

RESEARCH

Open Access



# A primal-dual fixed point algorithm for minimization of the sum of three convex separable functions

Peijun Chen<sup>1,2,3</sup>, Jianguo Huang<sup>1</sup> and Xiaoqun Zhang<sup>1,4\*</sup>

\*Correspondence:

xqzhang@sjtu.edu.cn

<sup>1</sup>Schools of Mathematical Sciences, and MOE-LSC, Shanghai Jiao Tong University, 800, Dongchuan Road, Shanghai, China

<sup>4</sup>Institute of Natural Sciences, Shanghai Jiao Tong University, 800, Dongchuan Road, Shanghai, China  
Full list of author information is available at the end of the article

## Abstract

Many problems arising in image processing and signal recovery with multi-regularization and constraints can be formulated as minimization of a sum of three convex separable functions. Typically, the objective function involves a smooth function with Lipschitz continuous gradient, a linear composite nonsmooth function, and a nonsmooth function. In this paper, we propose a primal-dual fixed point (PDFP) scheme to solve the above class of problems. The proposed algorithm for three-block problems is a symmetric and fully splitting scheme, only involving an explicit gradient, a linear transform, and the proximity operators which may have a closed-form solution. We study the convergence of the proposed algorithm and illustrate its efficiency through examples on fused LASSO and image restoration with non-negative constraint and sparse regularization.

**Keywords:** primal-dual fixed point algorithm; convex separable minimization; proximity operator; sparsity regularization

## 1 Introduction

In this paper, we aim to design a primal-dual fixed point algorithmic framework for solving the following minimization problem:

$$\min_{x \in \mathbb{R}^n} f_1(x) + (f_2 \circ B)(x) + f_3(x), \quad (1.1)$$

where  $f_1, f_2$ , and  $f_3$  are three proper lower semi-continuous convex functions, and  $f_1$  is differentiable on  $\mathbb{R}^n$  with a  $1/\beta$ -Lipschitz continuous gradient for some  $\beta \in (0, +\infty]$ , while  $B: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a linear transformation. This formulation covers a wide application in image processing and signal recovery with multi-regularization terms and constraints. For instance, in many imaging and data processing applications, the functional  $f_1$  corresponds to a data-fidelity term, and the last two terms are used for regularization. As a direct example of (1.1), we can consider the fused LASSO penalized problem [1, 2] defined by

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - a\|^2 + \mu_1 \|Bx\|_1 + \mu_2 \|x\|_1.$$

On the other hand, in the imaging science, total variation regularization with  $B$  being the discrete gradient operator together with  $\ell_1$  regularization has been adopted in some

image restoration applications, for example in [3]. Another useful application corresponds to  $f_3 = \chi_C$ , where  $\chi_C$  is the indicator function of a nonempty closed convex set  $C$ . In this case, the problem reduces to

$$\min_{x \in C} f_1(x) + (f_2 \circ B)(x). \tag{1.2}$$

As far as we know, Combettes and Pesquet first proposed a fully splitting algorithm in [4] to solve monotone operator inclusions problems, which include (1.1) as a special case. Condat [5] tackled the same problem and proposed a primal-dual splitting scheme. Extensions to multi-block composite functions are also discussed in detail. For the special case  $B = I$  ( $I$  denotes the usual identity operator), Davis and Yin [6] proposed a three-operator splitting scheme based on monotone operators. For the case that the problem (1.1) reduces to two-block separable functions, many splitting and proximal algorithms have been proposed and studied in the literature. Among them, extensive research have been conducted on the alternating direction of multiplier method (ADMM) [7] (also known as split Bregman [3]; see for example [8] and the references therein). The primal-dual hybrid gradient method (PDHG) [9–12], also known as the Chambolle-Pock algorithm [11], is another class of popular algorithm, largely adopted in imaging applications. In [13–16], several completely decoupled schemes, such as the inexact Uzawa solver and primal-dual fixed point algorithm, are proposed to avoid subproblem solving for some typical  $\ell_1$  minimization problems. Komodakis and Pesquet [17] recently gave a nice overview of recent primal-dual approaches for solving large-scale optimization problems (1.1). A general class of multi-step fixed point proximity algorithms is proposed in [18], which covers several existing algorithms [11, 12] as special cases. In the preparation of this paper, we notice that Li and Zhang [19] also studied the problem (1.1) and introduced a quasi-Newton and the overrelaxation strategies for accelerating the algorithms. Both algorithms can be viewed as a generalization of Condat’s algorithm [5]. The theoretical analysis is established based on the multi-step techniques present in [18].

In the following, we mainly review some most relevant work for a concise presentation. Problem (1.2) has been studied in [20] in the context of maximum a posterior ECT reconstruction, and a preconditioned alternating projection algorithm (PAPA) is proposed for solving the resulting regularization problem. For  $f_3 = 0$  in (1.1), we proposed the primal-dual fixed point algorithm PDFP<sup>2</sup>O (primal-dual fixed point algorithm based on proximity operator) in [15]. Based on the fixed point theory, we have shown the convergence of the scheme PDFP<sup>2</sup>O and the convergence rate of the iteration sequence under suitable conditions.

In this work, we aim to extend the ideas of the PDFP<sup>2</sup>O in [15] and the PAPA in [20] for solving (1.1) without subproblem solving and provide a convergence analysis on the primal-dual sequences. The specific algorithm, namely the primal-dual fixed point (PDFP) algorithm, is formulated as follows:

$$\text{(PDFP)} \quad \begin{cases} y^{k+1} = \text{prox}_{\gamma f_3}(x^k - \gamma \nabla f_1(x^k) - \lambda B^T v^k), \\ v^{k+1} = (I - \text{prox}_{\frac{\gamma}{\lambda} f_2})(B y^{k+1} + v^k), \\ x^{k+1} = \text{prox}_{\gamma f_3}(x^k - \gamma \nabla f_1(x^k) - \lambda B^T v^{k+1}), \end{cases} \tag{1.3}$$

where  $0 < \lambda < 1/\lambda_{\max}(BB^T)$ ,  $0 < \gamma < 2\beta$ . Here  $\text{prox}_f$  is the proximity operator [21] of a function  $f$ ; see (2.2). When  $f_3 = \chi_C$ , the proposed algorithm (1.3) is reduced to the PAPA proposed in [20]; see (4.1). For another special case,  $f_3 = 0$  in (1.3), we obtain the PDFP<sup>2</sup>O proposed in [15]; see (4.2). The convergence analysis of this PDFP algorithm is built upon fixed point theory on the primal and dual pairs. The overall scheme is completely explicit, which allows for an easy implementation and parallel computing for many large-scale applications. This will be further illustrated through application to the problems arising in statistics learning and image restoration. The PDFP has a symmetric form and it is different from Condat’s algorithm [5]. In addition, we point out that the ranges of the parameters in PDFP are larger than those of [5, 19] and the rules for the parameters in PDFP are well separated, which could be advantageous in practice compared to [5, 19].

The rest of the paper is organized as follows. In Section 2, we will present some preliminaries and notations, and deduce PDFP from the first order optimality condition. In Section 3, we will provide the convergence results and the linear convergence rate results for some special cases. In Section 4, we will make a comparison on the form of the PDFP algorithm (1.3) with some existing algorithms. In Section 5, we will show the numerical performance and the efficiency of PDFP through some examples on fused LASSO and pMRI (parallel magnetic resonance image) reconstruction.

## 2 Primal-dual fixed point algorithm

### 2.1 Preliminaries and notations

For the self completeness of this work, we list some relevant notations, definitions, assumption and lemmas in convex analysis. We refer the reader to [15, 22] and the references therein for more details.

For the ease of presentation, we restrict our discussion to Euclidean space  $\mathbb{R}^n$ , equipped with the usual inner product  $\langle \cdot, \cdot \rangle$  and norm  $\| \cdot \| = \langle \cdot, \cdot \rangle^{1/2}$ . We first assume that the problem (1.1) has at least one solution and  $f_2, f_3, B$  satisfy

$$0 \in \text{ri}(\text{dom}_{f_2} - B(\text{dom}_{f_3})), \tag{2.1}$$

where the symbol  $\text{ri}(\cdot)$  denotes the interior of a convex subset, and the effective domain of  $f$  is defined as  $\text{dom}_f = \{x \in \mathbb{R}^n \mid f(x) < +\infty\}$ .

