_{\frac{\alpha_i}{\beta} f_i}(x_i^k - \frac{\alpha_i}{\beta} A_i^\top y^k), & i \in \mathbb{N}_s, \\ y^{k+1} = y^k + \beta(\sum_{i=1}^s A_i(2x_i^{k+1} - x_i^k) - b), \end{cases}$$

and

$$(6.2) \quad \begin{cases} y^{k+1} = y^k + \beta(\sum_{i=1}^s A_i x_i^k - b), \\ x_i^{k+1} = \text{prox}_{\frac{\alpha_i}{\beta} f_i}(x_i^k - \frac{\alpha_i}{\beta} A_i^\top (2y^{k+1} - y^k)), & i \in \mathbb{N}_s. \end{cases}$$

We note that, algorithms (6.1) and (6.2) are actually special cases of the one-step first-order primal-dual algorithm [5, 12, 19], which solves the following optimization problem

$$(6.3) \quad \min\{f(x) + g(Ax) : x \in \mathbb{R}^n\}$$

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 $x := (x_1, \dots, x_s)$, $g := \iota_C$ and $f : \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_s} \rightarrow \bar{\mathbb{R}}$ defined at x as $f(x) := \sum_{i=1}^s f_i(x_i)$, then problem (6.3) is exactly the optimization problem (1.1). Clearly, algorithms (6.1) and (6.2) are special cases of the one-step first-order primal-dual algorithm [5, 19] by the fact that $\text{prox}_{f, P}(x) = (\text{prox}_{\frac{\alpha_1}{\beta} f_1}(x_1), \dots, \text{prox}_{\frac{\alpha_s}{\beta} f_s}(x_s))$. The corresponding convergence results are presented in the following theorem.

THEOREM 6.1. *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 $\{(x_1^k, \dots, x_s^k, y^k) : k \in \mathbb{N}\}$ generated from (6.1) or (6.2) for any $(x_1^0, \dots, x_s^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m$. If $\|AQ\|_2 < 1$, where $Q := \text{diag}(\sqrt{\alpha_1} \mathbf{1}_{n_1}, \dots, \sqrt{\alpha_s} \mathbf{1}_{n_s})$, then $\{(x_1^k, \dots, x_s^k, y^k) : k \in \mathbb{N}\}$ converges and the sequence $\{x^k : k \in \mathbb{N}\}$ converges to a solution of problem (1.1).*

We omit the proof since it can be obtained immediately by applying Lemma 6.2 in [19] and Theorem 4.5.

To close this subsection, we remark that both algorithms (6.1) and (6.2) do not take advantage of the separability of function f and vector x . More precisely, the information of x_j^{k+1} for $j = 1, \dots, i-1$ 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, \dots, 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 (3.8). We begin with constructing the set of matrices $\mathcal{M} := \{M_0, M_1, M_2\}$ by setting

$$(6.4) \quad M_0 := \begin{bmatrix} \frac{\beta}{\alpha_1}I & -\beta A_1^\top A_2 & -\beta A_1^\top A_3 & \dots & -\beta A_1^\top A_s & 0 \\ 0 & \frac{\beta}{\alpha_2}I & -\beta A_2^\top A_3 & \dots & -\beta A_2^\top A_s & 0 \\ 0 & 0 & \frac{\beta}{\alpha_3}I & \dots & -\beta A_3^\top A_s & 0 \\ \dots & \dots & \dots & \ddots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{\beta}{\alpha_s}I & 0 \\ 0 & 0 & 0 & \dots & 0 & \frac{1}{\beta}I \end{bmatrix},$$

$$(6.5) \quad M_1 := \begin{bmatrix} \frac{\beta}{\alpha_1}I & -2\beta A_1^\top A_2 & -2\beta A_1^\top A_3 & \dots & -2\beta A_1^\top A_s & 0 \\ 0 & \frac{\beta}{\alpha_2}I & -2\beta A_2^\top A_3 & \dots & -2\beta A_2^\top A_s & 0 \\ 0 & 0 & \frac{\beta}{\alpha_3}I & \dots & -2\beta A_3^\top A_s & 0 \\ \dots & \dots & \dots & \ddots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{\beta}{\alpha_s}I & 0 \\ 0 & 0 & 0 & \dots & 0 & \frac{1}{\beta}I \end{bmatrix},$$

$$(6.6) \quad M_2 := \begin{bmatrix} 0 & \beta A_1^\top A_2 & \beta A_1^\top A_3 & \dots & \beta A_1^\top A_s & 0 \\ 0 & 0 & \beta A_2^\top A_3 & \dots & \beta A_2^\top A_s & 0 \\ 0 & 0 & 0 & \dots & \beta A_3^\top A_s & 0 \\ \dots & \dots & \dots & \ddots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

With this choice of matrices M_0, M_1, M_2 , noting that $\text{prox}_{\beta \iota_C^*}(y) = y - \beta b$, iterative scheme (3.8) 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 (3.8), we obtain that

$$(6.7) \quad \begin{cases} x_j^{k+1} = \text{prox}_{\frac{\alpha_j}{\beta} f_j} \left(x_j^k - \alpha_j A_j^\top \left(\sum_{i=1}^j A_i x_i^{k+1} + \sum_{i=j+1}^s A_i (2x_i^k - x_i^{k-1}) - b \right) - \frac{\alpha_j}{\beta} A_j^\top y^k \right), j \in \mathbb{N}_s, \\ y^{k+1} = y^k + \beta \left(\sum_{i=1}^s A_i x_i^{k+1} - b \right). \end{cases}$$

