Two Modified Schemes for the Primal Dual Fixed Point Method

Ya-Nan Zhu[1] and Xiaoqun Zhang[2,*]

[1] *School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China.*
[2] *School of Mathematical Sciences, MOE-LSC, Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China.*

Abstract. The primal dual fixed point (PDFP) proposed in [7] was designed to solve convex composite optimization problems in imaging and data sciences. The algorithm was shown to have some advantages for simplicity and flexibility for divers applications. In this paper we study two modified schemes in order to accelerate its performance. The first one considered is an inertial variant of PDFP, namely inertial PDFP (iPDFP) and the second one is based on a prediction correction framework proposed in [20], namely Prediction Correction PDFP (PC-PDFP). Convergence analysis on both algorithms are provided. Numerical experiments on sparse signal recovery and CT image reconstruction using TV-L_2 model are present to demonstrate the acceleration of the two proposed algorithms compared to the original PDFP algorithm.

1 Introduction

We consider the optimization problem as follows:

$$\min_{x\in\mathbb{R}^d} f(x)+(g\circ B)(x), \tag{1.1}$$

where $g:\mathbb{R}^m\to\mathbb{R}\cup\{\infty\}$ is convex, lower semi-continuous (l.s.c) and may not be differentiable, $B:\mathbb{R}^d\to\mathbb{R}^m$ is a linear transform, and $f:\mathbb{R}^d\to\mathbb{R}\cup\{\infty\}$ is also convex l.s.c. with $\frac{1}{\beta}$ Lipschitz continuous gradient. This problem is widely considered in machine learning and signal/image processing. For instance, for linear regression $f(x)=\frac{1}{n}\sum_{i=1}^n(a_i^T x-b_i)^2$

Corresponding author. Email addresses: `zynsjtu@sjtu.edu.cn` (Y.-N. Zhu), `xqzhang@sjtu.edu.cn` (X. Zhang)

and for binary classification $f(x) = \frac{1}{n}\sum_{i=1}^{n}\log(1+\exp(-b_i a_i^T x))$ where b_i denotes the label of the i-th sample a_i. The second term $(g \circ B)(x)$, usually the so-called regularization term, which can be taken as $g(\cdot) := \|x\|_1, \|x\|_2^2$ and $B = I$ (I is identity). In some cases, B is chosen as inverse covariance matrix for obtaining some graph guided sparsity [3]. In signal and image processing, a classic example is total variation (TV) based image restoration model:

$$\min_{x \in \mathbb{R}^d} \|\mathcal{A}x - f\|_2^2 + g(\nabla x), \tag{1.2}$$

where $\mathcal{A}$ is some linear operator, for example convolution operator for image deconvolution and Radon transform for CT reconstruction, ∇ denotes the gradient operator and $g(\cdot)$ is the function that ensures sparsity.

Many algorithms are proposed to solve the problem (1.1). If $B = I$ one of the most popular methods is the proximal gradient method (PGM) (also known as proximal forward backward splitting (PFBS) [10]) and its acceleration variants [4,23]. The PGM proceeds as follows:

$$x_{k+1} = \text{Prox}_{\gamma g}(x_k - \gamma \nabla f(x_k)), \tag{1.3}$$

where $\gamma > 0$ is a parameter and the operator $\text{Prox}_{\gamma g}(\cdot)$ is defined as

$$\text{Prox}_{\gamma g}(\cdot) = \arg\min_{y \in \mathbb{R}^r} g(y) + \frac{1}{2\gamma}\|y - \cdot\|_2^2. \tag{1.4}$$

When $B \neq I$, PGM needs to handle the term $\text{Prox}_{g \circ B}(\cdot)$ which may be as difficult as the original problem. To overcome this difficulty, many algorithms based on augmented Lagrangian and Fenchel duality were designed, such as the split Bregman method [16,25] (a.k.a the alternating direction of multipliers method (ADMM) [12,19]), the primal dual hybrid gradient method (PDHG) [14,29] (also known as Chambolle-Pock algorithm [6]), Condat-Vu [11,27] algorithm, the fixed-point method based on proximity operator (FP^{2}O) [22] and the primal dual fixed point method (PDFP) [7]. In this paper we focus on PDFP as it can maximally decouple subproblems and it was shown to be effective with parallel implementation for many large scale imaging and data sciences problems [7–9]. The scheme of PDFP for solving (1.1) is given as follows:

Algorithm: Primal dual fixed point method

Step 1: set $x_1 \in \mathbb{R}^d, p_1 \in \mathbb{R}^m$ and $\gamma > 0, \lambda > 0$
Step 2: for $k = 1, 2, \cdots$

$$x_{k+\frac{1}{2}} = x_k - \gamma \nabla f(x_k)$$
$$p_{k+1} = \left(I - \text{Prox}_{\frac{\gamma}{\lambda}g}\right)\left(Bx_{k+\frac{1}{2}} + (I - \lambda BB^T)p_k\right)$$
$$x_{k+1} = x_{k+\frac{1}{2}} - \lambda B^T p_{k+1}$$

until the stop criterion is satisfied.

The method PDFP only involves the forward gradient of f, matrix vector multiplication and the proximal operation of g, which is generally easy to be solved. The convergence analysis of PDFP was established in [7] based on a fixed point iteration. Later its three-block and multi-block extensions were studied in [8, 9]. Recently it was extended to a stochastic setting for solving machine learning and image reconstruction problems in [30].

In this paper we reformulate the PDFP from the variational inequality framework and consider two types of acceleration schemes: one is based on the inertial variants [5, 21] and the other is from the prediction correction perspective [17, 18, 20]. Generally speaking, the idea of inertial variants utilizes a set of auxiliary variables based on two consecutive iterates of the main sequences for the acceleration. To be more specific, let (x_k, p_k) be the primal dual sequence of PDFP, two extra variables (z_k, v_k) are computed as follows:

$$\begin{cases} z_k = x_k + \theta_k(x_k - x_{k-1}), \\ v_k = p_k + \theta_k M(p_k - p_{k-1}), \end{cases} \tag{1.5}$$

where M is a positive definite matrix to be defined and θ_k is an inertial factor. The new iterate (x_{k+1}, p_{k+1}) is obtained by performing PDFP on (z_k, v_k). In [28], a different inertial step was applied to PDFP, which leads to a different convergence analysis. As PDFP recovers PGM when $B = I, \lambda = 1$, we will show that our proposed inertial PDFP can also reduce to the usual forms of inertial PGMs proposed in [2, 4, 23], while the method in [28] does not.

The second modified scheme is based on the prediction-correction type method proposed in [18, 20] to improve the performance of the original algorithm. The basic idea is to first generate a prediction sequence $(\tilde{x}_k, \tilde{p}_k)$ by first performing the basic algorithm on the current estimate (x_k, p_k). Then in the correction step the new iterate (x_{k+1}, p_{k+1}) is updated based on a weighted combination of two directions $d_1(x_k, \tilde{x}_k), d_2(p_k, \tilde{p}_k)$, that are computed based on the current optimality gap and the predefined estimate. The proposed two algorithms are called inertial PDFP (iPDFP) and prediction-correction PDFP (PC-PDFP) respectively. The convergence of the two algorithms are established. Numerically, through the examples on sparse signal and CT reconstruction, we can obverse that the two algorithms accelerate the original PDFP, both in terms of iterations number and computation time.

The paper is organized as following: In Section 2, we present Algorithm 1: inertial PDFP and Algorithm 2: prediction-correction PDFP. Convergence analysis is present in Section 3 and more details can be found in Appendix. Finally, a synthetic example for sparse signal recovery and CT reconstruction using TV-L_2 model are performed to demonstrate the performance of the two algorithms and the comparisons to the original PDFP algorithm.

2 Algorithms

In this section we present the details of the algorithms iPDFP and PC-PDFP. We first give some useful definitions.

Definition 2.1. *The conjugate function of $g(\cdot)$ at v is defined by*

$$g^*(v) = \sup_{y \in dom(g)} \{v^T y - g(y)\}, \tag{2.1}$$

where $v \in dom(g^*(\cdot)) = \{v | g^*(v) < \infty\} = V$.

Definition 2.2. *The following definitions are used throughout the paper:*

- *$\rho_{\min}(\cdot)$, $\rho_{\max}(\cdot)$ denote the minimum and maximum eigenvalues of a given matrix.*

- *Define $M = I - \lambda BB^T$, where $0 < \lambda < \frac{1}{\rho_{\max}(BB^T)}$ is a constant such that M is positive definite.*

Lemma 2.1. *The saddle point problem of (1.1) is*

$$\min_{x \in \mathbb{R}^d} \max_{p \in V} f(x) + \langle Bx, p \rangle - g^*(p), \tag{2.2}$$

where $g^*(\cdot)$ is the conjugate of g. The variational inequality reformulation of (2.2) is

$$\begin{cases} (x - x^*)^T (\nabla f(x^*) + B^T p^*) \geq 0, & \forall x \in \mathbb{R}^d, \\ (p - p^*)^T (s^* - Bx^*) \geq 0, & \forall p \in V, \ s^* \in \partial g^*(p^*), \end{cases} \tag{2.3}$$

where $(x^*, p^*) \in \mathbb{R}^d \times V$ is an optimal primal-dual solution of (2.2).