The  $\ell_1$  norm of a vector  $x \in \mathbb{R}^n$  is denoted by  $\| \cdot \|_1$  and the spectral norm of a matrix is denoted by  $\| \cdot \|_2$ . Let  $\Gamma_0(\mathbb{R}^n)$  be the collection of all proper lower semi-continuous convex functions from  $\mathbb{R}^n$  to  $(-\infty, +\infty]$ . For a function  $f \in \Gamma_0(\mathbb{R}^n)$ , the proximity operator of  $f$ :  $\text{prox}_f$  [21] is defined by

$$\text{prox}_f(x) = \arg \min_{y \in \mathbb{R}^n} f(y) + \frac{1}{2} \|x - y\|^2. \tag{2.2}$$

For a nonempty closed convex set  $C \subset \mathbb{R}^n$ , let  $\chi_C$  be the indicator function of  $C$ , defined by

$$\chi_C(x) = \begin{cases} 0, & x \in C, \\ +\infty, & x \notin C. \end{cases}$$

Let  $\text{proj}_C$  be the projection operator onto  $C$ , *i.e.*

$$\text{proj}_C(x) = \arg \min_{y \in C} \|x - y\|^2.$$

It is easy to see that  $\text{prox}_{\gamma \chi_C} = \text{proj}_C$  for all  $\gamma > 0$ , and the proximity operator is a generalization of projection operator. Note that many efficient splitting algorithms rely on the fact that  $\text{prox}_f$  has a closed-form solution. For example, when  $f = \gamma \|\cdot\|_1$ , the proximity solution is given by element-wise soft-shrinking. We refer the reader to [22] for more details as regards proximity operators. Let  $\partial f$  be the subdifferential of  $f$ , *i.e.*

$$\partial f(x) = \{v \in \mathbb{R}^n \mid \langle y - x, v \rangle \leq f(y) - f(x) \text{ for all } y \in \mathbb{R}^n\}, \tag{2.3}$$

and  $f^*$  be the convex conjugate function of  $f$ , defined by

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

An operator  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is nonexpansive if

$$\|Tx - Ty\| \leq \|x - y\| \quad \text{for all } x, y \in \mathbb{R}^n,$$

and  $T$  is firmly nonexpansive if

$$\|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y \rangle \quad \text{for all } x, y \in \mathbb{R}^n.$$

It is obvious that a firmly nonexpansive operator is nonexpansive. An operator  $T$  is  $\delta$ -strongly monotone if there exists a positive real number  $\delta$  such that

$$\langle Tx - Ty, x - y \rangle \geq \delta \|x - y\|^2 \quad \text{for all } x, y \in \mathbb{R}^n. \tag{2.4}$$

**Lemma 2.1** *For any two functions  $f_2 \in \Gamma_0(\mathbb{R}^m)$  and  $f_3 \in \Gamma_0(\mathbb{R}^n)$ , and a linear transformation  $B : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , satisfying that  $0 \in \text{ri}(\text{dom}_{f_2} - B(\text{dom}_{f_3}))$ , we have*

$$\partial(f_2 \circ B + f_3) = B^T \circ \partial f_2 \circ B + \partial f_3.$$

**Lemma 2.2** *Let  $f \in \Gamma_0(\mathbb{R}^n)$ . Then  $\text{prox}_f$  and  $I - \text{prox}_f$  are firmly nonexpansive. In addition,*

$$x = \text{prox}_f(y) \iff y - x \in \partial f(x) \quad \text{for a given } y \in \mathbb{R}^n, \tag{2.5}$$

$$\begin{aligned} y \in \partial f(x) &\iff x = \text{prox}_f(x + y) \\ &\iff y = (I - \text{prox}_f)(x + y) \quad \text{for } x, y \in \mathbb{R}^n, \end{aligned} \tag{2.6}$$

$$x = \text{prox}_{\gamma f}(x) + \gamma \text{prox}_{\frac{1}{\gamma} f^*}\left(\frac{1}{\gamma}x\right) \quad \text{for all } x \in \mathbb{R}^n \text{ and } \gamma > 0. \tag{2.7}$$

*If  $f$  has  $1/\beta$ -Lipschitz continuous gradient further, we have*

$$\beta \|\nabla f(x) - \nabla f(y)\|^2 \leq \langle \nabla f(x) - \nabla f(y), x - y \rangle \quad \text{for all } x, y \in \mathbb{R}^n. \tag{2.8}$$

**Lemma 2.3** *Let  $T$  be an operator and  $u^*$  be a fixed point of  $T$ . Let  $\{u^{k+1}\}$  be the sequence generated by the fixed point iteration  $u^{k+1} = T(u^k)$ . Suppose (i)  $T$  is continuous, (ii)  $\{\|u^k - u^*\|\}$  is non-increasing, (iii)  $\lim_{k \rightarrow +\infty} \|u^{k+1} - u^k\| = 0$ . Then the sequence  $\{u^k\}$  is bounded and converges to a fixed point of  $T$ .*

The proof of Lemma 2.3 is standard, and we refer the reader to the proof of Theorem 3.5 in [15] for more details.

Let  $\gamma$  and  $\lambda$  be two positive numbers. To simplify the presentation, we use the following notations:

$$T_0(v, x) = \text{prox}_{\gamma f_3}(x - \gamma \nabla f_1(x) - \lambda B^T v), \tag{2.9}$$

$$T_1(v, x) = (I - \text{prox}_{\frac{\gamma}{\lambda} f_2})(B \circ T_0(v, x) + v), \tag{2.10}$$

$$T_2(v, x) = \text{prox}_{\gamma f_3}(x - \gamma \nabla f_1(x) - \lambda B^T \circ T_1(v, x)), \tag{2.11}$$

$$T(v, x) = (T_1(v, x), T_2(v, x)). \tag{2.12}$$

Denote

$$g(x) = x - \gamma \nabla f_1(x) \quad \text{for all } x \in \mathbb{R}^n, \tag{2.13}$$

$$M = I - \lambda B B^T. \tag{2.14}$$

Let  $\lambda_{\max}(A)$  denote the largest eigenvalue of a square matrix  $A$ . When  $0 < \lambda < 1/\lambda_{\max}(B B^T)$ ,  $M$  is a symmetric and positive definite matrix, so we can define a norm

$$\|v\|_M = \sqrt{\langle v, Mv \rangle} \quad \text{for all } v \in \mathbb{R}^m. \tag{2.15}$$

For a pair  $u = (v, x) \in \mathbb{R}^m \times \mathbb{R}^n$ , we also define a norm on the product space  $\mathbb{R}^m \times \mathbb{R}^n$  as

$$\|u\|_\lambda = \sqrt{\lambda \|v\|^2 + \|x\|^2}. \tag{2.16}$$

### 2.2 Derivation of PDFP

On extending the ideas of the PAPA proposed in [20] and the PDFP<sup>2</sup>O proposed in [15], we derive the primal-dual fixed point algorithm (1.3) for solving the minimization problem (1.1).

Under the assumption (2.1), by using the first order optimality condition of (1.1) and Lemma 2.1, we have

$$0 \in \gamma \nabla f_1(x^*) + \lambda B^T \partial \left( \frac{\gamma}{\lambda} f_2 \right) (Bx^*) + \gamma \partial f_3(x^*),$$

where  $x^*$  is an optimal solution. Let

$$v^* \in \partial \left( \frac{\gamma}{\lambda} f_2 \right) (Bx^*). \tag{2.17}$$

By applying (2.6), we have

$$v^* = (I - \text{prox}_{\frac{\gamma}{\lambda} f_2})(Bx^* + v^*), \tag{2.18}$$

$$x^* = \text{prox}_{\gamma f_3}(x^* - \gamma \nabla f_1(x^*) - \lambda B^T v^*). \tag{2.19}$$

By inserting (2.19) into (2.18), we get

$$v^* = (I - \text{prox}_{\frac{\gamma}{\lambda} f_2})(B \circ \text{prox}_{\gamma f_3}(x^* - \gamma \nabla f_1(x^*) - \lambda B^T v^*) + v^*),$$

or equivalently,  $v^* = T_1(v^*, x^*)$ . Next, replacing  $v^*$  in (2.19) by  $T_1(v^*, x^*)$ , we can get  $x^* = T_2(v^*, x^*)$ . In other words  $u^* = T(u^*)$  for  $u^* = (v^*, x^*)$ . Meanwhile, if  $u^* = T(u^*)$ , we can see that  $x^*$  meets the first order optimality condition of (1.1) and thus  $x^*$  is a minimizer of (1.1).