We point out the connections of the proposed algorithm (6.7) with the proximal ADMM (PADMM). To this end, we introduce the augmented Lagrangian function for (1.1)

$$(6.8) \quad \mathcal{L}(x_1, \dots, x_s, y) := \sum_{i=1}^s f_i(x_i) + \frac{\beta}{2} \left\| \sum_{i=1}^s A_i x_i - b \right\|_2^2 + \langle y, \sum_{i=1}^s A_i x_i - b \rangle.$$

The PADMM for (1.1) reads as

$$(6.9) \quad \begin{cases} x_j^{k+1} = \arg \min \{ \mathcal{L}(x_1^{k+1}, \dots, x_{j-1}^{k+1}, x_j, x_{j+1}^k, \dots, x_s^k, y^k) + \frac{\beta}{2\alpha_j} \|x_j - x_j^k\|_2^2 : x_j \in \mathbb{R}^{n_j} \}, j \in \mathbb{N}_s, \\ y^{k+1} = y^k + \beta(\sum_{i=1}^s A_i x_i^{k+1} - b). \end{cases}$$

On the other hand, by the definition of proximity operator (2.1), the proposed algorithm (6.7) can be equivalently rewritten as

$$(6.10) \quad \begin{cases} x_j^{k+1} = \arg \min \{ \mathcal{L}(x_1^{k+1}, \dots, x_{j-1}^{k+1}, x_j, \tilde{x}_{j+1}^k, \dots, \tilde{x}_s^k, y^k) + \frac{\beta}{2\alpha_j} \|x_j - x_j^k\|_2^2 : x_j \in \mathbb{R}^{n_j} \}, j \in \mathbb{N}_s, \\ y^{k+1} = y^k + \beta(\sum_{i=1}^s A_i x_i^{k+1} - b), \\ \tilde{x}_j^{k+1} = 2x_j^{k+1} - x_j^k, j \in \mathbb{N}_s. \end{cases}$$

We can observe that our proposed algorithm (6.7) reduces to the PADMM if we set $\tilde{x}_j^{k+1} = x_j^{k+1}$ for $j \in \mathbb{N}_s$ in (6.10). As shown in [8], convergence of ADMM directly applied to problem (1.1) 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 (6.10) is ensured to converge without extra assumptions on f_i for $i \in \mathbb{N}_s$. We next establish the convergence result of algorithm (6.7).

PROPOSITION 6.2. *Let M_0, M_1, M_2 be defined as (6.4), (6.5) and (6.6). Let $\widetilde{M}_2 := \frac{1}{\beta} M_2$. If*

$$(6.11) \quad 0 < \alpha_i < \frac{1}{2\|\widetilde{M}_2\|_2}, \text{ for } i \in \mathbb{N}_s \text{ and } \beta > 0,$$

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 $H = \text{diag}(\frac{\beta}{\alpha_1} \mathbf{1}_{n_1}, \dots, \frac{\beta}{\alpha_s} \mathbf{1}_{n_s}, 1/\beta \mathbf{1}_m)$ 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

$$\|H^{-1/2} M_2 H^{-1/2}\|_2 = \|\widetilde{H}^{-1/2} M_2 \widetilde{H}^{-1/2}\|_2,$$

where $\widetilde{H} := \text{diag}(\frac{\beta}{\alpha_1} \mathbf{1}_{n_1}, \dots, \frac{\beta}{\alpha_s} \mathbf{1}_{n_s}, \mathbf{0}_m)$. By using hypothesis (6.11), we find that

$$\|\widetilde{H}^{-1/2} M_2 \widetilde{H}^{-1/2}\|_2 \leq \max\{\alpha_i : i \in \mathbb{N}_s\} \|\widetilde{M}_2\|_2 < \frac{1}{2},$$

which leads to Item (iii) of Condition-M. \square

The convergence results of algorithm (6.7) is presented below.

THEOREM 6.3. *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 the sequence $\{(x^k, y^k) : k \in \mathbb{N}\}$ be generated from the algorithm (6.10) for any $(x^0, y^0), (x^1, y^1) \in \mathbb{R}^n \times \mathbb{R}^m$. Let M_2 be defined as (6.6) and $\widetilde{M}_2 = \frac{1}{\beta} M_2$. If the condition (6.11) is satisfied, then the sequence $\{x^k : k \in \mathbb{N}\}$ converges to a solution of problem (1.1).*

Proof. By Proposition 6.2 and Theorem 4.5, it suffices to prove $T_{\mathcal{M}}$ is well-defined when M_0, M_1, M_2 are defined by (6.4), (6.5) and (6.6). In this case, if $w = T_{\mathcal{M}}(u, v)$, where $w := (w_1, \dots, w_s, w_y), u := (u_1, \dots, u_s, u_y), v := (v_1, \dots, v_s, v_y) \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_s} \times \mathbb{R}^m$, 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

$$w_i = \arg \min \{ f_i(x_i) + \frac{\beta}{2} \left\| \sum_{j=1}^{i-1} A_j w_j + A_i x_i + \sum_{j=i+1}^s A_j (2u_j - v_j) - b \right\|_2^2 + \langle u_y, A_i x_i \rangle + \frac{\beta}{2\alpha_i} \|x_i - u_i\|_2^2 : x_i \in \mathbb{R}^{n_i} \}.$$

Since the objective function of the above optimization problem is strongly convex, $T_{\mathcal{M}}$ is well-defined. \square

To end this subsection, we point out that compared with algorithms (6.1) and (6.2), algorithm (6.7) takes advantage of the separable structure of variable x and applies the block-wise Gauss-Seidel technique to blocks x_1, \dots, x_s, y . We also note that solving the subproblems involved in (6.7) may require inner iterations. In practice, it will affect the computational efficiency of the algorithm (6.7). 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, \dots, x_s, y .