Remark 2.1. If $g(\cdot)$ is l_1 norm, the conjugate $g^*(\cdot) = \iota_V(\cdot)$ where ι_V denotes the indicator function of set $V = \{p \in \mathbb{R}^m : \|p\|_{\inf} \leq 1\}$ and $\|\cdot\|_{\inf}$ is the maximum norm. Then the second inequality of (2.3) becomes

$$(p - p^*)^T (-Bx^*) \geq 0, \quad \forall p \in V. \tag{2.4}$$

In the following, we denote $\Omega = \mathbb{R}^d \times V$, $u = (x, p) \in \Omega$, $u^* = (x^*, p^*) \in \Omega^*$, where Ω^* is the optimal primal-dual solution set. Let

$$F(x, p) = \begin{pmatrix} \nabla f(x) + B^T p \\ s - Bx \end{pmatrix}, \quad s \in \partial g^*(p). \tag{2.5}$$

Then (2.3) can be rewritten as

$$(u - u^*)^T F(u^*) \geq 0, \quad \forall u \in \Omega. \tag{2.6}$$

We also denote the iterates in both algorithms as $u_k = (x_k, p_k)$ and $\tilde{u}_k = (\tilde{x}_k, \tilde{p}_k)$. Now we give a reformulation of PDFP by introducing an intermediate variable y_{k+1} as

$$\begin{cases} y_{k+1} = x_k - \gamma \nabla f(x_k) - \lambda B^T p_k, \\ p_{k+1} = (I - \text{Prox}_{\frac{\gamma}{\lambda}g})(By_{k+1} + p_k), \\ x_{k+1} = x_k - \gamma \nabla f(x_k) - \lambda B^T p_{k+1}. \end{cases} \tag{2.7}$$

According to Moreau decomposition [10], for $p \in \mathbb{R}^m$, one has

$$p = p_{\frac{\gamma}{\lambda}}^{\oplus} + p_{\frac{\gamma}{\lambda}}^{\ominus}, \tag{2.8}$$

where $p_{\frac{\gamma}{\lambda}}^{\oplus} = \text{Prox}_{\frac{\gamma}{\lambda}g}(p)$ and $p_{\frac{\gamma}{\lambda}}^{\ominus} = \frac{\gamma}{\lambda}\text{Prox}_{\frac{\lambda}{\gamma}g^*}(\frac{\lambda}{\gamma}p)$. Then the second equality of (2.7) can be rewritten as

$$p_{k+1} = (I - \text{Prox}_{\frac{\gamma}{\lambda}g})(By_{k+1} + p_k) = \frac{\gamma}{\lambda}\text{Prox}_{\frac{\lambda}{\gamma}g^*}\left(\frac{\lambda}{\gamma}By_{k+1} + \frac{\lambda}{\gamma}p_k\right). \tag{2.9}$$

Let $p_k := \frac{\lambda}{\gamma}p_k$, then (2.7) can be reformulated as

$$\begin{cases} y_{k+1} = x_k - \gamma \nabla f(x_k) - \gamma B^T p_k, \\ p_{k+1} = \text{Prox}_{\frac{\lambda}{\gamma}g^*}\left(\frac{\lambda}{\gamma}By_{k+1} + p_k\right), \\ x_{k+1} = x_k - \gamma \nabla f(x_k) - \gamma B^T p_{k+1}. \end{cases} \tag{2.10}$$

Based on this formulation, we present the inertial PDFP by using an extrapolated primal-dual pair (z_k, v_k) before performing PDFP steps, similar to [21]. The algorithm is described as follows:

Algorithm 1: Inertial Primal Dual Fixed Point Method (iPDFP)

Step 1: set $x_1 \in \mathbb{R}^d$, $p_1 \in \mathbb{R}^m$ and choose proper $\gamma, \lambda, 0 < \theta_k < 1$.

Step 2: for $k = 1, 2, \cdots$

$$z_k = x_k + \theta_k(x_k - x_{k-1})$$

$$v_k = p_k + \theta_k M(p_k - p_{k-1})$$

$$y_{k+1} = z_k - \gamma \nabla f(z_k) - \gamma B^T p_k$$

$$p_{k+1} = \text{Prox}_{\frac{\lambda}{\gamma}g^*}(\frac{\lambda}{\gamma}By_{k+1} + v_k)$$

$$x_{k+1} = z_k - \gamma \nabla f(z_k) - \gamma B^T p_{k+1}$$

until the stop criterion is satisfied.

We note that the proposed inertial step on the dual is different from the usual one, as there is a matrix M inside, where $M = I - \lambda BB^T$ as defined above. If $B = I$, $\lambda = 1$ iPDFP becomes inertial PGMs [2, 4, 23].

The PDFP algorithm can be also formulated as an approximate proximal point algorithm (APPA) [13, 26]. In [17, 18], a prediction correction framework was proposed to improve the performance of APPA. Following this line, we consider such prediction correction scheme on PDFP. In the following, we first present the framework for solving (2.6).

- **Step 1 (prediction):** Derive a prediction $\tilde{u}_k \in \Omega$, two directions $d_1(u_k, \tilde{u}_k)$, $d_2(u_k, \tilde{u}_k)$ and a function $\varphi(u_k, \tilde{u}_k)$ such that the following four conditions hold:

    - (1) $(u - \tilde{u}_k)^T (d_2(u_k, \tilde{u}_k) - d_1(u_k, \tilde{u}_k)) \geq 0$, $\forall u \in \Omega$.
    - (2) There exists $\kappa > 0$ such that
      $$\|d_1(u_k, \tilde{u}_k)\| \leq \kappa \|u_k - \tilde{u}_k\|.$$
    - (3) The function $\varphi(u_k, \tilde{u}_k)$ satisfy
      $$\varphi(u_k, \tilde{u}_k) \geq \delta \|u_k - \tilde{u}_k\|^2,$$
      for some $\delta > 0$ and $\varphi(u_k, \tilde{u}_k) = 0 \Leftrightarrow u_k = \tilde{u}_k$.
    - (4) For $u^* \in \Omega^*$, the following inequality holds:
      $$(\tilde{u}_k - u^*)^T d_2(u_k, \tilde{u}_k) \geq \varphi(u_k, \tilde{u}_k) - (u_k - \tilde{u}_k)^T d_1(u_k, \tilde{u}_k).$$

  The condition (1), (3), (4) are used to such that the two directions $d_1(u_k, \tilde{u}_k)$, $d_2(u_k, \tilde{u}_k)$ are descent directions with respect to the distant function $\|u - u^*\|^2$ at point u_k (see [17, Lemmas 2.1, 2.2]). The condition (2) is further used for the convergence. Based on the two descent directions d_1, d_2 in step I, the following contraction method [18] was proposed to improve the performance as a correction step.

- **Step 2 (correction):** The new update u_{k+1} is given by
  $$u_{k+1} = P_\Omega(u_k - \eta \alpha_k d(u_k, \tilde{u}_k)), \tag{2.11}$$
  where P_Ω is the projection onto the set Ω, $d(u_k, \tilde{u}_k) = (1-t)d_1(u_k, \tilde{u}_k) + t d_2(u_k, \tilde{u}_k)$, for some $t \in (0,1]$, $\eta \in [1,2)$ and the step size
  $$\alpha_k = \frac{\langle u_k - \tilde{u}_k, d_1(u_k, \tilde{u}_k) \rangle}{\|d_1(u_k, \tilde{u}_k)\|_2^2}. \tag{2.12}$$

Now in the case of PDFP, we propose to define $\tilde{u}_k$, $d_1(u_k, \tilde{u}_k)$, $d_2(u_k, \tilde{u}_k)$ and $\varphi(u_k, \tilde{u}_k)$ as follows:

– The prediction $\tilde{u}_k$ is generated by performing PDFP steps.
– The two directions $d_1(u_k,\tilde{u}_k), d_2(u_k,\tilde{u}_k)$ are defined as:

$$d_1(u_k,\tilde{u}_k) = \begin{pmatrix} \nabla f(\tilde{x}_k) - \nabla f(x_k) - \frac{1}{\gamma}(\tilde{x}_k - x_k) \\ G(p_k - \tilde{p}_k) \end{pmatrix}, \tag{2.13}$$

where $G = \frac{\gamma}{\lambda}M$ and

$$d_2(u_k,\tilde{u}_k) := F(\tilde{u}_k) = \begin{pmatrix} \nabla f(\tilde{x}_k) + B^T\tilde{p}_k \\ s_k - B\tilde{x}_k \end{pmatrix}, \qquad s_k \in \partial g^*(\tilde{p}_k). \tag{2.14}$$

– The function $\varphi(u_k,\tilde{u}_k)$ is defined as

$$\varphi(u_k,\tilde{u}_k) = (u_k - \tilde{u}_k)^T d_1(u_k,\tilde{u}_k). \tag{2.15}$$

We will show that $\tilde{u}_k$, $d_1(u_k,\tilde{u}_k)$, $d_2(u_k,\tilde{u}_k)$ and $\varphi(u_k,\tilde{u}_k)$ as defined above satisfy the four conditions in step I, thus the correction and the related convergence analysis proposed in [18] can be applied. To be more specific, the proposed prediction-correction PDFP is summarized as follows:

Algorithm 2: Prediction Correction Primal Dual Fixed Point Method

---

Step 1: Set $x_1 \in \mathbb{R}^d$, $p_1 \in \mathbb{R}^m$ and $u_1 = (x_1,p_1)$. Choose $\eta \in [1,2)$, $\lambda, \gamma > 0$, $t \in [0,1]$. and let $G = \frac{\gamma}{\lambda}M$.