To sum up, we have the following theorem.

**Theorem 2.1** *Suppose that  $x^*$  is a solution of (1.1) and  $v^* \in \mathbb{R}^m$  is defined as (2.17). Then we have*

$$\begin{cases} v^* = T_1(v^*, x^*), \\ x^* = T_2(v^*, x^*), \end{cases}$$

*i.e.  $u^* = (v^*, x^*)$  is a fixed point of  $T$ . Conversely, if  $u^* = (v^*, x^*) \in \mathbb{R}^m \times \mathbb{R}^n$  is a fixed point of  $T$ , then  $x^*$  is a solution of (1.1).*

It is easy to confirm that the sequence  $\{(v^{k+1}, x^{k+1})\}$  generated by the PDFP algorithm (1.3) is the Picard iteration  $(v^{k+1}, x^{k+1}) = T(v^k, x^k)$ . So we will use the operator  $T$  to analyze the convergence of the PDFP in Section 3.

### 3 Convergence analysis

In the following, let  $\{y^{k+1}\}$  and  $\{u^{k+1} = (v^{k+1}, x^{k+1})\}$  be the sequences generated by the PDFP algorithm (1.3), *i.e.*  $y^{k+1} = T_0(v^k, x^k)$  and  $(v^{k+1}, x^{k+1}) = T(v^k, x^k)$ . Let  $u^* = (v^*, x^*)$  be a fixed point of the operator  $T$ .

#### 3.1 Convergence

**Lemma 3.1** *We have the following estimates:*

$$\|v^{k+1} - v^*\|^2 \leq \|v^k - v^*\|^2 - \|v^{k+1} - v^k\|^2 + 2\langle B^T(v^{k+1} - v^*), y^{k+1} - x^* \rangle, \tag{3.1}$$

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - \|x^{k+1} - y^{k+1}\|^2 - \|x^k - y^{k+1}\|^2 \\ &\quad + 2\langle x^{k+1} - y^{k+1}, \gamma \nabla f_1(x^k) + \lambda B^T v^k \rangle \\ &\quad - 2\langle x^{k+1} - x^*, \gamma \nabla f_1(x^k) + \lambda B^T v^{k+1} \rangle + 2\gamma \langle f_3(x^*) - f_3(y^{k+1}) \rangle. \end{aligned} \tag{3.2}$$

*Proof* We first prove (3.1). By Lemma 2.2, we know  $I - \text{prox}_{\frac{\gamma}{\lambda} f_2}$  is firmly nonexpansive, and using (1.3)<sub>2</sub> and (2.18) we further have

$$\|v^{k+1} - v^*\|^2 \leq \langle v^{k+1} - v^*, (By^{k+1} + v^k) - (Bx^* + v^*) \rangle,$$

which implies

$$\langle v^{k+1} - v^*, v^{k+1} - v^k \rangle \leq \langle v^{k+1} - v^*, B(y^{k+1} - x^*) \rangle = \langle B^T(v^{k+1} - v^*), y^{k+1} - x^* \rangle.$$

Thus

$$\begin{aligned} \|v^{k+1} - v^*\|^2 &= \|v^k - v^*\|^2 - \|v^{k+1} - v^k\|^2 + 2\langle v^{k+1} - v^*, v^{k+1} - v^k \rangle \\ &\leq \|v^k - v^*\|^2 - \|v^{k+1} - v^k\|^2 + 2\langle B^T(v^{k+1} - v^*), y^{k+1} - x^* \rangle. \end{aligned}$$

Next we prove (3.2). By the optimality condition of (1.3)<sub>3</sub> (cf. (2.5)), we have

$$(x^k - \gamma \nabla f_1(x^k) - \lambda B^T v^{k+1}) - x^{k+1} \in \gamma \partial f_3(x^{k+1}).$$

By the property of subdifferentials (cf. (2.3)),

$$\langle x^* - x^{k+1}, (x^k - \gamma \nabla f_1(x^k) - \lambda B^T v^{k+1}) - x^{k+1} \rangle \leq \gamma (f_3(x^*) - f_3(x^{k+1})),$$

i.e.

$$\langle x^{k+1} - x^*, x^{k+1} - x^k \rangle \leq -\langle x^{k+1} - x^*, \gamma \nabla f_1(x^k) + \lambda B^T v^{k+1} \rangle + \gamma (f_3(x^*) - f_3(x^{k+1})).$$

Therefore,

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &= \|x^k - x^*\|^2 - \|x^{k+1} - x^k\|^2 + 2\langle x^{k+1} - x^*, x^{k+1} - x^k \rangle \\ &\leq \|x^k - x^*\|^2 - \|x^{k+1} - x^k\|^2 - 2\langle x^{k+1} - x^*, \gamma \nabla f_1(x^k) + \lambda B^T v^{k+1} \rangle \\ &\quad + 2\gamma (f_3(x^*) - f_3(x^{k+1})). \end{aligned} \tag{3.3}$$

On the other hand, by the optimality condition of (1.3)<sub>1</sub>, it follows that

$$(x^k - \gamma \nabla f_1(x^k) - \lambda B^T v^k) - y^{k+1} \in \gamma \partial f_3(y^{k+1}).$$

Thanks to the property of subdifferentials, we have

$$\langle x^{k+1} - y^{k+1}, (x^k - \gamma \nabla f_1(x^k) - \lambda B^T v^k) - y^{k+1} \rangle \leq \gamma (f_3(x^{k+1}) - f_3(y^{k+1})).$$

So

$$\langle x^{k+1} - y^{k+1}, x^k - y^{k+1} \rangle \leq \langle x^{k+1} - y^{k+1}, \gamma \nabla f_1(x^k) + \lambda B^T v^k \rangle + \gamma (f_3(x^{k+1}) - f_3(y^{k+1})).$$

Thus

$$\begin{aligned} -\|x^{k+1} - x^k\|^2 &= -\|x^{k+1} - y^{k+1}\|^2 - \|x^k - y^{k+1}\|^2 + 2\langle x^{k+1} - y^{k+1}, x^k - y^{k+1} \rangle \\ &\leq -\|x^{k+1} - y^{k+1}\|^2 - \|x^k - y^{k+1}\|^2 + 2\langle x^{k+1} - y^{k+1}, \gamma \nabla f_1(x^k) + \lambda B^T v^k \rangle \\ &\quad + 2\gamma (f_3(x^{k+1}) - f_3(y^{k+1})). \end{aligned}$$

Replacing the term  $-\|x^{k+1} - x^k\|^2$  in (3.3) with the right side term of the above inequality, we immediately obtain (3.2). □

**Lemma 3.2** *We have*

$$\begin{aligned} \|u^{k+1} - u^*\|_\lambda^2 &\leq \|u^k - u^*\|_\lambda^2 - \lambda \|v^{k+1} - v^k\|_M^2 - \|x^{k+1} - y^{k+1} + \lambda B^T(v^{k+1} - v^k)\|^2 \\ &\quad - \|(x^k - y^{k+1}) - (\gamma \nabla f_1(x^k) - \gamma \nabla f_1(x^*))\|^2 \\ &\quad - \gamma(2\beta - \gamma) \|\nabla f_1(x^k) - \nabla f_1(x^*)\|^2. \end{aligned} \tag{3.4}$$