We begin with specifying the set of matrices \mathcal{M} . We set

$$(6.12) \quad M_0 := \begin{bmatrix} \frac{\beta}{\alpha_1}I - \beta A_1^\top A_1 & -\beta A_1^\top A_2 & -\beta A_1^\top A_3 & \dots & -\beta A_1^\top A_s & 0 \\ 0 & \frac{\beta}{\alpha_2}I - \beta A_2^\top A_2 & -\beta A_2^\top A_3 & \dots & -\beta A_2^\top A_s & 0 \\ 0 & 0 & \frac{\beta}{\alpha_3}I - \beta A_3^\top A_3 & \dots & -\beta A_3^\top A_s & 0 \\ \dots & \dots & \dots & \ddots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{\beta}{\alpha_s}I - \beta A_s^\top A_s & 0 \\ 0 & 0 & 0 & \dots & 0 & \frac{1}{\beta}I \end{bmatrix},$$

$$(6.13) \quad M_1 := \begin{bmatrix} \frac{\beta}{\alpha_1}I - \beta A_1^\top A_1 & -2\beta A_1^\top A_2 & -2\beta A_1^\top A_3 & \dots & -2\beta A_1^\top A_s & 0 \\ 0 & \frac{\beta}{\alpha_2}I - \beta A_2^\top A_2 & -2\beta A_2^\top A_3 & \dots & -2\beta A_2^\top A_s & 0 \\ 0 & 0 & \frac{\beta}{\alpha_3}I - \beta A_3^\top A_3 & \dots & -2\beta A_3^\top A_s & 0 \\ \dots & \dots & \dots & \ddots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{\beta}{\alpha_s}I - \beta A_s^\top A_s & 0 \\ 0 & 0 & 0 & \dots & 0 & \frac{1}{\beta}I \end{bmatrix},$$

and let M_2 be defined as in (6.6). We can obtain an implicit algorithm by directly substituting (6.12), (6.13) and (6.6) into the iterative scheme (3.8). As the same as the algorithm (6.7), it implies

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

As in subsection 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

$$(6.14) \quad (2\text{SFPPA}) \quad \begin{cases} x_j^{k+1} = \text{prox}_{\frac{\alpha_j}{\beta} f_j} \left(x_j^k - \alpha_j A_j^\top \left(\sum_{i=1}^{j-1} A_i x_i^{k+1} + A_j x_j^k + \sum_{i=j+1}^s A_i (2x_i^k - x_i^{k-1}) - b \right) - \frac{\alpha_j}{\beta} A_j^\top y^k \right), & j \in \mathbb{N}_s, \\ y^{k+1} = y^k + \beta \left(\sum_{i=1}^s A_i x_i^{k+1} - b \right). \end{cases}$$

We point out here the relationship between the proposed algorithm (6.14) and the LADMM. To this end, we first review the exact extension of LADMM to problem (1.1). For $j \in \mathbb{N}_s$, let $J_j : \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_s} \times \mathbb{R}^m \rightarrow \mathbb{R}^{n_j}$ defined, at $(x_1, x_2, \dots, x_s, y)$, as $J_j(x_1, \dots, x_s, y) := \beta(A_j^\top \sum_{i=1}^s A_i x_i - b) + A_j^\top y$. The direct extension of LADMM to the multi-block problem is as follows

$$(6.15) \quad \begin{cases} x_j^{k+1} = \arg \min \{ f_j(x_j) + \langle J_j(x_1^{k+1}, \dots, x_{j-1}^{k+1}, x_j^k, \dots, x_s^k, y^k), x_j - x_j^k \rangle \\ \quad + \frac{\beta}{2\alpha_j} \|x_j - x_j^k\|_2^2 : x_j \in \mathbb{R}^{n_j} \}, & j \in \mathbb{N}_s, \\ y^{k+1} = y^k + \beta \left(\sum_{i=1}^s A_i x_i^{k+1} - b \right). \end{cases}$$

Using the above notations and the definition of proximity operators (2.1), the algorithm (6.14) can be rewritten in its equivalent form

$$(6.16) \quad \begin{cases} x_j^{k+1} = \arg \min \{ f_j(x_j) + \langle J_j(x_1^{k+1}, \dots, x_{j-1}^{k+1}, x_j^k, \tilde{x}_{j+1}^k, \dots, \tilde{x}_s^k, y^k), x_j - x_j^k \rangle \\ \quad + \frac{\beta}{2\alpha_j} \|x_j - x_j^k\|_2^2 : x_j \in \mathbb{R}^{n_j} \}, j \in \mathbb{N}_s, \\ y^{k+1} = y^k + \beta(\sum_{i=1}^s A_i x_i^{k+1} - b), \\ \tilde{x}^{k+1} = 2x^{k+1} - x^k. \end{cases}$$

Obviously, our proposed algorithm (6.14) reduces to the LADMM if we set $\tilde{x}^{k+1} = x^{k+1}$ in (6.16). As mentioned in [16], the direct extension of LADMM to the multi-block problem (1.1) is not necessarily convergent. Nevertheless, the convergence of the proposed algorithm (6.14) is guaranteed. Next we present the convergence results of the algorithm (6.14).