Step 2: For $k = 1,2,\cdots$

• Prediction:

$$\tilde{y}_k = x_k - \gamma\nabla f(x_k) - \gamma B^T p_k$$
$$\tilde{p}_k = \text{Prox}_{\frac{\lambda}{\gamma}g^*}(\frac{\lambda}{\gamma}B\tilde{y}_k + p_k)$$
$$\tilde{x}_k = x_k - \gamma\nabla f(x_k) - \gamma B^T\tilde{p}_k$$

Let $\tilde{u}_k = (\tilde{x}_k,\tilde{p}_k)$

• Compute the two directions:

$$d_1(u_k,\tilde{u}_k) = \begin{pmatrix} \nabla f(\tilde{x}_k) - \nabla f(x_k) - \frac{1}{\gamma}(\tilde{x}_k - x_k) \\ G(p_k - \tilde{p}_k) \end{pmatrix}$$

$$d_2(u_k,\tilde{u}_k) = \begin{pmatrix} \nabla f(\tilde{x}_k) + B^T\tilde{p}_k \\ s_k - B\tilde{x}_k \end{pmatrix}, \quad s_k \in \partial g^*(\tilde{p}_k)$$

• Correction:

$$u_{k+1} = \begin{pmatrix} x_{k+1} \\ p_{k+1} \end{pmatrix} = P_\Omega(u_k - \eta\alpha_k((1-t)d_1(u_k,\tilde{u}_k) + td_2(u_k,\tilde{u}_k)))$$

where $\alpha_k = \frac{\langle u_k - \tilde{u}_k, d_1(u_k,\tilde{u}_k)\rangle}{\|d_1(u_k,\tilde{u}_k)\|_2^2}$.

until the stop criterion is satisfied.

3 Convergence analysis

3.1 Convergence of iPDFP

The following theorem states the convergence of the sequence generated by iPDFP algorithm.

Theorem 3.1. *Suppose $f(x)$ has $\frac{1}{\beta}$ Lipschitz continuous gradient. Then if the parameters satisfy*

- $0 < \lambda < \frac{1}{\rho_{\max}(BB^T)}$ *and* $0 < \gamma < 2\beta$;

- θ_k *is non-decreasing and* $0 \leq \theta_k < \min\left\{\frac{-(2c+1)+\sqrt{8c+1}}{2(1-c)}, \frac{1}{3}\right\}$ *for* $c = \frac{2\beta-\gamma}{2\beta}$,

then the sequence $u_k = (x_k, p_k)$ generated by iPDFP converges to a solution $u^* = (x^*, p^*)$.

Before we prove this theorem, we first define a normed space for the primal dual iterates and present some necessary lemmas.

Definition 3.1. *For two vectors $u_1 = (x_1, p_1)$, $u_2 = (x_2, p_2)$ in $\mathcal{H} = \mathbb{R}^{d+m}$, define the inner product as followed:*

$$\langle u_1, u_2 \rangle_{\mathcal{H}} = \langle x_1, x_2 \rangle + \langle p_1, Rp_2 \rangle, \tag{3.1}$$

where $R = \frac{\gamma^2}{\lambda}M$ and $\langle \cdot, \cdot \rangle$ denote the usual inner product in Euclidean space. It can be seen that $\mathcal{H}$ is a Hilbert space and the induced norm is

$$\|u\|_{\mathcal{H}}^2 = \|x\|_2^2 + \|p\|_R^2, \tag{3.2}$$

where $\|p\|_R^2 = \langle p, Rp \rangle$.

Lemma 3.1. [1] *Let $(\varphi_k)_{k \in \mathbb{N}}$, $(\delta_k)_{k \in \mathbb{N}}$ and $(\alpha_k)_{k \in \mathbb{N}}$ be the sequences in $[0, +\infty)$ such that $\varphi_{k+1} \leq \varphi_k + \alpha_k(\varphi_k - \varphi_{k-1}) + \delta_k$ for all $k \geq 1$, $\sum_{k \in \mathbb{N}} \delta_k < \infty$ and there exists a real number α with $0 \leq \alpha_k \leq \alpha < 1$ for all $k \in \mathbb{N}$. Then the following holds:*

- $\sum_{n \geq 1}[\varphi_k - \varphi_{k-1}]_+ < \infty$, *where* $[a]_+ = \max\{t, 0\}$;

- *there exists $\varphi^* \in [0, \infty)$ such that $\lim_{k \to \infty} \varphi_k = \varphi^*$.*

Lemma 3.2. (Opial [24]) *Let C be a nonempty set of $\mathcal{H}$ and $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$ such that the following two conditions hold:*

- *for every $x \in C$, $\lim_{n \to +\infty} \|x_n - x\|$ exists;*

- *every sequential weak cluster point of $(x_n)_{n \in \mathbb{N}}$ is in C.*

Then the sequence $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in C.

Lemma 3.3. *Suppose $\overline{x} \in \mathbb{R}^d$ and $\overline{p} \in V$ satisfy*

$$\begin{cases} \overline{p} = \text{Prox}_{\frac{\lambda}{\gamma} g^*} \left(\frac{\lambda}{\gamma} B\overline{x} + \overline{p} \right), \\ \overline{x} = \overline{x} - \gamma \nabla f(\overline{x}) - \gamma B^T \overline{p}, \end{cases} \tag{3.3}$$

then $(\overline{x}, \overline{p}) \in \Omega^*$.

Lemma 3.4. *The following inequality holds:*

$$\gamma(B(x_{k+1} - x^*))^T (p_{k+1} - p^*) + E_x^{k+1} \le \theta_k E_x^k + \frac{\gamma - 2\beta}{4\beta} \|x_{k+1} - z_k\|_2^2 + \frac{\theta_k + \theta_k^2}{2} \|x_k - x_{k-1}\|_2^2, \tag{3.4}$$

where $E_x^k = \frac{1}{2} \|x_k - x^*\|_2^2 - \frac{1}{2} \|x_{k-1} - x^*\|_2^2$.

Lemma 3.5. *The following inequality holds:*

$$-\gamma(p_{k+1} - p^*)^T (B(x_{k+1} - x^*)) + E_p^{k+1} \le \theta_k E_p^k + \frac{\theta_k + \theta_k^2}{2} \|p_k - p_{k-1}\|_R^2, \tag{3.5}$$

where $E_p^k = \frac{1}{2} \|p_k - p^*\|_R^2 - \frac{1}{2} \|p_{k-1} - p^*\|_R^2$.

Lemma 3.6. *Let $c = \frac{2\beta - \gamma}{2\beta}$, choose $0 < \gamma < 2\beta$, $0 < \lambda < \frac{1}{\rho_{\max}(BB^T)}$ and a nondecreasing sequence θ_k such that*

$$\begin{aligned} c(1 - \theta_k)^2 - \theta_k(1 + \theta_k) &\ge \varepsilon, \\ 1 - 3\theta_k &\ge \varepsilon, \end{aligned} \tag{3.6}$$

for some $\varepsilon > 0$. Then

$$\sum_{k=1}^{\infty} \|u_k - u_{k-1}\|_{\mathcal{H}}^2 < \infty. \tag{3.7}$$

    The above four lemmas are essential for the proof of Theorem 3.1, for their proofs can be found in Appendix. In the following, we present the proof for Theorem 3.1.

Proof of Theorem 3.1. The proof of Theorem 3.1 can be divided into three steps.

STEP I: By the choice of θ_k, it can be verified that the condition in Eq. (3.6) in Lemma 3.6 is satisfied, thus $\|u_k - u_{k-1}\|_{\mathcal{H}}^2 \to 0$ and

$$\sum_{k=1}^{\infty} \theta_k \|u_k - u_{k-1}\|_{\mathcal{H}}^2 < \sum_{k=1}^{\infty} \|u_k - u_{k-1}\|_{\mathcal{H}}^2 < \infty. \tag{3.8}$$

Let $E_k = E_x^k + E_p^k = \frac{1}{2} \|u_k - u^*\|_{\mathcal{H}}^2 - \frac{1}{2} \|u_{k-1} - u^*\|_{\mathcal{H}}^2$. By adding Eq. (3.4) and Eq. (3.5), we obtain

$$\begin{aligned} E_{k+1} &\le \theta_k E_k + \frac{\gamma - 2\beta}{4\beta} \|x_{k+1} - z_k\|_2^2 + \frac{\theta_k + \theta_k^2}{2} \left(\|x_k - x_{k-1}\|_2^2 + \|p_k - p_{k-1}\|_R^2 \right) \\ &= \theta_k E_k + \frac{\gamma - 2\beta}{4\beta} \|x_{k+1} - z_k\|_2^2 + \frac{\theta_k + \theta_k^2}{2} \|u_k - u_{k-1}\|_{\mathcal{H}}^2. \end{aligned} \tag{3.9}$$

Let $\Delta_k = \theta_k \|u_k - u_{k-1}\|_{\mathcal{H}}^2$. Using the fact that $0 < \gamma < 2\beta$, $0 \le \theta_k < 1$ and $\frac{\theta_k + \theta_k^2}{2} \le \theta_k$, we have

$$E_{k+1} \le \theta_k E_k + \frac{\theta_k + \theta_k^2}{2} \|u_k - u_{k-1}\|_{\mathcal{H}}^2 \le \theta_k E_k + \Delta_k. \tag{3.10}$$

According to Eq. (3.8), we can see that the second term of the right hand side of (3.10) is summable. Let $\varphi_k = \frac{1}{2} \|u_k - u^*\|_{\mathcal{H}}^2$, $\alpha_k = \theta_k$, $\delta_k = \Delta_k$ in Lemma 3.1, it follows that $\lim_{k \to \infty} \|u_k - u^*\|_{\mathcal{H}}^2$ exists.