*Proof* Summing the two inequalities (3.1) and (3.2) and re-arranging the terms, we have

$$\begin{aligned} &\lambda \|v^{k+1} - v^*\|^2 + \|x^{k+1} - x^*\|^2 \\ &\leq \lambda \|v^k - v^*\|^2 + \|x^k - x^*\|^2 - \lambda \|v^{k+1} - v^k\|^2 - \|x^{k+1} - y^{k+1}\|^2 - \|x^k - y^{k+1}\|^2 \\ &\quad + 2\langle \lambda B^T(v^{k+1} - v^*), y^{k+1} - x^* \rangle + 2\langle x^{k+1} - y^{k+1}, \gamma \nabla f_1(x^k) + \lambda B^T v^k \rangle \\ &\quad - 2\langle x^{k+1} - x^*, \gamma \nabla f_1(x^k) + \lambda B^T v^{k+1} \rangle + 2\gamma(f_3(x^*) - f_3(y^{k+1})) \\ &= \lambda \|v^k - v^*\|^2 + \|x^k - x^*\|^2 - \lambda \|v^{k+1} - v^k\|^2 - \|x^{k+1} - y^{k+1}\|^2 - \|x^k - y^{k+1}\|^2 \\ &\quad + 2\langle \lambda B^T(v^{k+1} - v^k), y^{k+1} - x^{k+1} \rangle + 2\langle x^k - y^{k+1}, \gamma \nabla f_1(x^k) - \gamma \nabla f_1(x^*) \rangle \\ &\quad - 2\langle x^k - x^*, \gamma \nabla f_1(x^k) - \gamma \nabla f_1(x^*) \rangle \\ &\quad + 2\langle (y^{k+1} - x^*, -\gamma \nabla f_1(x^*) - \lambda B^T v^*) + \gamma(f_3(x^*) - f_3(y^{k+1})) \rangle \\ &= \lambda \|v^k - v^*\|^2 + \|x^k - x^*\|^2 - \lambda \|v^{k+1} - v^k\|_M^2 - \|x^{k+1} - y^{k+1} + \lambda B^T(v^{k+1} - v^k)\|^2 \\ &\quad - \|(x^k - y^{k+1}) - (\gamma \nabla f_1(x^k) - \gamma \nabla f_1(x^*))\|^2 + \|\gamma \nabla f_1(x^k) - \gamma \nabla f_1(x^*)\|^2 \\ &\quad - 2\langle x^k - x^*, \gamma \nabla f_1(x^k) - \gamma \nabla f_1(x^*) \rangle \\ &\quad + 2\langle (y^{k+1} - x^*, -\gamma \nabla f_1(x^*) - \lambda B^T v^*) + \gamma(f_3(x^*) - f_3(y^{k+1})) \rangle, \end{aligned} \tag{3.5}$$

where  $\|\cdot\|_M$  is given in (2.14) and (2.15). Meanwhile, by the optimality condition of (2.19), we have

$$-\gamma \nabla f_1(x^*) - \lambda B^T v^* \in \gamma \partial f_3(x^*),$$

which implies

$$\langle y^{k+1} - x^*, -\gamma \nabla f_1(x^*) - \lambda B^T v^* \rangle + \gamma(f_3(x^*) - f_3(y^{k+1})) \leq 0. \tag{3.6}$$

On the other hand, it follows from (2.8) that

$$-\langle x^k - x^*, \nabla f_1(x^k) - \nabla f_1(x^*) \rangle \leq -\beta \|\nabla f_1(x^k) - \nabla f_1(x^*)\|^2. \tag{3.7}$$

Recalling (2.16), we immediately obtain (3.4) in terms of (3.5)-(3.7). □

**Lemma 3.3** *Let  $0 < \lambda < 1/\lambda_{\max}(BB^T)$  and  $0 < \gamma < 2\beta$ . Then the sequence  $\{\|u^k - u^*\|_\lambda\}$  is non-increasing and  $\lim_{k \rightarrow +\infty} \|u^{k+1} - u^k\|_\lambda = 0$ .*

*Proof* If  $0 < \lambda < 1/\lambda_{\max}(BB^T)$  and  $0 < \gamma < 2\beta$ , it follows from (3.4) that  $\|u^{k+1} - u^*\|_\lambda \leq \|u^k - u^*\|_\lambda$ , i.e. the sequence  $\{\|u^k - u^*\|_\lambda\}$  is non-increasing. Moreover, summing the inequalities

(3.4) from  $k = 0$  to  $k = +\infty$ , we get

$$\lim_{k \rightarrow +\infty} \|v^{k+1} - v^k\|_M = 0, \tag{3.8}$$

$$\lim_{k \rightarrow +\infty} \|x^{k+1} - y^{k+1} + \lambda B^T(v^{k+1} - v^k)\| = 0, \tag{3.9}$$

$$\lim_{k \rightarrow +\infty} \|(x^k - y^{k+1}) - (\gamma \nabla f_1(x^k) - \gamma \nabla f_1(x^*))\| = 0, \tag{3.10}$$

$$\lim_{k \rightarrow +\infty} \|\nabla f_1(x^k) - \nabla f_1(x^*)\| = 0. \tag{3.11}$$

The combination of (3.10) and (3.11) gives

$$\lim_{k \rightarrow +\infty} \|x^k - y^{k+1}\| = 0. \tag{3.12}$$

Noting that  $0 < \lambda < 1/\lambda_{\max}(BB^T)$ , we know  $M$  is symmetric and positive definite, so (3.8) is equivalent to

$$\lim_{k \rightarrow +\infty} \|v^{k+1} - v^k\| = 0. \tag{3.13}$$

Hence, we have from the above inequality and (3.9) that

$$\lim_{k \rightarrow +\infty} \|x^{k+1} - y^{k+1}\| = 0. \tag{3.14}$$

The combination of (3.12) and (3.14) then gives rise to

$$\lim_{k \rightarrow +\infty} \|x^{k+1} - x^k\| = 0. \tag{3.15}$$

According to (3.13), (3.15), and (2.16), we have  $\lim_{k \rightarrow +\infty} \|u^{k+1} - u^k\|_\lambda = 0$ . □

As a direct consequence of Lemma 3.3 and Lemma 2.3, we obtain the convergence of the PDFP as follows.

**Theorem 3.1** *Let  $0 < \lambda < 1/\lambda_{\max}(BB^T)$  and  $0 < \gamma < 2\beta$ . Then the sequence  $\{u^k\}$  is bounded and converges to a fixed point of  $T$ , and both  $\{x^k\}$  and  $\{y^k\}$  converge to a solution of (1.1).*

*Proof* By Lemma 2.2, both  $\text{prox}_{\gamma f_3}$  and  $I - \text{prox}_{\frac{\gamma}{\lambda} f_2}$  are firmly nonexpansive, thus the operator  $T$  defined by (2.9)-(2.12) is continuous. From Lemma 3.3, we know that the sequence  $\{\|u^k - u^*\|_\lambda\}$  is non-increasing and  $\lim_{k \rightarrow +\infty} \|u^{k+1} - u^k\|_\lambda = 0$ . By using Lemma 2.3, we know that the sequence  $\{u^k\}$  is bounded and converges to a fixed point of  $T$ . By using Theorem 2.1 and (3.14), we can conclude that both  $\{x^k\}$  and  $\{y^k\}$  converge to a solution of (1.1). □

**Remark 3.1** For the special case  $f_3 = 0$ , the PDFP reduces naturally to the PDFP<sup>2</sup>O (4.2) proposed in [15], where the conditions for the parameters are  $0 < \lambda \leq 1/\lambda_{\max}(BB^T)$ ,  $0 < \gamma < 2\beta$ . In the proof of Lemma 3.3, we utilize the positive definitiveness of  $M$  to obtain (3.13) from (3.8). So the condition for the parameter  $\lambda$  is slightly more restricted as  $0 < \lambda < 1/\lambda_{\max}(BB^T)$  in Lemma 3.3 and Theorem 3.1. When  $f_3 = 0$ , the conditions in the proof

of Lemma 3.3 can also be relaxed to  $0 < \lambda \leq 1/\lambda_{\max}(BB^T)$ . As a matter of fact, it is easy to check by the definition of  $y^{k+1}$  (see (1.3)<sub>1</sub>) and the optimality condition of (2.19) that

$$\|(x^k - y^{k+1}) - (\gamma \nabla f_1(x^k) - \gamma \nabla f_1(x^*))\|^2 = \|\lambda B^T(v^k - v^*)\|^2. \tag{3.16}$$

Observing that  $\|v^{k+1} - v^k\|^2 = \|v^{k+1} - v^k\|_M^2 + \lambda \|B^T(v^{k+1} - v^k)\|^2$ , we have by (3.16), (3.10) and (3.8) that  $\lim_{k \rightarrow +\infty} \|v^{k+1} - v^k\| = 0$ . Therefore we can derive the convergence whenever  $M$  is semi-positive definite for  $f_3 = 0$ .