PROPOSITION 6.4. *Let M_0, M_1 and M_2 be defined as in (6.12), (6.13) and (6.6). Let $\widetilde{M}_2 := \frac{1}{\beta} M_2$. If*

$$(6.17) \quad 0 < \alpha_i < \frac{1}{\|A_i\|_2^2 + 2\|\widetilde{M}_2\|_2}, \text{ for } i \in \mathbb{N}_s \text{ and } \beta > 0,$$

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 $H = \text{diag}(\frac{\beta}{\alpha_1} I - \beta A_1^\top A_1, \dots, \frac{\beta}{\alpha_s} I - \beta A_s^\top A_s, 1/\beta I)$ is symmetric. In light of (6.17), we have that $H \in \mathbb{S}_+^{n+m}$. Hence, Item (ii) of Condition-M holds. We next show Item (iii) of Condition-M. Similar to the proof of Proposition 6.2,

$$\|H^{-1/2} M_2 H^{-1/2}\|_2 = \|\widetilde{H}^{-1/2} M_2 \widetilde{H}^{-1/2}\|_2,$$

where $\widetilde{H} := \text{diag}(\frac{\beta}{\alpha_1} I - \beta A_1^\top A_1, \dots, \frac{\beta}{\alpha_s} I - \beta A_s^\top A_s, 0)$. Hypothesis (6.17) leads to

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

This completes the proof. \square

The following theorem regards the convergence of algorithm (6.14).

THEOREM 6.5. *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 (6.14) for any $(x^0, y^0), (x^1, y^1) \in \mathbb{R}^n \times \mathbb{R}^m$. Let M_2 be defined as (6.6) and $\widetilde{M}_2 = \frac{1}{\beta} M_2$. If condition (6.17) is satisfied, then the sequence $\{x^k : k \in \mathbb{N}\}$ converges to a solution of problem (1.1).*

Proof. By Theorem 4.5 and Proposition 6.4, we only need to prove $T_{\mathcal{M}}$ is well-defined. In this case, from algorithm (6.14), it is obvious that $T_{\mathcal{M}}$ can be computed explicitly. Therefore $T_{\mathcal{M}}$ is well-defined. \square

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 (6.14) may be more efficient than the two-step algorithm (6.7). This is because the two-step algorithm (6.7) may require inner iterations to solve the subproblems involved, while each step of algorithm (6.14) can be implemented efficiently by making use of the closed form.

6.4. Variants of algorithms (6.7) and (6.14). There is a wide variety of the choices of $\{M_0, M_1, M_2\}$ satisfying condition-M, including those of algorithms (6.7) and (6.14). 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 (3.8) reduces to a class of new algorithms, which can be viewed as variants of algorithms (6.7) and (6.14).

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 (6.5) can be

chosen as $((\theta + 1)\frac{\beta}{\alpha_1}\mathbf{1}_{n_1}, \dots, (\theta + 1)\frac{\beta}{\alpha_s}\mathbf{1}_{n_s}, \frac{1}{\beta}\mathbf{1}_m)$ with $\theta \in [0, 1)$ and correspondingly, the diagonal entries of M_2 in (6.6) should be $(-\theta\frac{\beta}{\alpha_1}\mathbf{1}_{n_1}, \dots, -\theta\frac{\beta}{\alpha_s}\mathbf{1}_{n_s}, \mathbf{0}_m)$. With such a choice of $\{M_0, M_1, M_2\}$, iterative scheme (3.8) reduces to a variant of algorithm (6.7)

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

As a second example, the diagonal blocks of M_1 in (6.13) can be chosen as $(\frac{\beta}{\alpha_1}I - 2\beta A_1^\top A_1, \dots, \frac{\beta}{\alpha_s}I - 2\beta A_s^\top A_s, \frac{1}{\beta}I)$. Accordingly, the diagonal blocks of M_2 in (6.6) should be $(\beta A_1^\top A_1, \dots, \beta A_s^\top A_s, \mathbf{0})$ to make $M_0 = M_1 + M_2$. These matrices leads to a variant of algorithm (6.14)

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

Modifications of nondiagonal blocks: We change the (i, j) -th block of M_0 (defined by (6.4) or (6.12)) 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 (6.6). 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 (3.8)

$$(6.18) \quad \begin{cases} x_j^{k+1} = \text{prox}_{\frac{\alpha_j}{\beta}f_j}(x_j^k - \alpha_j A_j^\top (\sum_{i=1}^{j-1} A_i (x_i^{k+1} + \theta(x_i^{k+1} - x_i^k)) + A_j x_j^{k+1} \\ + \sum_{i=j+1}^s A_i ((2 + \theta)x_i^k - (\theta + 1)x_i^{k-1}) - b) - \frac{\alpha_j}{\beta} A_j^\top y^k), j \in \mathbb{N}_s, \\ y^{k+1} = y^k + \beta(\sum_{i=1}^s A_i x_i^{k+1} - b), \end{cases}$$

$$(6.19) \quad \begin{cases} x_j^{k+1} = \text{prox}_{\frac{\alpha_j}{\beta}f_j}(x_j^k - \alpha_j A_j^\top (\sum_{i=1}^{j-1} A_i (x_i^{k+1} + \theta(x_i^{k+1} - x_i^k)) + A_j x_j^k \\ + \sum_{i=j+1}^s A_i ((2 + \theta)x_i^k - \theta x_i^{k-1}) - b) - \frac{\alpha_j}{\beta} A_j^\top y^k), j \in \mathbb{N}_s, \\ y^{k+1} = y^k + \beta(\sum_{i=1}^s A_i x_i^{k+1} - b). \end{cases}$$