STEP II: Suppose there exists a subsequence $u_{i_k} = (x_{i_k}, p_{i_k}) \to (\overline{x}, \overline{p}) = \overline{u}$. Since $\|u_k - u_{k-1}\|_{\mathcal{H}}^2 \to 0$ we have $x_{i_k} - x_{i_{k-1}} \to 0$, $p_{i_k} - p_{i_{k-1}} \to 0$ and

$$\begin{aligned} z_{i_k} &= x_{i_k} + \theta_k(x_{i_k} - x_{i_{k-1}}) \to \overline{x}, \\ v_{i_k} &= p_{i_k} + \theta_k M(p_{i_k} - p_{i_{k-1}}) \to \overline{p}. \end{aligned} \tag{3.11}$$

For p_{i_k+1}, x_{i_k+1}, we have

$$\begin{aligned} \|p_{i_k+1} - \overline{p}\| &\le \|p_{i_k+1} - p_{i_k}\| + \|p_{i_k} - \overline{p}\| \to 0, \\ \|x_{i_k+1} - \overline{x}\| &\le \|x_{i_k+1} - x_{i_k}\| + \|x_{i_k} - \overline{x}\| \to 0. \end{aligned} \tag{3.12}$$

Thus $p_{i_k+1} \to \overline{p}, x_{i_k+1} \to \overline{x}$.

Then by the update of Algorithm 1, we have

$$\begin{aligned} x_{i_k+1} &= z_{i_k} - \gamma \nabla f(z_{i_k}) - \gamma B^T p_{i_k+1} \\ &\to \overline{x} = \overline{x} - \gamma \nabla f(\overline{x}) - \gamma B^T \overline{p}. \end{aligned} \tag{3.13}$$

Further the update of y_{i_k+1} leads to $y_{i_k+1} \to \overline{x}$. And

$$\begin{aligned} p_{i_k+1} &= \mathrm{Prox}_{\frac{\lambda}{\gamma} g^*} \left(\frac{\lambda}{\gamma} B y_{i_k+1} + v_{i_k} \right) \\ &\to \overline{p} = \mathrm{Prox}_{\frac{\lambda}{\gamma} g^*} \left(\frac{\lambda}{\gamma} B \overline{x} + \overline{p} \right). \end{aligned} \tag{3.14}$$

We then have

$$\begin{cases} \overline{p} = \mathrm{Prox}_{\frac{\lambda}{\gamma} g^*} \left(\frac{\lambda}{\gamma} B \overline{x} + \overline{p} \right), \\ \overline{x} = \overline{x} - \gamma \nabla f(\overline{x}) - \gamma B^T \overline{p}. \end{cases} \tag{3.15}$$

It follows that $(\overline{x}, \overline{p})$ is optimal by using Lemma 3.3.

STEP III: Let $C = \Omega^*$ in Lemma 3.2 and combining the results in Step I and Step II, we get the convergence of iPDFP. $\qquad \square$

3.2 Convergence of PC-PDFP

The following lemma verifies that the chosen $\tilde{u}_k$, $d_1(u_k,\tilde{u}_k)$, $d_2(u_k,\tilde{u}_k)$ and $\varphi(u_k,\tilde{u}_k)$ in Algorithm 2 satisfy the four conditions described above.

Lemma 3.7. *Recall the definition of $\tilde{u}_k$, $d_1(u_k,\tilde{u}_k)$, $d_2(u_k,\tilde{u}_k)$ and $\varphi(u_k,\tilde{u}_k)$. Choose $0<\gamma<\beta$ and $0<\lambda<\frac{1}{\rho_{\max}(BB^T)}$, then the following results hold:*

- $(u-\tilde{u}_k)^T(d_2(u_k,\tilde{u}_k)-d_1(u_k,\tilde{u}_k))\geq 0$, $\forall u\in\Omega$.

- $\|d_1(u_k,\tilde{u}_k)\|_2^2\leq\kappa\|u_k-\tilde{u}_k\|_2^2$, *for* $\kappa=\max\left\{\frac{2\beta^2+2\gamma^2}{\beta^2\gamma^2},\frac{\gamma}{\lambda}-\gamma\rho_{\min}(BB^T)\right\}$.

- $(\tilde{u}_k-u^*)^T d_2(u_k,\tilde{u}_k)\geq\varphi(u_k,\tilde{u}_k)-(u_k-\tilde{u}_k)^T d_1(u_k,\tilde{u}_k)$.

- $\varphi(u_k,\tilde{u}_k)\geq\delta\|u_k-\tilde{u}_k\|_2^2$, *where* $\delta=\min\left\{\frac{\beta-\gamma}{\gamma\beta},\frac{\gamma}{\lambda}-\gamma\rho_{\max}(BB^T)\right\}$.

Proof. By the optimality condition of the second equation of (2.9), we have

$$(p-\tilde{p}_k)^T\left(\frac{\lambda}{\gamma}\tilde{s}_k+\tilde{p}_k-p_k-\frac{\lambda}{\gamma}B\tilde{y}_k\right)\geq 0,\quad\forall\,\tilde{s}_k\in\partial g^*(p_k),\quad p\in V^*$$

$$\Leftrightarrow (p-\tilde{p}_k)^T\left(\frac{\lambda}{\gamma}\tilde{s}_k+\tilde{p}_k-p_k-\frac{\lambda}{\gamma}B(x_k-\gamma\nabla f(x_k)-\gamma B^T p_k)\right)\geq 0$$

$$\Leftrightarrow (p-\tilde{p}_k)^T\left(\tilde{s}_k-B\tilde{x}_k+G(\tilde{p}_k-p_k)\right)\geq 0,\tag{3.16}$$

where $G=\frac{\gamma}{\lambda}M$. Then the update of variable x and p in prediction step can be rewritten as

$$(x-\tilde{x}_k)^T\left(\nabla f(\tilde{x}_k)+B^T\tilde{p}_k-(\nabla f(\tilde{x}_k)-\nabla f(x_k)-\frac{1}{\gamma}(\tilde{x}_k-x_k))\right)\geq 0,\quad\forall x\in\mathbb{R}^d,$$
$$(p-\tilde{p}_k)^T\left(\tilde{s}_k-B\tilde{x}_k+G(\tilde{p}_k-p_k)\right)\geq 0,\quad\forall p\in V.\tag{3.17}$$

(1): The first argument is given as follows: recall

$$d_1(u_k,\tilde{u}_k)=\begin{pmatrix}\nabla f(\tilde{x}_k)-\nabla f(x_k)-\frac{1}{\gamma}(\tilde{x}_k-x_k)\\ G(p_k-\tilde{p}_k)\end{pmatrix},\tag{3.18}$$

$$d_2(u_k,\tilde{u}_k)=F(\tilde{u}_k)=\begin{pmatrix}\nabla f(\tilde{x}_k)+B^T\tilde{p}_k\\ s_k-B\tilde{x}_k\end{pmatrix},\tag{3.19}$$

then the inequality (3.17) can be rewritten as

$$(u-\tilde{u}_k)^T(d_2(u_k,\tilde{u}_k)-d_1(u_k,\tilde{u}_k))\geq 0,\quad\forall u\in\Omega.\tag{3.20}$$

(2): The second argument is given by the following estimate:

$$\|d_1(u_k,\tilde{u}_k)\|_2^2 = \|\nabla f(\tilde{x}_k) - \nabla f(x_k) - \frac{1}{\gamma}(\tilde{x}_k - x_k)\|_2^2 + \|G(p_k - \tilde{p}_k)\|_2^2$$

$$\leq 2\|\nabla f(\tilde{x}_k) - \nabla f(x_k)\|_2^2 + \frac{2}{\gamma^2}\|\tilde{x}_k - x_k\|_2^2 + \left(\frac{\gamma}{\lambda} - \gamma\rho_{\min}(BB^T)\right)\|p_k - \tilde{p}_k\|_2^2$$

$$\leq \frac{2}{\beta^2}\|\tilde{x}_k - x_k\|_2^2 + \frac{2}{\gamma^2}\|\tilde{x}_k - x_k\|_2^2 + \left(\frac{\gamma}{\lambda} - \gamma\rho_{\min}(BB^T)\right)\|p_k - \tilde{p}_k\|_2^2$$

$$\leq \max\left\{\frac{2\beta^2 + 2\gamma^2}{\beta^2\gamma^2}, \frac{\gamma}{\lambda} - \gamma\rho_{\min}(BB^T)\right\}\|u_k - \tilde{u}_k\|_2^2$$

$$= \kappa\|u_k - \tilde{u}_k\|_2^2. \tag{3.21}$$

(3): For the third inequality: Let $\varphi(u_k,\tilde{u}_k) = (u_k - \tilde{u}_k)^T d_1(u_k,\tilde{u}_k)$, we then have

$$(\tilde{u}_k - u^*)^T d_2(u_k,\tilde{u}_k)$$
$$= (\tilde{u}_k - u^*)^T F(\tilde{u}_k)$$
$$\geq (\tilde{u}_k - u^*)^T F(u^*)$$
$$\geq 0 = \varphi(u_k,\tilde{u}_k) - (u_k - \tilde{u}_k)^T d_1(u_k,\tilde{u}_k). \tag{3.22}$$