**Remark 3.2** For the special case  $f_1 = 0$ , the problem (1.1) only corresponds to two proper lower semi-continuous convex functions. The convergence condition  $0 < \gamma < 2\beta$  in the PDFP becomes  $0 < \gamma < +\infty$ . Although  $\gamma$  is an arbitrary positive number in theory, the range of  $\gamma$  will affect the convergence speed and it is also a difficult problem to choose a best value in practice.

### 3.2 Linear convergence rate for special cases

In the following, we will show the convergence rate results with some additional assumptions on the basic problem (1.1). In particular, for  $f_3 = 0$ , the algorithm reduces to the PDFP<sup>2</sup>O proposed in [15]. The conditions for a linear convergence given there as Condition 3.1 in [15] is as follows: for  $0 < \lambda \leq 1/\lambda_{\max}(BB^T)$  and  $0 < \gamma < 2\beta$ , there exist  $\eta_1, \eta_2 \in [0, 1)$  such that

$$\begin{aligned} \|I - \lambda BB^T\|_2 &\leq \eta_1^2, \\ \|g(x) - g(y)\| &\leq \eta_2 \|x - y\| \quad \text{for all } x, y \in \mathbb{R}^n, \end{aligned} \tag{3.17}$$

where  $g(x)$  is given in (2.13). It is easy to see that a strongly convex function  $f_1$  satisfies the condition (3.17). For a general  $f_3$ , we need stronger conditions on the functions.

**Theorem 3.2** *Suppose that (3.17) holds and  $f_2^*$  is strongly convex. Then we have*

$$\|u^{k+1} - u^*\|_{(1+\lambda\delta/\gamma)\lambda} \leq \eta \|u^k - u^*\|_{(1+\lambda\delta/\gamma)\lambda},$$

where  $0 < \eta < 1$  is the convergence rate (indicated in the proof) and  $\delta > 0$  is a parameter describing the strongly monotone property of  $\partial f_2^*$  (cf. (2.4)).

*Proof* Use Moreau’s identity (cf. (2.7)) to get

$$(I - \text{prox}_{\frac{\gamma}{\lambda} f_2})(By^{k+1} + v^k) = \frac{\gamma}{\lambda} \text{prox}_{\frac{\lambda}{\gamma} f_2^*} \left( \frac{\lambda}{\gamma} By^{k+1} + \frac{\lambda}{\gamma} v^k \right).$$

So (1.3)<sub>2</sub> is equivalent to

$$\frac{\lambda}{\gamma} v^{k+1} = \text{prox}_{\frac{\lambda}{\gamma} f_2^*} \left( \frac{\lambda}{\gamma} By^{k+1} + \frac{\lambda}{\gamma} v^k \right). \tag{3.18}$$

According to the optimality condition of (3.18),

$$\frac{\lambda}{\gamma} By^{k+1} + \frac{\lambda}{\gamma} v^k - \frac{\lambda}{\gamma} v^{k+1} \in \frac{\lambda}{\gamma} \partial f_2^* \left( \frac{\lambda}{\gamma} v^{k+1} \right). \tag{3.19}$$

Similarly, according to the optimality condition of (2.18),

$$\frac{\lambda}{\gamma} Bx^* \in \frac{\lambda}{\gamma} \partial f_2^* \left( \frac{\lambda}{\gamma} v^* \right). \tag{3.20}$$

Observing that  $\partial f_2^*$  is  $\delta$ -strongly monotone, we have by (3.19) and (3.20)

$$\langle v^{k+1} - v^*, (By^{k+1} + v^k - v^{k+1}) - Bx^* \rangle \geq \frac{\lambda}{\gamma} \delta \|v^{k+1} - v^*\|^2,$$

i.e.

$$\langle v^{k+1} - v^*, v^{k+1} - v^k \rangle \leq \langle B^T(v^{k+1} - v^*), y^{k+1} - x^* \rangle - \frac{\lambda}{\gamma} \delta \|v^{k+1} - v^*\|^2.$$

Thus

$$\begin{aligned} \|v^{k+1} - v^*\|^2 &= \|v^k - v^*\|^2 - \|v^{k+1} - v^k\|^2 + 2\langle v^{k+1} - v^*, v^{k+1} - v^k \rangle \\ &\leq \|v^k - v^*\|^2 - \|v^{k+1} - v^k\|^2 + 2\langle B^T(v^{k+1} - v^*), y^{k+1} - x^* \rangle \\ &\quad - \frac{\lambda}{\gamma} \delta \|v^{k+1} - v^*\|^2. \end{aligned} \tag{3.21}$$

Summing the two inequalities (3.21) and (3.2), and then using the same argument for driving (3.5), we arrive at

$$\begin{aligned} \left(1 + \frac{\lambda}{\gamma} \delta\right) \lambda \|v^{k+1} - v^*\|^2 + \|x^{k+1} - x^*\|^2 &\leq \lambda \|v^k - v^*\|^2 + \|g(x^k) - g(x^*)\|^2 \\ &\leq \lambda \|v^k - v^*\|^2 + \eta_2^2 \|x^k - x^*\|^2, \end{aligned} \tag{3.22}$$

where we have also used the condition (3.17) and the inequality (3.6).

Let  $\eta_3 = 1/\sqrt{1 + \lambda\delta/\gamma}$  and  $\eta = \max\{\eta_2, \eta_3\}$ . It is clear that  $0 < \eta < 1$ . Hence, according to the notation (2.16), the estimate (3.22) can be rewritten as required.  $\square$

We note that a linear convergence rate for strongly convex  $f_2^*$  and  $f_3$  are obtained in [19]. They introduced two preconditioned operators for accelerating the algorithm, while a clear relation between the convergence rate and the preconditioned operators is still missing. Meanwhile, introducing preconditioned operators could be beneficial in practice, and we can also introduce a preconditioned operator to deal with  $\nabla f_1$  in our scheme. Since the analysis is rather similar to the current one, we will omit it in this paper.

#### 4 Connections to other algorithms

In this section, we present the connections of the PDFP algorithm to some algorithms proposed previously in the literature.

In particular, when  $f_3 = \chi_C$ , due to  $\text{prox}_{\gamma f_3} = \text{proj}_C$ , the proposed algorithm (1.3) is reduced to the PAPA proposed in [20]

$$\text{(PDFP)} \quad \begin{cases} y^{k+1} = \text{proj}_C(x^k - \gamma \nabla f_1(x^k) - \lambda B^T v^k), \\ v^{k+1} = (I - \text{prox}_{\frac{\gamma}{\lambda} f_2})(By^{k+1} + v^k), \\ x^{k+1} = \text{proj}_C(x^k - \gamma \nabla f_1(x^k) - \lambda B^T v^{k+1}), \end{cases} \tag{4.1}$$

where  $0 < \lambda < 1/\lambda_{\max}(BB^T)$ ,  $0 < \gamma < 2\beta$ . We note that the conditions of the parameters for the convergence of the PDFP are larger than those in [20]. Here we still refer to (4.1) as the PDFP, since the PAPA originally proposed in [20] incorporates other techniques such as diagonal preconditioning.

For the special case  $f_3 = 0$ , due to  $\text{prox}_{\gamma f_3} = I$ , we obtain the PDFP<sup>2</sup>O scheme proposed in [15]

$$(\text{PDFP}^2\text{O}) \begin{cases} y^{k+1} = x^k - \gamma \nabla f_1(x^k) - \lambda B^T v^k, \\ v^{k+1} = (I - \text{prox}_{\frac{\gamma}{\lambda} f_2})(By^{k+1} + v^k), \\ x^{k+1} = x^k - \gamma \nabla f_1(x^k) - \lambda B^T v^{k+1}, \end{cases} \tag{4.2}$$

where  $0 < \lambda \leq 1/\lambda_{\max}(BB^T)$ ,  $0 < \gamma < 2\beta$ . Recently, we notice that the PDFP<sup>2</sup>O reduces to the algorithm previously proposed for  $f_1(x) = \frac{1}{2} \|Ax - a\|^2$  by Loris and Verhoeven in [14]. The convergence and the convergence rate of the objective function were established in [14], but the convergence conditions are slightly more restrictive than the ones given in [15]. On the other hand, we emphasize that the PDFP<sup>2</sup>O algorithm can also be interpreted from the point of view of forward-backward operator splitting, as shown in [15, 16]. Moreover, the multi-block formulation was devised and analyzed in [16].

