CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.2020-0042

Vol. **2**, No. 1, pp. 108-130
March 2021

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.*

Received 18 July 2020; Accepted 25 November 2020

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.

AMS subject classifications: 65K10, 49M29, 68U10, 94A08

Key words: Inertial iteration, prediction-correction, primal dual fixed point method, acceleration, composite optimization, image restoration.

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 $f(x)=\frac{1}{n}\sum_{i=1}^n(a_i^Tx-b_i)^2$

*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 $k=1,2,\cdots$

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

until the stop criterion is satisfied.