Numer. Math. Theor. Meth. Appl. Vol. **11**, No. 1, pp. 187-210
doi: 10.4208/nmtma.2018.m1621 February 2018

Adaptive Parallel Primal-Dual Method for Saddle Point Problems

Xiayang Zhang*

Department of Mathematics, Nanjing University, Nanjing210093, China

Received 22 June 2016; Accepted (in revised version) 24 March 2017

Abstract. The primal-dual hybrid gradient method is a classic way to tackle saddle-point problems. However, its convergence is not guaranteed in general. Some restrictions on the step size parameters, e.g., $\tau\sigma \le 1/\|A^TA\|$, are imposed to guarantee the convergence. In this paper, a new convergent method with no restriction on parameters is proposed. Hence the expensive calculation of $\|A^TA\|$ is avoided. This method produces a predictor like other primal-dual methods but in a parallel fashion, which has the potential to speed up the method. This new iterate is then updated by a simple correction to guarantee the convergence. Moreover, the parameters are adjusted dynamically to enhance the efficiency as well as the robustness of the method. The generated sequence monotonically converges to the solution set. A worst-case $\mathcal{O}(1/t)$ convergence rate in ergodic sense is also established under mild assumptions. The numerical efficiency of the proposed method is verified by applications in LASSO problem and Steiner tree problem.

AMS subject classifications: 49K35, 49M27, 90C25, 65K10

Key words: Adaptive, Parallel, Primal-dual method, Saddle-point problem, LASSO.

1. Introduction

This paper is concerned about solving the following saddle-point problem:

$$\min_{x\in X}\max_{y\in Y}\ g(x)+y^TAx-f^*(y), \tag{1.1}$$

where $A\in\mathfrak{R}^{m\times n}$, $X\subset\mathfrak{R}^n$, $Y\subset\mathfrak{R}^m$ are closed convex sets, g, f^* are convex functions and f^* is the conjugate function of a convex function f, $f^*(x)=\sup_{x\in domf}\big(y^Tx-f(x)\big)$. Note that this saddle-point problem is a primal-dual formulation of the nonlinear primal problem

$$\min_{x\in X}\ g(x)+f(Ax). \tag{1.2}$$

*Corresponding author. *Email address:* 369318324@qq.com (X. Y. Zhang)

 ©2018 Global-Science Press

The formulation (1.1) has a wide range of applications including image denoising [6, 24], statistical learning [21], compressive sensing [9] etc.

In many problems of practical interest, g and f^* do not share common properties, making it difficult to derive numerical schemes for (1.1) that address both terms simultaneously. Fortunately, it frequently occurs in practice that efficient algorithms exist for minimizing g and f^* separately. The primal-dual hybrid gradient (PDHG) method was first mentioned in [24] to tackle total variation (TV) minimization problems. This method removes the coupling between g and f^*, enabling each term to be addressed separately. Because it decouples g and f^*, the steps of PDHG can often be written explicitly, as opposed to other splitting methods that require expensive minimization sub-problems. As a result PDHG shows high numerical efficiency when applied to total variation (TV) minimization problems. However the convergence of PDHG is highly dependent on the choice of parameters. In [4], Chambolle and Pock's (CP) method improved PDHG method by a change on dual variable update. Their method is convergent and numerically competitive. CP method was further studied by He and Yuan in [12] . They explained the method from the aspect of Proximal Point Algorithm (PPA) and a relaxation factor was also introduced to PPA scheme to accelerate the convergence. More recently, Goldstein et al. introduced the Adaptive Primal-Dual Splitting (APD)method in [9] which tunes the step size parameters automatically for the CP method. The primal-dual decomposition method was proposed by O'Connor and Vandenberghe in [18] which applied the Douglas-Rachford splitting method to various splitting of the primal-dual optimality conditions.

More specifically, a general framework of some existing primal-dual methods solve the saddle-point problems (1.1) by the following procedures:

$$\begin{cases} x^{k+1} = \arg\min_{x\in X} \{g(x) + x^T A^T y^k + \frac{1}{2\tau}\|x - x^k\|^2\}, \\ \bar{x}^k = x^{k+1} + \theta(x^{k+1} - x^k), \\ y^{k+1} = \arg\min_{y\in Y} \{f^*(y) - y^T A\bar{x}^k + \frac{1}{2\sigma}\|y - y^k\|^2\}. \end{cases} \tag{1.3}$$

In (1.3), θ is called the combination parameter, $\sigma > 0$ and $\tau > 0$ are proximal parameters of the regularization terms, also referred as step size parameters in e.g. [9]. In [4], it was shown that the primal-dual procedure (1.3) is closely related to the extrapolational gradient methods in [15, 20], the Douglas-Rachford splitting method in [8, 16] and the alternating direction method of multipliers (ADMM) in [5]. With specific choice of parameters in (1.3), some existing primal-dual algorithms for (1.1) are recovered, and their convergence can be guaranteed when certain restriction are imposed on these parameters. Below are some examples.

- When $\theta = 0$, the primal-dual procedure in (1.3) reduces to the PDHG scheme in [24] which is indeed the Arrow-Hurwicz algorithm in [1]. This scheme has shown numerical efficiencies in [24] for TV image restoration problems. In [6], the convergence of the PDHG method has been studied insightfully by imposing additional restrictions ensuring that the parameters $\sigma > 0$ and $\tau > 0$ are small. However, a counter example has been given in [11] to show that PDHG method could be divergent even if $\sigma > 0$ and $\tau > 0$ are fixed at very small values.