Based on the PDFP<sup>2</sup>O, we also proposed the PDFP<sup>2</sup>O<sub>C</sub> in [23] for  $f_3 = \chi_C$  as

$$(\text{PDFP}^2\text{O}_C) \begin{cases} y^{k+1} = x^k - \gamma \nabla f_1(x^k) - \lambda B^T v_1^k - \lambda v_2^k, \\ v_1^{k+1} = (I - \text{prox}_{\frac{\gamma}{\lambda} f_2})(By^{k+1} + v_1^k), \\ v_2^{k+1} = (I - \text{proj}_C)(y^{k+1} + v_2^k), \\ x^{k+1} = x^k - \gamma \nabla f_1(x^k) - \lambda B^T v_1^{k+1} - \lambda v_2^{k+1}, \end{cases} \tag{4.3}$$

where  $0 < \lambda \leq 1/(\lambda_{\max}(BB^T) + 1)$ ,  $0 < \gamma < 2\beta$ . A similar technique of extension to multi-composite functions has also been used in [5, 18, 24]. Compared to the PDFP (4.1), the algorithm PDFP<sup>2</sup>O<sub>C</sub> introduces an extra variable, while the PDFP requires two times projections. Most importantly, the primal variable at each iterate of the PDFP is feasible, but maybe not for that of the PDFP<sup>2</sup>O<sub>C</sub>. In addition, the permitted ranges of the parameters are also tighter in the PDFP<sup>2</sup>O<sub>C</sub>.

Another interesting special case is  $f_1 = 0$ . The scheme (1.3) reduces to

$$(\text{PDFP}) \begin{cases} y^{k+1} = \text{prox}_{\gamma f_3}(x^k - \lambda B^T v^k), \\ v^{k+1} = (I - \text{prox}_{\frac{\gamma}{\lambda} f_2})(By^{k+1} + v^k), \\ x^{k+1} = \text{prox}_{\gamma f_3}(x^k - \lambda B^T v^{k+1}), \end{cases} \tag{4.4}$$

where  $0 < \lambda < 1/\lambda_{\max}(BB^T)$ ,  $0 < \gamma < +\infty$ . It is easy to see that (4.4) is different from the PDHG method in [10, 11] and (4.4) has a symmetric step (4.4)<sub>1</sub> compared to the extrapolation step in the PDHG method.

Combettes and Pesquet first proposed a fully split algorithm in [4] to solve monotone inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators, which include (1.1) as a special case. The problem is recast as two-block inclusions and then solved with an error-tolerant primal-dual forward-backward-forward algorithm

**Table 1** The comparison between Condat ( $\rho_k = 1$ ) and PDFP

	Condat ( $\rho_k = 1$ )	PDFP
Form	$\bar{v}^{k+1} = \text{prox}_{\sigma f_2^*}(\sigma Bx^k + \bar{v}^k),$ $x^{k+1} = \text{prox}_{\tau f_3}(x^k - \tau \nabla f_1(x^k) - \tau B^T(2\bar{v}^{k+1} - \bar{v}^k))$	$y^{k+1} = \text{prox}_{\gamma f_3}(x^k - \gamma \nabla f_1(x^k) - \gamma B^T \bar{v}^k),$ $\bar{v}^{k+1} = \text{prox}_{\frac{\lambda}{\gamma} f_2^*}(\frac{\lambda}{\gamma} B y^{k+1} + \bar{v}^k),$ $x^{k+1} = \text{prox}_{\gamma f_3}(x^k - \gamma \nabla f_1(x^k) - \gamma B^T \bar{v}^{k+1})$
$f_1 \neq 0$	$\sigma \tau \lambda_{\max}(BB^T) + \tau/(2\beta) \leq 1$	$0 < \lambda < 1/\lambda_{\max}(BB^T), 0 < \gamma < 2\beta$
$f_1 = 0$	$0 < \sigma \tau \leq 1/\lambda_{\max}(BB^T)$	$0 < \lambda < 1/\lambda_{\max}(BB^T), 0 < \gamma < +\infty$
Relation	$\sigma = \lambda/\gamma, \tau = \gamma$	

as studied in [25]. Condat [5] tackled the same problem as given in (1.1) and proposed a primal-dual splitting scheme. For the special case with  $f_1 = 0$ , Condat’s algorithm reduces to the PDHG method in [11]. By grouping the multi-block as two blocks, the authors in [24] extended the PDHG algorithm [12] to the minimization of sum of multi-composite functions. The authors in [18] proposed a class of multi-step fixed point proximity algorithms, including several existing algorithms as special examples, for example the algorithms in [11, 12]. In [6], Davis and Yin proposed a three-operator splitting method for solving three-block monotone inclusions in a very tricky way. When solving the problem (1.1) with  $B = I$ , the scheme is different from Condat’s algorithm and PDFP algorithm. But it requires sub-problem solving if  $B \neq I$ . Li and Zhang [19] studied (1.1) based on the techniques present in [18] and including Condat’s algorithm in [5] as a special case, and further introduced quasi-Newton and the overrelaxation strategies to accelerate the algorithms.

In the following, we mainly compare PDFP to the basic Algorithm 3.2 proposed by Condat in [5] to simplify the presentation. We first change the form of the PDFP algorithm (1.3) by using Moreau’s identity, see (2.7), *i.e.*

$$(I - \text{prox}_{\frac{\gamma}{\lambda} f_2})(By^{k+1} + v^k) = \frac{\gamma}{\lambda} \text{prox}_{\frac{\lambda}{\gamma} f_2^*} \left( \frac{\lambda}{\gamma} B y^{k+1} + \bar{v}^k \right),$$

where  $\bar{v}^k = \frac{\lambda}{\gamma} v^k$ . A direct comparison is presented in Table 1. From Table 1, we can see that the ranges of the parameters in Condat’s algorithm are relatively smaller than PDFP. Also since the condition for Condat’s algorithm is mixed with all the parameters, it is not always easy to choose them in practice. This is also pointed out in [5]. While the rules for the parameters in PDFP are separate, and they can be chosen independently according to the Lipschitz constant and the operator norm of  $BB^T$ . In this sense, our parameter rules are relatively more practical. In the numerical experiments, we can set  $\lambda$  to be close to  $1/\lambda_{\max}(BB^T)$  and  $\gamma$  to be close to  $2\beta$  for most of tests. Moreover, the results of  $x^k - \gamma \nabla f_1(x^k)$  and  $\lambda B^T v^{k+1}$  can be stored as two intermediate variables that can be reused in (1.3)<sub>1</sub> and (1.3)<sub>3</sub> during the iterations. Nevertheless, PDFP has an extra step (1.3)<sub>1</sub> compared to Condat’s algorithm and the computation cost may increase due to the computation of  $\text{prox}_{\gamma f_3}$ . In practice, this step is often related to the  $\ell_1$  shrinkage or projection operation being easy to implement, so the cost could be still ignorable in practice.

**5 Numerical experiments**

In this section, we will apply the PDFP algorithm to solve two problems: the fused LASSO penalized problem and parallel magnetic resonance imaging (pMRI) reconstruction. All the experiments are implemented under MATLAB 7.00 (R14) and conducted on a computer with Intel (R) core (TM) i5-4300U CPU@1.90G.

### 5.1 The fused LASSO penalized problem

The fused LASSO (least absolute shrinkage and selection operator) penalized problem is proposed for group variable selection, and we refer the reader to [1, 2, 26] for more details for the applications of this model. It can be described as

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - a\|^2 + \mu_1 \sum_{i=1}^{n-1} |x_{i+1} - x_i| + \mu_2 \|x\|_1.$$

Here  $A \in \mathbb{R}^{r \times n}$ ,  $a \in \mathbb{R}^r$ . The row of  $A$ :  $A_i$  for  $i = 1, 2, \dots, r$  represent the  $i$ th observation of the independent variables and  $a_i$  denotes the response variable, and the vector  $x \in \mathbb{R}^n$  is the regression coefficient to recover. The first term is corresponding to the data-fidelity term, and the last two terms aim to ensure the sparsity in both  $x$  and their successive differences in  $x$ . Let

$$B = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix}.$$

Then the forgoing problem can be reformulated as

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - a\|^2 + \mu_1 \|Bx\|_1 + \mu_2 \|x\|_1. \tag{5.1}$$

For this example, we can set  $f_1(x) = \frac{1}{2} \|Ax - a\|^2, f_2 = \mu_1 \|\cdot\|_1, f_3 = \mu_2 \|\cdot\|_1$ . We want to show that the PDFP algorithm (1.3) can be applied to solve this generic class of problems (5.1) directly and easily.