Hybrids of both algorithms: Both algorithms (6.7) and (6.14) share the same matrix M_2 . Matrices M_0 for algorithms (6.7) and (6.14) 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 (6.10) 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 (6.10) 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 $\frac{\beta}{\alpha_i}I - \beta A_i^\top A_i$ for $i \in S_2$. The nondiagonal blocks of M_0 are chosen to be the same as in (6.4) and (6.12). We further choose the matrix M_2 as in (6.6). Accordingly, the matrix M_1 is determined by $M_1 = M_0 - M_2$. Then we obtain the following hybrid algorithm

$$\begin{cases} x_j^{k+1} = \text{prox}_{\frac{\alpha_j}{\beta}f_j}(x_j^k - \alpha_j A_j^\top (\sum_{i=1}^j A_i x_i^{k+1} \\ + \sum_{i=j+1}^s A_i (2x_i^k - x_i^{k-1}) - b) - \frac{\alpha_j}{\beta} A_j^\top y^k), \text{ if } j \in S_1, \\ x_j^{k+1} = \text{prox}_{\frac{\alpha_j}{\beta}f_j}(x_j^k - \alpha_j A_j^\top (\sum_{i=1}^{j-1} A_i x_i^{k+1} + A_j x_j^k + \\ \sum_{i=j+1}^s A_i (2x_i^k - x_i^{k-1}) - b) - \frac{\alpha_j}{\beta} A_j^\top y^k), \text{ if } j \in S_2, \\ y^{k+1} = y^k + \beta(\sum_{i=1}^s A_i x_i^{k+1} - b). \end{cases}$$

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 (6.7) and (6.14).

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 [23]. We shall compare the performances of the proposed 2SFPPA with those 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 $F(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$ is a sparse-promoting function. It is well-known that superior image reconstruction can be obtained when $F(\cdot)$ is chosen to be the hybrid of total variation and the ℓ_1 -norm of the Haar wavelet transform. Denote the Haar wavelet transform matrix by $W \in \mathbb{R}^{q \times d}$ and define the $q \times q$ diagonal matrix $\Lambda := \text{diag}(\lambda_1, \dots, \lambda_q)$ with $\lambda_i \geq 0, i \in \mathbb{N}_q$. We turn to considering the following specific sparse MRI problem

$$(7.1) \quad \min\{\mu\|u\|_{\text{TV}} + \|\Lambda W u\|_1 : u \in \mathbb{R}^d, Ku = b\},$$

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 (7.1). First, we rewrite $\|\cdot\|_{\text{TV}}$ to a function composed with a linear mapping. To this end, we recall the $r \times r$ difference matrix D_r by

$$(7.2) \quad D_r := \begin{bmatrix} 1 & & & & -1 \\ -1 & 1 & & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}.$$

Through the matrix Kronecker product \otimes , we define the $2d \times d$ matrix B by

$$(7.3) \quad B := \begin{bmatrix} I_{d_2} \otimes D_{d_1} \\ D_{d_2} \otimes I_{d_1} \end{bmatrix}.$$

Moreover, we define function $\psi : \mathbb{R}^{2d} \rightarrow \mathbb{R}$ at $y \in \mathbb{R}^{2d}$ as

$$(7.4) \quad \psi(y) := \sum_{i=1}^d \|[y_i, y_{d+i}]^\top\|_2.$$

With the definition of matrix B (7.3) and the convex function ψ (7.4), the (isotropic) total variation of an image x can be represented by

$$(7.5) \quad \|x\|_{\text{TV}} = \psi(Bx).$$

Moreover, we define $\varphi : \mathbb{R}^q \rightarrow \mathbb{R}$ at $y \in \mathbb{R}$ as $\varphi(y) := \|\Lambda y\|_1$. Then with help of the formula (7.5), function φ and the indicator function $\iota_{\{b\}}$, problem (7.1) can be equivalently reformulated as

$$(7.6) \quad \min\{\mu\psi(Bu) + \varphi(Wu) + \iota_{\{b\}}(Ku) : u \in \mathbb{R}^d\}.$$

Recall the dual problem of (7.6) has a form of

$$(7.7) \quad \min\{(\mu\psi)^*(x_1) + \varphi^*(x_2) + \iota_{\{b\}}^*(x_3) : B^\top x_1 + W^\top x_2 + K^\top x_3 = 0, x_1 \in \mathbb{R}^{2d}, x_2 \in \mathbb{R}^q, x_3 \in \mathbb{R}^p\}.$$

By the definition of the Fenchel conjugate function, one can easily check that the Fenchel conjugate functions in (7.7) 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

$$(7.8) \quad \min\{\iota_{S_1}(x_1) + \iota_{S_2}(x_2) + \langle b, x_3 \rangle : B^\top x_1 + W^\top x_2 + K^\top x_3 = 0, x_1 \in \mathbb{R}^{2d}, x_2 \in \mathbb{R}^q, x_3 \in \mathbb{R}^p\}.$$