(4): The final estimate is given as follows:

$$\varphi(u_k,\tilde{u}_k) = (u_k - \tilde{u}_k)^T d_1(u_k,\tilde{u}_k)$$

$$= (x_k - \tilde{x}_k)^T\left(\nabla f(\tilde{x}_k) - \nabla f(x_k) - \frac{1}{\gamma}(\tilde{x}_k - x_k)\right)$$

$$\quad + (p_k - \tilde{p}_k)^T G(p_k - \tilde{p}_k)$$

$$\geq \frac{1}{\gamma}\|\tilde{x}_k - x_k\|_2^2 - \langle\nabla f(\tilde{x}_k) - \nabla f(x_k), \tilde{x}_k - x_k\rangle$$

$$\quad + \left(\frac{\gamma}{\lambda} - \gamma\rho_{\max}(BB^T)\right)\|\tilde{p}_k - p_k\|_2^2$$

$$\geq \min\left\{\frac{\beta - \gamma}{\gamma\beta}, \frac{\gamma}{\lambda} - \gamma\rho_{\max}(BB^T)\right\}\|u_k - \tilde{u}_k\|_2^2$$

$$= \delta\|u_k - \tilde{u}_k\|_2^2. \tag{3.23}$$

The condition $0 < \gamma < \beta$ yields $\delta > 0$. This completes the proof. $\quad\square$

Theorem 3.2. *Choose $0 < \gamma \leq \beta$ and $0 < \lambda < \frac{1}{\rho_{\max}(BB^T)}$, $1 \leq \eta < 2$, $u^* = (x^*,p^*) \in \Omega^*$, then the sequence generated by PC-PDFP is Féjer monotone, i.e.*

$$\|u_{k+1} - u^*\|_2^2 \leq \|u_k - u^*\|_2^2 - \eta(2 - \eta)\frac{\delta^2}{\kappa^2}\|u_k - \tilde{u}_k\|_2^2, \tag{3.24}$$

and u_k converge to the optimal solution u^*.

Proof. The Féjer monotonicity follows by using Lemma 3.7 and results in [17,18]. Since u_k is Féjer monotone and it is bounded, thus there exits a subsequence u_{i_k} such that $u_{i_k} \to \overline{u}$. Then it can be verified that $d_1(u_{i_k}, \tilde{u}_{i_k}) \to 0$, $d_2(u_{i_k}, \tilde{u}_{i_k}) \to F(\overline{u})$ and from the first condition in Lemma 3.7, we have

$$(u - \overline{u})^T F(\overline{u}) \geq 0, \quad \forall u \in \Omega, \tag{3.25}$$

which implies $\overline{u}$ is optimal. Using the fact that

$$\|u_{k+1} - \overline{u}\|_2^2 \leq \|u_k - \overline{u}\|_2^2, \tag{3.26}$$

we obtain the convergence of u_k.   □

4 Numerical experiments

In this section, we consider numerical experiments on sparse signal recovery and CT image reconstruction using TV-L_2 model.

4.1 Sparse signal recovery

The optimization problem is given as follows:

$$\underset{x \in \mathbb{R}^d}{\operatorname{argmin}} \frac{1}{2}\|Ax - b\|_2^2 + \mu \|Bx\|_1, \tag{4.1}$$

where $A \in \mathbb{R}^{n \times d}$ is a random matrix whose entries are drawn i.i.d from normal distribution and we choose $n = 1000$, $d = 200$ here. The matrix $B \in \mathbb{R}^{(d-1) \times d}$ is a discrete gradient operator whose diagonal entry are -1 and the upper diagonal entry is 1 and all the other entries are zeros. For the ground truth vector x_0, we first simulate a vector with all entries 1, then we randomly perturb 5% of its entries and the vector b is computed by $b = Ax_0 + \varepsilon$ where ε is unit random Gaussian noise. The regularization parameter is set as $\mu = 10^{-4}$.

The settings of the comparison are given as follows:

- For PC-PDFP, we consider the direction d_2 in the correction step i.e. we set $t = 1$ in Algorithm 2 as it gives the best performance.

- The comparisons between PDFP and iPDFP are performed by fixing the step sizes γ of the two methods, and choosing proper inertial parameters θ_k for iPDFP. In Table 1, we list the parameters and the name of the methods. We note that for some cases the inertial parameters θ_k can be chosen larger than the bound obtained in Theorem 3.1 and they can lead to faster convergence.

- The parameter λ is fixed as $\lambda = \frac{1}{4}$ (which meets the conditions in the convergence analysis) for all the methods.

Table 1: Summary of the parameters for PDFP and iPDFP.

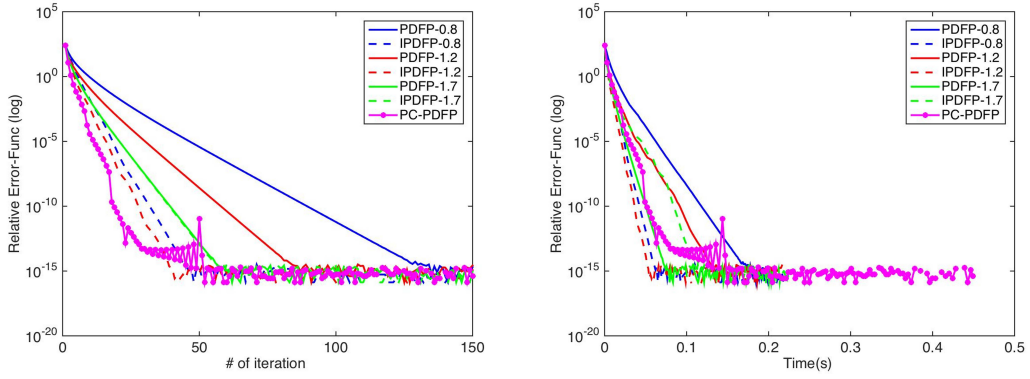| γ / θ_k	$\gamma = 0.8\beta$	$\gamma = 1.2\beta$	$\gamma = 1.7\beta$
$\theta_k = 0$	PDFP-0.8	PDFP-1.2	PDFP-1.7
$\theta_k = 0.5$	iPDFP-0.8	iPDFP-1.2	–
$\theta_k = 0.01$	–	–	iPDFP-1.7

Figure 1: Relative error of function value v.s. iterations number (left) and v.s. time (right).
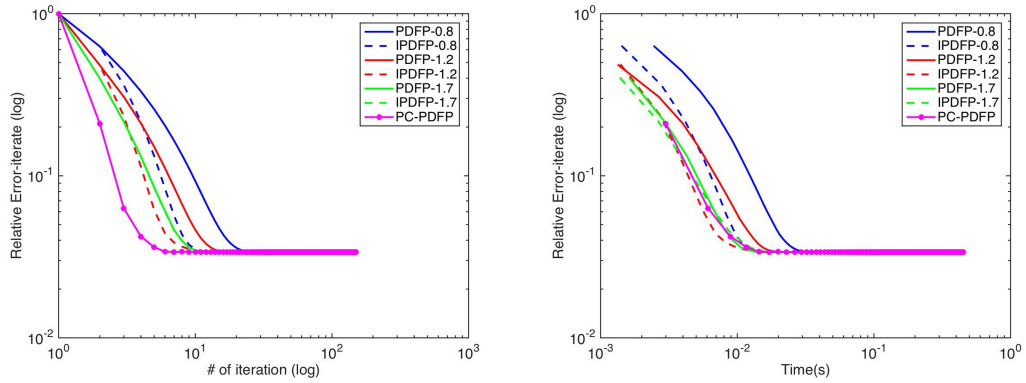

Figure 2: Relative error of iterate v.s. time (left) and v.s. iterations number (right).

- The relative error of the objective value is computed with respect to the optimal one, which is obtained by running PDFP for 10000 step when the convergence is observed. And the relative errors of the iterates are computed with respect to the ground truth signal.

Figs. 1 and 2 give the relative error of both function value and iterate over iteration number and time respectively. The observations are summarized as follows:

- Both iPDFP and PC-PDFP accelerate the original PDFP algorithm. In particular, with the same step size, iPDFP is faster than PDFP (see PDPF-0.8 *v.s.* iPDFP-0.8; PDPF-1.2 *v.s.* iPDFP-1.2). Moreover, the gain is more obvious with smaller stepsize γ. When the stepsize γ is approaching the theoretical upper bound, the improvement is marginal.

- It can be seen that iPDFP-1.2 is still faster than the best of PDFP algorithm: PDFP-1.7. This implies that using small stepsize and combining with the inertial step, iPDFP can achieve a better performance than the best of PDFP. In practice the best convergence of PDFP usually occurs near the upper bound 2β. But when the Lipschitz constant $\frac{1}{\beta}$ is over-estimated which leads to a small stepsize, applying iPDFP can improve the convergence speed.