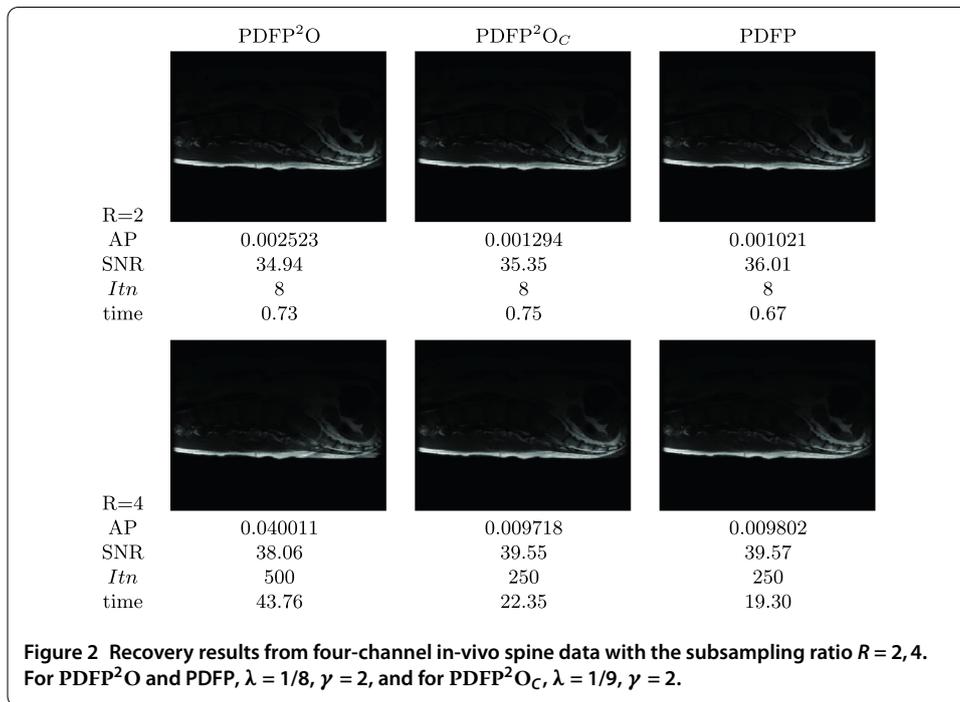
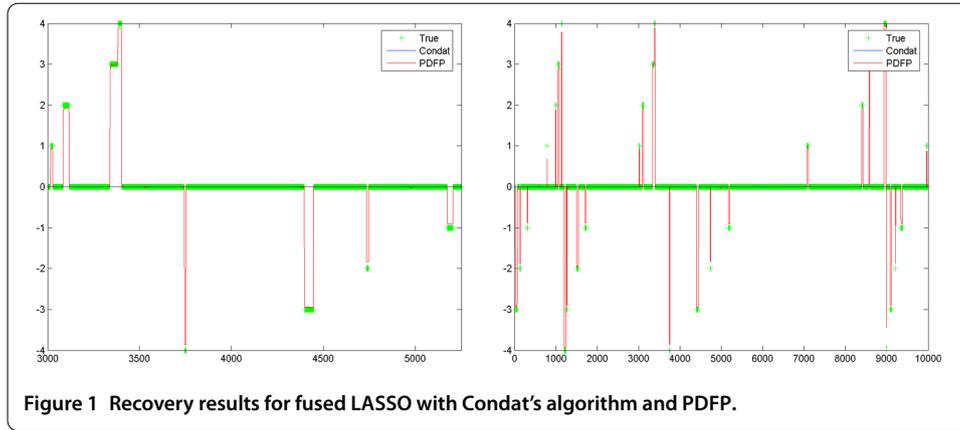
The following tests are designed for the simulation. We set  $r = 500, n = 10,000$ , and the data  $a$  is generated as  $Ax + \alpha e$ , where  $A$  and  $e$  are random matrices whose elements are normally distributed with zero mean and variance 1, and  $\alpha = 0.01$ , and  $x$  is a generated sparse vector, whose nonzero elements are showed in Figure 1 by green '+'. We set  $\mu_1 = 200, \mu_2 = 20$ , and the maximum iteration number as  $Itn = 1,500$ .

We compare the PDFP algorithm with Condat's algorithm [5]. For the PDFP algorithm, the parameters  $\lambda$  and  $\gamma$  are chosen according to Theorem 3.1. In practice, we set  $\lambda$  to be close to  $1/\lambda_{\max}(BB^T)$  and  $\gamma$  to be close to  $2\beta$ . Here we set  $\lambda = 1/4$  as the  $n - 1$  eigenvalues of  $BB^T$  can be analytically computed as  $2 - 2 \cos(i\pi/n), i = 1, 2, \dots, n - 1$  and  $\gamma = 1.99/\lambda_{\max}(A^T A)$ . For Condat's algorithm, we set  $\lambda = 0.19/4, \gamma = 1.9/\lambda_{\max}(A^T A)$ , which is chosen for a relative better numerical performance. The computation time, the attained objective function values, and the relative errors to the true solution are close for Condat's algorithm and PDFP. From Figure 1, we see that both Condat's algorithm and PDFP can quite correctly recover the positions of the non-zeros and the values.

### 5.2 Image restoration with non-negative constraint and sparse regularization

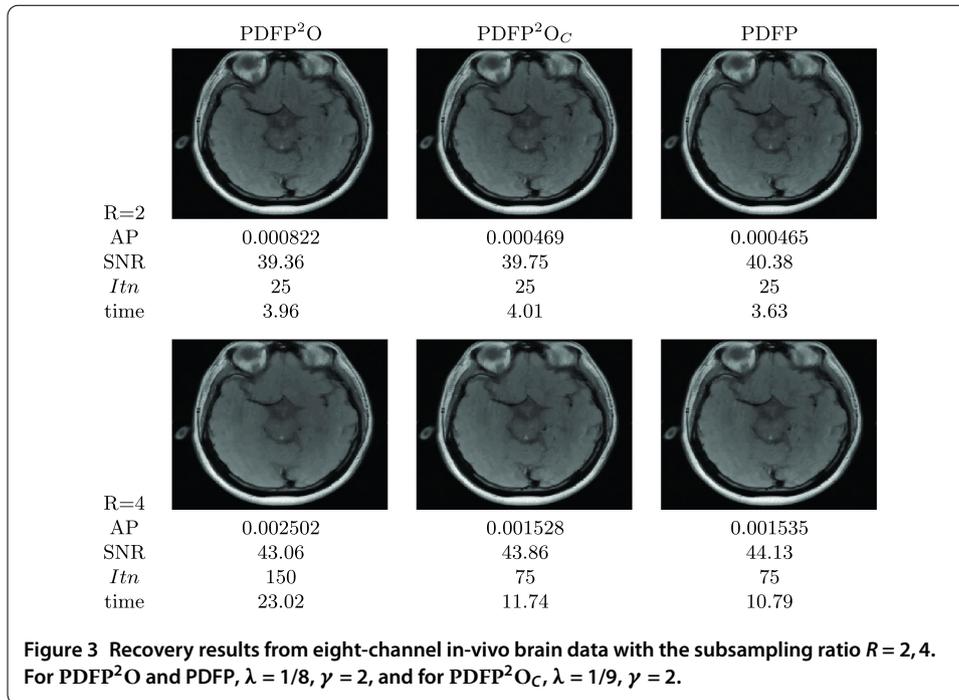
A general image restoration problem with non-negative constraint and sparse regularization can be written as

$$\min_{x \in C} \frac{1}{2} \|Ax - a\|^2 + \mu \|Bx\|_1, \tag{5.2}$$



where  $A$  is some linear operator describing the image formation process,  $\|Bx\|_1$  is the usual  $\ell_1$  based regularization in order to promote sparsity under the transform  $B$ ,  $\mu > 0$  is the regularization parameter. Here we use the isotropic total variation as the regularization functional, thus the matrix  $B$  represents the discrete gradient operator. For this example, we can set  $f_1(x) = \frac{1}{2} \|Ax - a\|^2, f_2 = \mu \|\cdot\|_1$ , and  $f_3 = \chi_C$ .