Obviously, problem (7.8) is a special case of the multi-block problem (1.1) with the block number $s = 3$. Thus we can directly apply 2SFPPA to solving problem (7.8). In particular, all the proximity operators of the convex functions involved in (7.8) have closed forms. More precisely, the proximity operators $\text{prox}_{\frac{\alpha_1}{\beta} \iota_{S_1}}$ and $\text{prox}_{\frac{\alpha_2}{\beta} \iota_{S_2}}$ are exactly the projection operator onto the sets S_1 and S_2 respectively. The proximity operator $\text{prox}_{\frac{\alpha_3}{\beta} \langle \cdot, b \rangle}$ 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: $x_1^0 = K^\top b$, $x_2^0 = x_2^{-1} = 0$, $x_3^0 = x_3^{-1} = 0$, $y_0 = 0$.
 - 3: **repeat**
 - 4: Step 1: $x_1^{k+1} \leftarrow \text{Proj}_{S_1}(x_1^k - \alpha_1 B(B^\top x_1^k + W^\top(2x_2^k - x_2^{k-1}) + K^\top(2x_3^k - x_3^{k-1}) + \frac{1}{\beta}y^k))$,
 - 5:
 - 6: Step 2: $x_2^{k+1} \leftarrow \text{Proj}_{S_2}(x_2^k - \alpha_2 W(B^\top x_1^{k+1} + W^\top x_2^k + K^\top(2x_3^k - x_3^{k-1}) + \frac{1}{\beta}y^k))$,
 - 7:
 - 8: Step 3: $x_3^{k+1} \leftarrow x_3^k - \alpha_3 K(B^\top x_1^{k+1} + W^\top x_2^{k+1} + K^\top x_3^k + \frac{1}{\beta}y^k) - \frac{\alpha_3}{\beta}b$,
 - 9:
 - 10: Step 4: $y^{k+1} \leftarrow y^k + \beta(B^\top x_1^{k+1} + W^\top x_2^{k+1} + K^\top x_3^{k+1})$.
 - 11: **until** “convergence”
 - 12: Write the output of $-y^k$ from the above loop as u^∞ .
-

7.2. Numerical results. In this subsection, we shall compare numerical results of the proposed 2SFPPA with those of the Jacobi-type LADMM (JADMM) (6.2), 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.

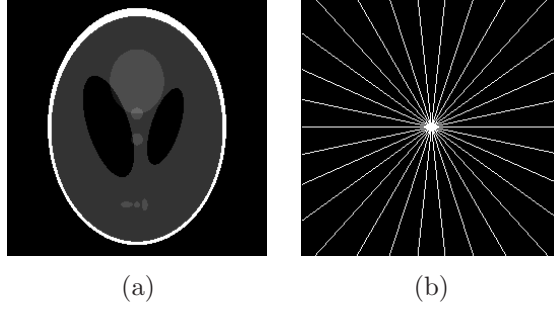


FIG. 1. (a) Shepp-Logan phantom (b) Sample pattern.

In the experiment, we select the 256×256 ‘‘Shepp-Logan’’ phantom as the test image, see Fig.1 (a). The observed data b is obtained by sampling the discrete Fourier transform of the phantom along 17 pseudo-radial lines, as shown in Fig.1 (b). 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 Λ as follows

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

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 (7.6). 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

$$\epsilon_1^k := (F(u^k) + \tau \|Ku^k - b\|_2 - F^*)/F^*,$$

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

$$\epsilon_2^k := \frac{\|y^k - y^{k-1}\|_2}{\|y^k\|_2}.$$

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^∞ is the recovered image vector.

For the JLADMM, we set

$$(7.9) \quad \alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{8} \text{ and } \beta = 1.$$

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

$$(7.10) \quad \alpha_1 = \frac{1}{8}, \alpha_2 = \frac{0.999999}{\|W\|_2^2}, \alpha_3 = \frac{0.999999}{\|K\|_2^2}, \text{ and } \beta = 1.$$

TABLE 1

Performance comparison for the sparse MRI. For a given error tolerance ϵ , the first column in the bracket represents the first iteration number k such that $e_1^k < \epsilon$, the second column and the third column in the bracket show the corresponding PSNR and CPU time.

	$\epsilon = 10^{-4}$	$\epsilon = 10^{-5}$	$\epsilon = 10^{-6}$
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)

TABLE 2

Performance comparison for the sparse MRI. For a given error tolerance ϵ , the first column in the bracket represents the first iteration number k such that $e_2^k < \epsilon$, the second column and the third column in the bracket show the corresponding PSNR and CPU time.

	$\epsilon = 5 \times 10^{-5}$	$\epsilon = 5 \times 10^{-6}$	$\epsilon = 5 \times 10^{-7}$
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)

Besides, as suggested in [16], the parameter θ involved in LADMMG is set to be 1. With such choice of parameters, all the four algorithms achieve their best performance in terms of the convergence speed.

Table 1 and Table 2 summarize the numbers of iteration, 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 LADMM and 2SFPPA due to the Gaussian back substitution step which ensures convergence of the algorithm. 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 that from JLADMM and LADMMG. Similarly, the sequence of PSNR values from 2SFPPA grows faster to the maximum value than that from JLADMM and LADMMG. Overall, we conclude that 2SFPPA performs as efficiently as LADMM and much better than JLADMM and LADMMG.

8. Conclusions. In this paper, we study the multi-block separable convex problem, which minimizes the sum of several convex functions with linear constraints. We develop a two-step fixed-point iterative scheme for solving the problem. We prove that the iterative scheme is convergent and has the convergence rate of $O(\frac{1}{k})$ in the ergodic sense and the sense of the partial primal-dual gap, where k denotes the iteration number. Based on the iterative scheme, we propose a class of convergent two-step algorithms for the multi-block separable convex problem. Convergence analysis for the specific algorithms can be carried out by verifying conditions on the matrices used to construct the algorithms. In the numerical experiments, we applied our two-step algorithms to the sparse MRI problems. Numerical results show that our proposed algorithms perform as efficiently as LADMM and outperform the JLADMM and LADMMG.