- From Figs. 1 and 2, we can also see that PC-PDFP gets the best performance in terms of iteration number among three methods but it is not as fast as iPDFP in terms of computation time. This is due to the fact that PC-PDFP needs extra computation in the correction step. We can also observe some oscillation near the optimal, which may due to the adaptive step-size.

4.2 Image reconstruction

In this section we use the TV-L_2 model for CT image reconstruction with total variation regularization. The model still takes a similar form:

$$\min_{x \in \mathbb{R}^d} \|\mathcal{A}x - f\|_2^2 + \nu \|\nabla x\|_1, \tag{4.2}$$

where x is the image to be reconstructed and its dimension is $d = 512 \times 512$. The operator $\mathcal{A}$ is the X-ray transform and here we use fan beam scanning geometry [15] for which the number of detectors is 512 and the number of viewers is 180. The measured projection f is computed by applying $\mathcal{A}$ to the ground truth and adding the Gaussian noise with variance 0.3. The transform ∇ is the 2D discrete gradient operator and $\nu = 1$ is the regularization parameter. The experiments are performed on TITAN RTX GPU with 4608 cores and the version of Matlab is 2018b.

The parameters setting is listed in Table 2 for the comparisons of PDFP and iPDFP. Here γ^* is the (numerical) upper bound for the stepsize γ that ensures the convergence of the algorithms. The parameter λ is set as $\lambda = 1/8$ for both PDFP and iPDFP, $1/20$ for PC-PDFP. And the rest of parameters are tuned such that all the methods achieve the best performance.

Figs. 3 and 4 show the plot of PSNRs and the objective values of different methods over time and iterations number respectively. Some similar observations can be obtained. In particular, both PC-PDFP and iPDFP perform better than the original PDFP. In terms of iterations number, PC-PDFP and iPDFP with proper parameters ($\theta_k = 0.6$) perform the best; in terms of computation time, the objective values decreases faster and PSNR

Table 2: Summary of parameters of PDFP and iPDFP.

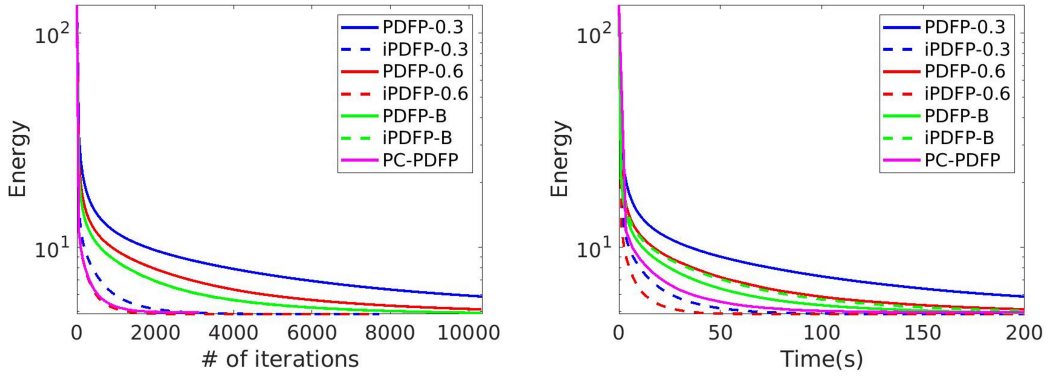| γ θ_k	$\gamma = \gamma^*$	$\gamma = \frac{1}{3}\gamma^*$	$\gamma = \frac{2}{3}\gamma^*$
$\theta_k = 0$	PDFP-B	PDFP-0.3	PDFP-0.6
$\theta_k = 0.001$	iPDFP-B	—	—
$\theta_k = 0.9$	—	iPDFP-0.3	iPDFP-0.6

Figure 3: Objective value v.s. iterations number (left) and v.s. time (right).
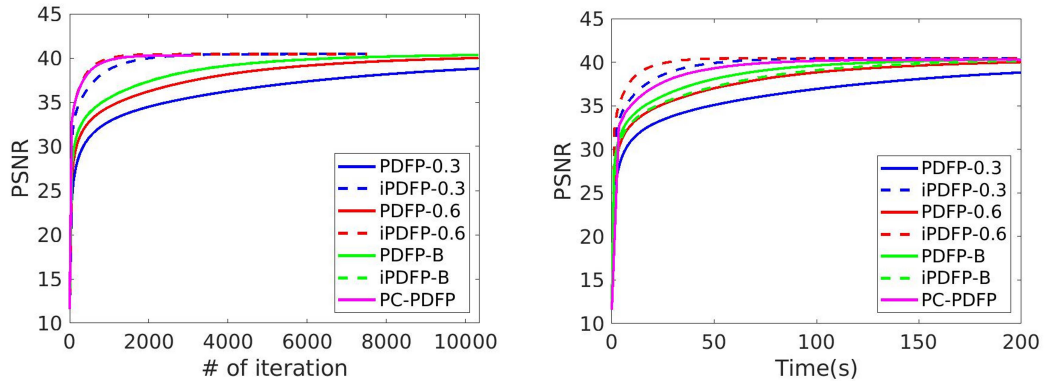

Figure 4: PSNR v.s. iteration (left) and v.s. time (right).

increases faster with iPDFP than that of PC-PDFP. Fig. 5 present the reconstructed images of the three methods and the respective computation times. With the same image quality, we can see that both iPDFP and PC-PDFP converge faster and iPDFP is the fastest.

(a) Ground truth   (b) PDFP, PSNR= 40.2 (110.0 s)

(c) IPDFP, PSNR= 40.2 (30.0 s)   (d) PC-PDFP, PSNR= 40.2 (78.6 s)

Figure 5: Reconstructed images and the respective computation times.

5   Conclusions

In this paper we discussed two acceleration schemes for PDFP: iPDFP and PC-PDFP. The convergence analysis of both algorithms were established. Numerical results on sparse signal reconstruction and CT reconstruction were present to demonstrate the performance of the two algorithms. We can observe that both algorithms are faster than PDFP and generally the energy of PC-PDFP decreases the fastest in terms of iterations number while iPDFP with proper inertial parameter is the fastest in terms of computation time, as iPDFP needs less computation at each iteration. Both algorithms can be potentially useful for some real applications in signal and image processing.

Acknowledgments

Appendix

In this section we give some details of the convergence analysis of iPDFP and PC-PDFP.

Proof of Lemma 3.3. By the optimal condition in the first equation of (3.3), we have

$$(p-\overline{p})^T\left(\frac{\lambda}{\gamma}\overline{s}-\frac{\lambda}{\gamma}B\overline{x}\right)\geq 0, \quad \overline{s}\in\partial g^*(\overline{p})$$
$$\Leftrightarrow (p-\overline{p})^T(\overline{s}-B\overline{x})\geq 0, \quad \overline{s}\in\partial g^*(\overline{p}), \quad \forall p\in V. \tag{A.1}$$

Using the second equation of (3.3), we get

$$\nabla f(\overline{x})+B^T\overline{p}=0. \tag{A.2}$$

Recall Eq. (2.2) we get the desired result. □

Proof of Lemma 3.4. By the update formula of the primal variable x, one has

$$(x-x_{k+1})^T(\gamma\nabla f(z_k)+\gamma B^T p_{k+1}+x_{k+1}-z_k)\geq 0, \quad \forall x\in\mathbb{R}^d. \tag{A.3}$$

Let $x=x^*$, then
$$(x^*-x_{k+1})^T(\gamma\nabla f(z_k)+\gamma B^T p_{k+1}+x_{k+1}-z_k)\geq 0. \tag{A.4}$$

By the optimality condition of (1.1): $\nabla f(x^*)+B^T p^*=0$, it follows that

$$(x^*-x_{k+1})^T(\gamma(\nabla f(z_k)-\nabla f(x^*))+\gamma B^T(p_{k+1}-p^*)+x_{k+1}-z_k)\geq 0. \tag{A.5}$$

Using the definition of z_k and the identity $\langle a,b\rangle=\frac{1}{2}\|a+b\|_2^2-\frac{1}{2}\|a\|_2^2-\frac{1}{2}\|b\|_2^2$, one gets

$$(x^*-x_{k+1})^T(\gamma(\nabla f(z_k)-\nabla f(x^*))+\gamma B^T(p_{k+1}-p^*)+x_{k+1}-x_k-\theta_k(x_k-x_{k-1}))\geq 0$$
$$\Leftrightarrow \gamma(x^*-x_{k+1})^T(\nabla f(z_k)-\nabla f(x^*))+\gamma(B(x^*-x_{k+1}))^T(p_{k+1}-p^*)+\frac{1}{2}\|x_k-x^*\|_2^2$$
$$-\theta_k(x^*-x_{k+1})^T(x_k-x_{k-1})\geq\frac{1}{2}\|x_{k+1}-x^*\|_2^2+\frac{1}{2}\|x_{k+1}-x_k\|_2^2$$
$$\Leftrightarrow \gamma(x^*-x_{k+1})^T(\nabla f(z_k)-\nabla f(x^*))-\frac{1}{2}\|x_{k+1}-x_k\|_2^2+\frac{\theta_k}{2}\|x_k-x_{k-1}\|_2^2$$
$$+\theta_k(x_{k+1}-x_k)^T(x_k-x_{k-1})$$
$$\geq\frac{1}{2}\|x_{k+1}-x^*\|_2^2-\frac{1}{2}\|x_k-x^*\|_2^2-\theta_k\left(\frac{1}{2}\|x_k-x^*\|_2^2-\frac{1}{2}\|x_{k-1}-x^*\|_2^2\right)$$
$$+\gamma(B(x_{k+1}-x^*))^T(p_{k+1}-p^*). \tag{A.6}$$