We consider pMRI reconstruction, where  $A = (A_1^T, A_2^T, \dots, A_N^T)^T$  for each  $A_j$  is composed of a diagonal downsampling operator  $D$ , the Fourier transform  $F$ , and a diagonal coil sensitivity mapping  $S_j$  for receiver  $j$ , i.e.  $A_j = DFS_j$  and  $S_j$  are often estimated in advance. It is well known in the total variation application that  $\lambda_{\max}(BB^T) = 8$ . The related Lipschitz constant of  $\nabla f_1$  can be estimated as  $\beta = 1$ . Therefore the two parameters in PDFP are set as  $\lambda = 1/8$  and  $\gamma = 2$ . The same simulation setting as in [15] is used in this experiment and we still use the artifact power (AP) and the two-region signal to noise ratio (SNR) to measure the image quality. We refer the reader to [15, 27] for more details.



In the following, we compare PDFFP algorithm (4.1) with the previously proposed algorithms PDFP<sup>2</sup>O (4.2) and PDFP<sup>2</sup>O<sub>C</sub> (4.3). From Figures 2 and 3, we can first see that the introduction of non-negative constraint in the model (5.2) is beneficial and we can recover a better solution with higher two-region SNR and lower AP value. The non-negative constraint leads to a faster convergence for a stable recovery. Second, PDFP<sup>2</sup>O<sub>C</sub> and PDFFP are both efficient. For a subsampling rate  $R = 2$ , PDFP<sup>2</sup>O<sub>C</sub> and PDFFP can both recover better solutions in terms of AP values compared to PDFP<sup>2</sup>O under the same iterative numbers. For  $R = 4$ , the solutions of PDFP<sup>2</sup>O<sub>C</sub> and PDFFP have better AP values than those of PDFP<sup>2</sup>O, but only use half iteration numbers of PDFP<sup>2</sup>O. The computation time for PDFFP is slightly less than PDFP<sup>2</sup>O<sub>C</sub>. Finally, the iterative solutions of PDFFP are always feasible, which could be useful in practice.

### 6 Conclusion

We have extended the algorithm PAPA [20] and PDFP<sup>2</sup>O [15] to derive a primal-dual fixed point algorithm PDFFP (see (1.3)) for solving the minimization problem of three-block convex separable functions (1.1). The proposed PDFFP algorithm is a symmetric and fully splitting scheme, only involving explicit gradient and linear operators without any inversion and subproblem solving, when the proximity operator of nonsmooth functions can easily be handled. The scheme can easily be adapted to a variety of inverse problems involving many terms minimization and it is suitable for large-scale parallel implementation. In addition, the parameter range determined by the convergence analysis is rather simple and clear, and it could be useful for practical applications. Finally, as discussed in Section 5 in [5], we can also extend the current PDFFP algorithm to solve multi-block composite (more than three) minimization problems. Preconditioning operators, as proposed in [16, 19, 24, 28], can also be introduced to accelerate the PDFFP, which could be a future work for some specific applications.

**Competing interests**

The authors declare that they have no competing interests.

**Authors' contributions**

All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

**Author details**

<sup>1</sup>Schools of Mathematical Sciences, and MOE-LSC, Shanghai Jiao Tong University, 800, Dongchuan Road, Shanghai, China. <sup>2</sup>School of Biomedical Engineering, Shanghai Jiao Tong University, 800, Dongchuan Road, Shanghai, China. <sup>3</sup>Department of Mathematics, Taiyuan University of Science and Technology, Taiyuan, China. <sup>4</sup>Institute of Natural Sciences, Shanghai Jiao Tong University, 800, Dongchuan Road, Shanghai, China.

**Acknowledgements**

P Chen was partially supported by the PhD research startup foundation of Taiyuan University of Science and Technology (No. 20132024). J Huang was partially supported by NSFC (No. 11571237). X Zhang was partially supported by NSFC (Nos. 91330102 and GZ1025) and 973 program (No. 2015CB856004). We thank the reviewer for pointing out the references [4, 14, 16] and for the pertinent comments and suggestions, which greatly improved the early version of this paper.

Received: 29 November 2015 Accepted: 12 April 2016 Published online: 26 April 2016

**References**

- Tibshirani, R, Saunders, M, Rosset, S, Zhu, J, Knight, K: Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **67**(1), 91-108 (2005)
- Yuan, M, Lin, Y: Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **68**(1), 49-67 (2006)
- Goldstein, T, Osher, S: The split Bregman method for L1-regularized problems. *SIAM J. Imaging Sci.* **2**(2), 323-343 (2009)
- Combettes, PL, Pesquet, J-C: Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal.* **20**(2), 307-330 (2012)
- Condat, L: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.* **158**(2), 460-479 (2013)
- Davis, D, Yin, W: A three-operator splitting scheme and its optimization applications (2015). arXiv:1504.01032
- Fortin, M, Glowinski, R: Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems. North-Holland, Amsterdam (1983)
- Boyd, S, Parikh, N, Chu, E, Peleato, B, Eckstein, J: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1-122 (2011)
- Zhu, M, Chan, T: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM report 08-34, UCLA (2008)
- Esser, E, Zhang, X, Chan, TF: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.* **3**(4), 1015-1046 (2010)
- Chambolle, A, Pock, T: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120-145 (2011)
- Pock, T, Chambolle, A: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: 2011 International Conference on Computer Vision (ICCV), pp. 1762-1769. IEEE Press, New York (2011)
- Zhang, X, Burger, M, Bresson, X, Osher, S: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM J. Imaging Sci.* **3**(3), 253-276 (2010)
- Loris, I, Verhoeven, C: On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty. *Inverse Probl.* **27**(12), 125007 (2011)
- Chen, P, Huang, J, Zhang, X: A primal-dual fixed point algorithm for convex separable minimization with applications to image restoration. *Inverse Probl.* **29**(2), 025011 (2013)
- Combettes, PL, Condat, L, Pesquet, J-C, Vu, BC: A forward-backward view of some primal-dual optimization methods in image recovery. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 4141-4145. IEEE Press, New York (2014)
- Komodakis, N, Pesquet, J-C: Playing with duality: an overview of recent primal-dual approaches for solving large-scale optimization problems (2014). arXiv:1406.5429
- Li, Q, Shen, L, Xu, Y, Zhang, N: Multi-step fixed-point proximity algorithms for solving a class of optimization problems arising from image processing. *Adv. Comput. Math.* **41**(2), 387-422 (2015)
- Li, Q, Zhang, N: Fast proximity-gradient algorithms for structured convex optimization problems. Preprint (2015)
- Krol, A, Li, S, Shen, L, Xu, Y: Preconditioned alternating projection algorithms for maximum a posteriori ECT reconstruction. *Inverse Probl.* **28**(11), 115005 (2012)
- Moreau, J-J: Fonctions convexes duales et points proximaux dans un espace Hilbertien. *C. R. Acad. Sci. Paris, Sér. A Math.* **255**, 2897-2899 (1962)
- Combettes, PL, Wajs, VR: Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **4**(4), 1168-1200 (2005)
- Chen, P, Huang, J, Zhang, X: A primal-dual fixed point algorithm based on proximity operator for convex set constrained separable problem. *J. Nanjing Norm. Univ. Nat. Sci. Ed.* **36**(3), 1-5 (2013) (in Chinese)
- Tang, Y-C, Zhu, C-X, Wen, M, Peng, J-G: A splitting primal-dual proximity algorithm for solving composite optimization problems (2015). arXiv:1507.08413
- Briceno-Arias, LM, Combettes, PL: A monotone+skew splitting model for composite monotone inclusions in duality. *SIAM J. Control Optim.* **21**(4), 1230-1250 (2011)
- Liu, J, Yuan, L, Ye, J: An efficient algorithm for a class of fused lasso problems. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 323-332. ACM, New York (2010)

27. Ji, JX, Son, JB, Rane, SD: PULSAR: a Matlab toolbox for parallel magnetic resonance imaging using array coils and multiple channel receivers. *Concepts Magn. Reson., Part B Magn. Reson. Eng.* **31**(1), 24-36 (2007)
28. Chen, P: Primal-dual fixed point algorithms for convex separable minimization and their applications. PhD thesis, Shanghai Jiao Tong University (2013) (in Chinese)

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](http://springeropen.com)

---