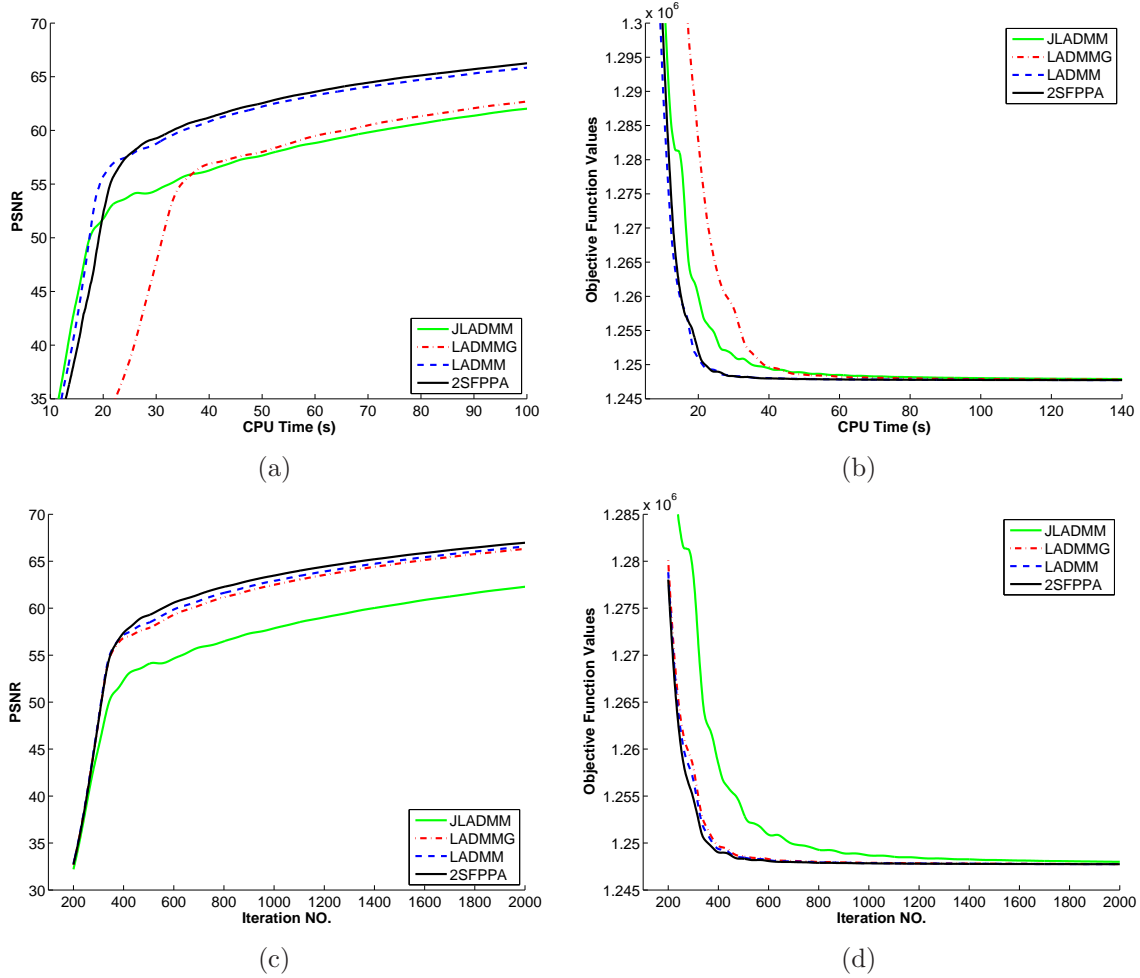


FIG. 2. (a) PSNR versus computational time, (b) objective function value versus computational time, (c) PSNR versus number of iterations, (d) objective function value versus number of iterations.

- [1] HEDY ATTOUCH, LUIS M. BRICENO-ARIAS, AND PATRICK L. COMBETTES, *A parallel splitting method for coupled monotone inclusions*, SIAM Journal on Control and Optimization, 48 (2010), pp. 3246–3270.
- [2] HEINZ H. BAUSCHKE AND PATRICK L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, AMS Books in Mathematics, Springer, New York, 2011.
- [3] JIANFENG CAI, RAYMOND CHAN, AND ZUOWEI SHEN, *A framelet-based image inpainting algorithm*, Applied and Computational Harmonic Analysis, 24 (2007), pp. 131–149.
- [4] JIANFENG CAI, STANLEY OSHER, AND ZUOWEI SHEN, *Linearized Bregman iteration for frame based image deblurring*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 226–252.
- [5] ANTONIN CHAMBOLLE AND THOMAS 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.
- [6] RAYMOND CHAN, TONY CHAN, LIXIN SHEN, AND ZUOWEI SHEN, *Wavelet algorithms for high-resolution image reconstruction*, SIAM Journal on Scientific Computing, 24 (2003), pp. 1408–1432.
- [7] RAYMOND CHAN, SHERMAN D. RIEMENSCHNEIDER, LIXIN SHEN, AND ZUOWEI SHEN, *Tight frame: The efficient way for high-resolution image reconstruction*, Applied and Computational Harmonic Analysis, 17 (2004), pp. 91–115.
- [8] CAIHUA CHEN, BINGSHENG HE, YINYU YE, AND XIAOMING YUAN, *The direct extension of admm for multi-block convex minimization problems is not necessarily convergent*, Mathematical Programming, (2014), pp. 1–23.
- [9] CORINNA CORTES AND VLADIMIR VAPNIK, *Support-vector networks*, Machine learning, 20 (1995), pp. 273–297.
- [10] DAMEK DAVIS AND WOTAO YIN, *A three-operator splitting scheme and its optimization applications*. UCLA CAM Report 15-13.
- [11] WEI DENG, MING-JUN LAI, ZHIMIN PENG, AND WOTAO YIN, *Parallel multi-block admm with $o(1/k)$ convergence*, UCLA