Estimate the first term of the left hand side of the last inequality of (A.6), we obtain

$$
\begin{aligned}
&(x^* - x_{k+1})^T(\nabla f(z_k) - \nabla f(x^*)) \\
&= (x^* - z_k)^T(\nabla f(z_k) - \nabla f(x^*)) + (z_k - x_{k+1})^T(\nabla f(z_k) - \nabla f(x^*)) \\
&\leq -\beta\|\nabla f(z_k) - \nabla f(x^*)\|_2^2 + \beta\|\nabla f(z_k) - \nabla f(x^*)\|_2^2 + \frac{1}{4\beta}\|z_k - x_{k+1}\|_2^2 \\
&= \frac{1}{4\beta}\|x_{k+1} - z_k\|_2^2.
\end{aligned}
\tag{A.7}
$$

Moreover

$$
\begin{aligned}
&\theta_k(x_{k+1} - x_k)^T(x_k - x_{k-1}) \\
&= (x_{k+1} - x_k))^T(z_k - x_k) \\
&= -\frac{1}{2}\|x_{k+1} - z_k\|_2^2 + \frac{1}{2}\|x_{k+1} - x_k\|_2^2 + \frac{\theta_k^2}{2}\|x_k - x_{k-1}\|_2^2.
\end{aligned}
\tag{A.8}
$$

Recall $E_x^{k+1} = \frac{1}{2}\|x_{k+1} - x^*\|_2^2 - \frac{1}{2}\|x_k - x^*\|_2^2$ and combine Eq. (A.6), Eq. (A.7) and Eq. (A.8), we have

$$
\gamma(B(x_{k+1} - x^*))^T(p_{k+1} - p^*) + E_x^{k+1} \leq \theta_k E_x^k + \frac{\gamma - 2\beta}{4\beta}\|x_{k+1} - z_k\|_2^2 + \frac{\theta_k + \theta_k^2}{2}\|x_k - x_{k-1}\|_2^2. \tag{A.9}
$$

This completes the proof. □

Proof of Lemma 3.5. Recall the update of p_{k+1}, we have

$$
(p - p_{k+1})^T\left(\frac{\lambda}{\gamma}s_{k+1} + p_{k+1} - v_k - \frac{\lambda}{\gamma}By_{k+1}\right) \geq 0, \quad s_{k+1} \in \partial g^*(p_{k+1}), \quad \forall p \in V. \tag{A.10}
$$

Let $p = p^*$, and use the definition of v_k, x_{k+1} and M, we then get

$$
\begin{aligned}
&(p^* - p_{k+1})^T\left(\frac{\lambda}{\gamma}s_{k+1} + p_{k+1} - v_k - \frac{\lambda}{\gamma}By_{k+1}\right) \geq 0 \\
&\Leftrightarrow (p^* - p_{k+1})^T\left(\frac{\lambda}{\gamma}s_{k+1} + p_{k+1} - v_k - \frac{\lambda}{\gamma}B(z_k - \gamma\nabla f(z_k) - \gamma B^T p_k)\right) \geq 0 \\
&\Leftrightarrow (p^* - p_{k+1})^T\left(\gamma s_{k+1} - \gamma Bx_{k+1} + R(p_{k+1} - p_k) - \theta_k R(p_k - p_{k-1})\right) \geq 0,
\end{aligned}
\tag{A.11}
$$

where $R = \frac{\gamma^2}{\lambda}M$.

Since $g(\cdot)$ is closed convex and $\partial g(Bx^*) \in p^*$, we have $Bx^* \in \partial g^*(p^*)$. Therefore the inequality (A.11) can be rewritten as

$$
(p^* - p_{k+1})^T\left(\gamma s_{k+1} - \gamma s^* - \gamma B(x_{k+1} - x^*) + R(p_{k+1} - p_k) - \theta_k R(p_k - p_{k-1})\right) \geq 0, \quad s^* \in \partial g^*(p^*). \tag{A.12}
$$

By the monotonicity of $\partial g^*(\cdot)$, we have

$$(p^* - p_{k+1})^T\left(-\gamma B(x_{k+1} - x^*) + R(p_{k+1} - p_k) - \theta_k R(p_k - p_{k-1})\right) \geq 0. \tag{A.13}$$

Using the identity $\langle a, b \rangle = \frac{1}{2}\|a+b\|_2^2 - \frac{1}{2}\|a\|_2^2 - \frac{1}{2}\|b\|_2^2$ and a similar argument as before we get

$$
\begin{aligned}
&\frac{1}{2}\|p_{k+1} - p^*\|_R^2 - \frac{1}{2}\|p_k - x^*\|_R^2 - \theta_k\left(\frac{1}{2}\|p_k - p^*\|_R^2 - \frac{1}{2}\|p_{k-1} - p^*\|_R^2\right) \\
&\quad -\gamma(p_{k+1} - p^*)^T(B(x_{k+1} - x^*)) \\
&\leq -\frac{1}{2}\|p_{k+1} - p_k\|_R^2 + \frac{\theta_k}{2}\|p_k - p_{k-1}\|_R^2 + \theta_k\langle p_{k+1} - p_k, R(p_k - p_{k-1})\rangle \\
&\leq -\frac{1}{2}\|p_{k+1} - p_k\|_R^2 + \frac{\theta_k}{2}\|p_k - p_{k-1}\|_R^2 + \frac{1}{2}\|p_{k+1} - p_k\|_R^2 + \frac{\theta_k^2}{2}\|p_k - p_{k-1}\|_R^2 \\
&= \frac{\theta_k + \theta_k^2}{2}\|p_k - p_{k-1}\|_R^2,
\end{aligned}
\tag{A.14}
$$

where the last inequality follows from $\langle a, Rb \rangle \leq \frac{1}{2}\|a\|_R^2 + \frac{1}{2}\|b\|_R^2$.

Recall $E_p^{k+1} = \frac{1}{2}\|p_{k+1} - p^*\|_R^2 - \frac{1}{2}\|p_k - p^*\|_R^2$, then Eq. (A.14) can be rewritten as

$$-\gamma(p_{k+1} - p^*)^T(B(x_{k+1} - x^*)) + E_p^{k+1} \leq \theta_k E_p^k + \frac{\theta_k + \theta_k^2}{2}\|p_k - p_{k-1}\|_R^2. \tag{A.15}$$

This completes the proof. $\qquad\qquad\square$

Proof of Lemma 3.6. Let

$$R_k = \begin{pmatrix} c(\theta_k - 1)^2 - \theta_k(1 + \theta_k) & 0 \\ 0 & (1 - 3\theta_k)R \end{pmatrix}. \tag{A.16}$$

The condition (3.6) indicates

$$R_k \geq \varepsilon \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix}. \tag{A.17}$$

Recall Eq. (A.14), one has

$$
\begin{aligned}
&E_p^{k+1} - E_p^k - \gamma(p_{k+1} - p^*)^T(B(x_{k+1} - x^*)) \\
&\leq -\frac{1}{2}\|p_{k+1} - p_k\|_R^2 + \frac{\theta_k}{2}\|p_k - p_{k-1}\|_R^2 + \theta_k\langle p_{k+1} - p_k, R(p_k - p_{k-1})\rangle \\
&\leq -\frac{1}{2}\|p_{k+1} - p_k\|_R^2 + \frac{\theta_k}{2}\|p_k - p_{k-1}\|_R^2 + \frac{\theta_k}{2}\|p_{k+1} - p_k\|_R^2 + \frac{\theta_k}{2}\|p_k - p_{k-1}\|_R^2 \\
&= \frac{\theta_k - 1}{2}\|p_{k+1} - p_k\|_R^2 + \theta_k\|p_k - p_{k-1}\|_R^2.
\end{aligned}
\tag{A.18}
$$

Recall $c = \frac{2\beta - \gamma}{2\beta}$, combining Eq. (A.18) and Eq. (3.4), we get

$$
\begin{aligned}
E_{k+1} - \theta_k E_k \leq & -\frac{c}{2}\|x_{k+1} - z_k\|_2^2 + \frac{\theta_k + \theta_k^2}{2}\|x_k - x_{k-1}\|_2^2 + \frac{\theta_k - 1}{2}\|p_{k+1} - p_k\|_R^2 \\
& + \theta_k \|p_k - p_{k-1}\|_R^2 \\
\leq & (\theta_k - 1)\frac{c}{2}\|x_{k+1} - x_k\|_2^2 + (\theta_k - \theta_k^2)\frac{c}{2}\|x_k - x_{k-1}\|_2^2 + \frac{\theta_k + \theta_k^2}{2}\|x_k - x_{k-1}\|_2^2 \\
& + \frac{\theta_k - 1}{2}\|p_{k+1} - p_k\|_R^2 + \theta_k \|p_k - p_{k-1}\|_R^2.
\end{aligned}
\tag{A.19}
$$

Let $\Delta u_{k+1} = (x_{k+1} - x_k, p_{k+1} - p_k)^T$, then Eq. (A.19) can be rewritten as

$$
E_{k+1} - \theta_k E_k \leq (\theta_k - 1)\frac{1}{2}\|\Delta u_{k+1}\|_A^2 + \frac{\theta_k}{2}\|\Delta u_k\|_{B_k}^2,
\tag{A.20}
$$

where

$$
A = \begin{pmatrix} cI & 0 \\ 0 & R \end{pmatrix}, \quad B_k = \begin{pmatrix} ((1-\theta_k)c + 1 + \theta_k)I & 0 \\ 0 & 2R \end{pmatrix}.
\tag{A.21}
$$

It can be seen that $-R_k = (\theta_k - 1)A + \theta_k B_k$.