- CAM 13-64, (2014).
- [12] ERNIE ESSER, XIAOQUN ZHANG, AND TONY F. CHAN, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, Siam Journal on Imaging Sciences, 3 (2010), pp. 1015–1046.
 - [13] DANIEL GABAY AND BERTRAND MERCIER, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation. comput. math. appl.* 2(1), 17-40, Computers and Mathematics with Applications, 2 (1976), pp. 17–40.
 - [14] TOM GOLDSTEIN AND STANLEY OSHER, *The split Bregman method for ℓ^1 regularization problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 323–343.
 - [15] BINGSHENG HE, MIN TAO, AND XIAOMING YUAN, *Alternating direction method with gaussian back substitution for separable convex programming*, SIAM Journal on Optimization, 22 (2012), pp. 313–340.
 - [16] BINGSHENG HE AND XIAOMING YUAN, *Linearized alternating direction method of multipliers with gaussian back substitution for seperable convex programming*, Numerical Algebra Control and Optimization, 22 (2013), pp. 247–260.
 - [17] ———, *The direct extension of admm for three-block separable convex minimization models is convergent when one function is strongly convex*, manuscript, (2014).
 - [18] MIN LI, DEFENG SUN, AND KIM-CHUAN TOH, *A convergent 3-block semi-proximal admm for convex minimization problems with one strongly convex block*, Asia-Pacific Journal of Operational Research, (2015).
 - [19] QIA LI, LIXIN SHEN, YUESHENG XU, AND NA ZHANG, *Multi-step fixed-point proximity algorithms for solving a class of convex optimization problems arising from image processing*, Advances in Computational Mathematics, 41 (2015), pp. 387–422.
 - [20] QIA LI, LIXIN SHEN, AND LIHUA YANG, *Split-bregman iteration for framelet based image inpainting*, Applied and Computational Harmonic Analysis, 32 (2012), pp. 145–154.
 - [21] QIA LI AND NA ZHANG, *Fast proximity-gradient algorithms for structured convex optimization problems*, Applied and Computational Harmonic Analysis, (2015, accepted).
 - [22] TIANYI LIN, SHIQIAN MA, AND SHUZHONG ZHANG, *On the convergence rate of multi-block admm*, manuscript, (2014).
 - [23] MICHAEL LUSTIG, DAVID DONOHO, AND JOHN M. PAULY, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnetic Resonance in Medicine, 58 (2007), pp. 1182–1195.
 - [24] CHARLES A. MICCHELLI, LIXIN SHEN, AND YUESHENG XU, *Proximity algorithms for image models: Denoising*, Inverse Problems, 27 (2011), p. 045009(30pp).
 - [25] JEAN 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.
 - [26] MILLA NIKOLOVA, *Local strong homogeneity of a regularized estimator*, SIAM Journal of Applied Mathematics, 61 (2000), pp. 633–658.
 - [27] R. TYRRELL ROCKAFELLAR, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization, 14 (1976), pp. 877–898.
 - [28] LEONID I. RUDIN AND STANLEY OSHER, *Total variation based image restoration with free local constraints*, in IEEE International Conference on Image Processing, 1994, pp. 31–35.
 - [29] ANDRZEJ RUSZCZYNSKI, *Parallel decomposition of multistage stochastic programming problems*, Mathematical Programming, 58 (1993), pp. 201–228.
 - [30] ALEX SAWATZKY, QI XU, CARSTEN O. SCHIRRA, AND MARK A. ANASTASIO, *Proximal admm for multi-channel image reconstruction in spectral x-ray ct*, Medical Imaging IEEE Transactions on, 33 (2014), pp. 1657 – 1668.
 - [31] WEI SHI, WEI LING, WEI WU, AND WOTAO YIN, *Extra: an exact first-order algorithm for decentralized consensus optimization*, SIAM Journal on Optimization, (2015). To appear.
 - [32] DEFENG SUN, KIM CHUAN TOH, AND LIUQIN YANG, *A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type of constraints*, SIAM Journal on Optimization, 25 (2014).
 - [33] ROBERT TIBSHIRANI, MICHAEL SAUNDERS, SAHARON ROSSET, JI ZHU, AND KEITH KNIGHT, *Sparsity and smoothness via the fused lasso*, Journal of the Royal Statistical Society, 67 (2005), pp. 91–108.
 - [34] ROBERT TIBSHIRANI, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society, 58 (1996), pp. 267–288.
 - [35] ZAIWEN WEN, DONALD GOLDFARB, AND WOTAO YIN, *Alternating direction augmented lagrangian methods for semidefinite programming*, Mathematical Programming Computation, 2 (2010), pp. 203–230.