Let

$$
\begin{aligned}
\Phi_k &= \frac{1}{2}\|u_k - u^*\|_{\mathcal{H}}^2 - \frac{\theta_k}{2}\|u_{k-1} - x^*\|_{\mathcal{H}}^2 + \frac{\theta_k}{2}\|\Delta u_k\|_{\mathcal{B}_k}^2 \\
&= Q_k - \theta_k Q_{k-1} + \frac{\theta_k}{2}\|\Delta u_k\|_{\mathcal{B}_k}^2.
\end{aligned}
\tag{A.22}
$$

Using the nondecreasing property of θ_k, we obtain

$$
\begin{aligned}
\Phi_{k+1} - \Phi_k &\leq E^{k+1} - \theta_k E^k + \frac{\theta_{k+1}}{2}\|\Delta u_{k+1}\|_{B_{k+1}}^2 - \frac{\theta_k}{2}\|\Delta u_k\|_{B_k}^2 \\
&\leq (\theta_k - 1)\frac{1}{2}\|\Delta u_{k+1}\|_A^2 + \frac{\theta_k}{2}\|\Delta u_k\|_{B_k}^2 + \frac{\theta_{k+1}}{2}\|\Delta u_{k+1}\|_{B_{k+1}}^2 - \frac{\theta_k}{2}\|\Delta u_k\|_{B_k}^2 \\
&\leq \frac{1}{2}\|\Delta u_{k+1}\|_{(\theta_{k+1}-1)A + \theta_{k+1}B_{k+1}}^2 \\
&= -\frac{1}{2}\|\Delta u_{k+1}\|_{R_{k+1}}^2.
\end{aligned}
\tag{A.23}
$$

Thus Φ_k is non-increasing and

$$
\frac{1}{2}\|u_k - u^*\|_{\mathcal{H}}^2 - \theta\frac{1}{2}\|u_{k-1} - u^*\|_{\mathcal{H}}^2 \leq \frac{1}{2}\|u_k - u^*\|_{\mathcal{H}}^2 - \theta_k\frac{1}{2}\|u_{k-1} - u^*\|_{\mathcal{H}}^2 \leq \Phi_k \leq \Phi_1.
\tag{A.24}
$$

It follows that

$$
\frac{1}{2}\|u_k - u^*\|_{\mathcal{H}}^2 \leq \theta^k \frac{1}{2}\|u_0 - u^*\|_{\mathcal{H}}^2 + \Phi_1 \sum_{i=0}^{k-1} \theta^i \leq \frac{\theta^k}{2}\|u_0 - u^*\|_{\mathcal{H}}^2 + \frac{\Phi_1}{1 - \theta}.
\tag{A.25}
$$

Meanwhile sum Eq. (A.23) for $k=1,\cdots,K$,

$$\sum_{i=1}^{K}\frac{1}{2}\|\Delta u_{k+1}\|_{R_{k+1}}^{2} \leq \Phi_{1}-\Phi_{K}$$

$$\leq \Phi_{1}+\frac{\theta_{K}}{2}\|u_{K-1}-u^{*}\|_{\mathcal{H}}^{2}$$

$$\leq \frac{\theta^{K}}{2}\|u_{0}-u^{*}\|_{\mathcal{H}}^{2}+\frac{\Phi_{1}}{1-\theta}+\Phi_{1}. \tag{A.26}$$

According to the condition

$$R_{k}\geq\varepsilon\begin{pmatrix} I & 0 \\ 0 & R \end{pmatrix}, \tag{A.27}$$

we get

$$\sum_{k=1}^{\infty}\|u_{k}-u^{*}\|_{\mathcal{H}}^{2} = \sum_{k=1}^{\infty}\|x_{k}-x_{k-1}\|_{2}^{2}+\|p_{k}-p_{k-1}\|_{R}^{2} < \infty. \tag{A.28}$$

This completes the proof. $\qquad\square$

References

[1] F. ALVAREZ, *Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in hilbert space*, SIAM Journal on Optimization, 14 (2004), pp. 773–782.

[2] F. ALVAREZ AND H. ATTOUCH, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Analysis, 9 (2001), pp. 3–11.

[3] O. BANERJEE, L. EL GHAOUI, AND A. D'ASPREMONT, *Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data*, J. Mach. Learn. Res., 9 (2008), pp. 485–516.

[4] A. BECK AND M. TEBOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.

[5] R. I. BOT AND E. R. CSETNEK, *An inertial alternating direction method of multipliers*, arXiv preprint arXiv:1404.4582, (2014).

[6] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems withapplications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 120–145.

[7] P. CHEN, J. HUANG, AND X. ZHANG, *A primaldual fixed point algorithm for convex separable minimization with applications to image restoration*, Inverse Problems, 29 (2013).

[8] P. CHEN, J. HUANG, AND X. ZHANG, *A primal-dual fixed point algorithm for multi-block convex minimization*, arXiv e-prints, (2016), arXiv:1602.00414, p. arXiv:1602.00414.

[9] P. CHEN, J. HUANG, AND X. ZHANG, *A primal-dual fixed point algorithm for minimization of the sum of three convex separable functions*, Fixed Point Theory and Applications, 2016 (2016), p. 54.

[10] P. L. COMBETTES AND V. R. WAJS, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling & Simulation, 4 (2005), pp. 1168–1200.

[11] L. CONDAT, *A primaldual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms*, Journal of Optimization Theory and Applications, 158 (2013), pp. 460–479.

[12] W. DENG AND W. YIN, *On the global and linear convergence of the generalized alternating direction method of multipliers*, Journal of Scientific Computing, 66 (2015), pp. 889–916.

[13] J. ECKSTEIN, *Approximate iterations in bregman-function-based proximal algorithms*, Mathematical Programming, 83 (1998), pp. 113–123.

[14] E. ESSER, X. ZHANG, AND T. 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.

[15] H. GAO, *Fast parallel algorithms for the x-ray transform and its adjoint*, Med Phys, 39 (2012), pp. 7110–20.

[16] T. GOLDSTEIN AND S. OSHER, *The split bregman method for l1-regularized problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 323–343.

[17] B. HE, L.-Z. LIAO, AND X. WANG, *Proximal-like contraction methods for monotone variational inequalities in a unified framework i: Effective quadruplet and primary methods*, Computational Optimization and Applications, 51 (2012), pp. 649–679.

[18] B. HE, L.-Z. LIAO, AND X. WANG, *Proximal-like contraction methods for monotone variational inequalities in a unified frameworkii: general methods and numerical experiments*, Computational Optimization and Applications, 51 (2012), pp. 681–708.

[19] B. HE AND X. YUAN, *On the $o(1/n)$ convergence rate of the douglasrachford alternating direction method*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 700–709.

[20] B. HE, X. YUAN, AND J. J. Z. ZHANG, *Comparison of two kinds of prediction-correction methods for monotone variational inequalities*, Computational Optimization and Applications, 27 (2004), pp. 247–267.

[21] D. A. LORENZ AND T. POCK, *An inertial forward-backward algorithm for monotone inclusions*, Journal of Mathematical Imaging and Vision, 51 (2015), pp. 311–325.

[22] C. A. MICCHELLI, L. SHEN, AND Y. XU, *Proximity algorithms for image models: denoising*, Inverse Problems, 27 (2011).

[23] Y. NESTEROV, *A method for solving the convex programming problem with convergence rate $\mathcal{O}(1/k^2)$*, Proceedings of the USSR Academy of Sciences, 269 (1983), pp. 543–547.

[24] Z. OPIAL, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc., 73 (1967), pp. 591–597.

[25] S. SETZER, *Split bregman algorithm, douglas-rachford splitting and frame shrinkage*, in Scale Space and Variational Methods in Computer Vision, X.-C. Tai, K. Mørken, M. Lysaker, and K.-A. Lie, eds., Berlin, Heidelberg, 2009, Springer Berlin Heidelberg, pp. 464–476.

[26] M. V. SOLODOV AND B. F. SVAITER, *Error bounds for proximal point subproblems and associated inexact proximal point algorithms*, Mathematical Programming, 88 (2000), pp. 371–389.

[27] B. C. V, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Advances in Computational Mathematics, 38 (2013), pp. 667–681.

[28] M. WEN, Y.-C. TANG, AND J. PENG, *An inertial primal-dual fixed point algorithm for composite optimization problems*, 2016.

[29] M. ZHU AND T. CHAN, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, UCLA CAM Report, 34 (2008).

[30] Y.-N. ZHU AND X. ZHANG, *Stochastic primal dual fixed point method for composite optimization*, Journal of Scientific Computing, 84 (2020), p. 16.