Numer. Math. Theor. Meth. Appl. Vol. **13**, No. 1, pp. 176-199
doi: 10.4208/nmtma.OA-2019-0030 February 2020

Several Variants of the Primal-Dual Hybrid Gradient Algorithm with Applications

Jianchao Bai[1,2,*], Jicheng Li[2] and Zhie Wu[3]

[1] *Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an 710129, Shanxi, China*
[2] *School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China*
[3] *School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, Shaanxi, China*

Received 18 February 2019; Accepted (in revised version) 28 April 2019

Abstract. By reviewing the primal-dual hybrid gradient algorithm (PDHG) proposed by He, You and Yuan (SIAM J. Image Sci., 7(4) (2014), pp. 2526–2537), in this paper we introduce four improved schemes for solving a class of saddle-point problems. Convergence properties of the proposed algorithms are ensured based on weak assumptions, where none of the objective functions are assumed to be strongly convex but the step-sizes in the primal-dual updates are more flexible than the previous. By making use of variational analysis, the global convergence and sublinear convergence rate in the ergodic/nonergodic sense are established, and the numerical efficiency of our algorithms is verified by testing an image deblurring problem compared with several existing algorithms.

AMS subject classifications: 65K10, 65Y20, 90C90

Key words: Saddle-point problem, primal-dual hybrid gradient algorithm, variational inequality, convergence complexity, image deblurring.

1. Introduction

Let $\mathbb{R}$ be the set of real numbers, $\mathbb{R}^{m\times n}$ be the space of $m\times n$ dimensional real matrices, and $\mathbb{R}^n$ be the space of n dimensional real column vectors equipped with inner product $\langle\cdot,\cdot\rangle$ and Euclidean norm $\|z\|_2=\sqrt{\langle z,z\rangle}$ for any $z\in\mathbb{R}^n$. Consider the following general saddle-point problem

$$\min_{x_i\in\mathcal{X}_i}\max_{y_j\in\mathcal{Y}_j} F(x,y):=\sum_{i=1}^p f_i(x_i)-\sum_{i=1}^p\sum_{j=1}^p\langle y_j,A_ix_i\rangle-\sum_{j=1}^p g_j(y_j), \tag{1.1}$$

*Corresponding author. *Email address:* bjc1987@163.com (J. C. Bai)

where $x = (x_1^\top, x_2^\top, \cdots, x_p^\top)^\top$, $y = (y_1^\top, y_2^\top, \cdots, y_p^\top)^\top$ are grouped variables; $p \geq 1$ is any positive integer; $\mathcal{X}_i \subset \mathbb{R}^{m_i}$, $\mathcal{Y}_j \subset \mathbb{R}^n$ are structured and closed convex sets; $f_i(x_i) : \mathbb{R}^{m_i} \to \mathbb{R}$, $g_j(y_j) : \mathbb{R}^n \to \mathbb{R}$ are proper closed convex functions but possibly nonsmooth; and all $A_i \in \mathbb{R}^{n \times m_i}$ are given matrices. Throughout the discussions, the solution set of the problem (1.1) is assumed to be nonempty.

Minimization problems in the form of (1.1) arises in many possible applications, such as the 2D image denoising [12] and machine learning [3, Problem (2)]. In the past several years, a number of first-order algorithms had been developed for solving the problem (1.1) with case $p = 1$. For instance, Zhu and Chan [13] firstly proposed the primal-dual hybrid gradient algorithm (PDHG), whose iteration alternates between the primal and dual formulations, for solving total variation (TV) minimizations with applications in 2D image processing. Later, this PDHG was extended by Esser et al. [6] to solve a broader class of convex optimization models, and the modified version of PDHG was analyzed to have a similarly good empirical convergence rate for TV minimization problems. In 2011, Chambolle and Pock [4] showed an accelerated version of PDHG for non-smooth convex optimization problems with known saddle-point structure. In particular, their algorithm had $\mathcal{O}(1/t)$ convergence rate for non-smooth problems, and $\mathcal{O}(1/t^2)$ convergence rate for problems where either the primal or dual objective is uniformly convex. Here t denotes the iteration number. To better understand how to choose the step-sizes of the primal-dual updates, He et al. [9] revisited convergence of PDHG by an extremely simple example that it is not necessarily convergent when the step-sizes are fixed as tiny constants. The modified PDHG in [9], that is,

$$\begin{cases} x^{k+1} = \arg\min_{x \in \mathcal{X}} \left\{ F(x, y^k) + \dfrac{r}{2} \left\| x - x^k \right\|_2^2 \right\}, \\ y^{k+1} = \arg\max_{y \in \mathcal{Y}} \left\{ F(x^{k+1}, y) - \dfrac{s}{2} \left\| y - y^k \right\|_2^2 \right\}, \end{cases} \tag{1.2}$$

is indeed globally convergent under the following conditions:

- (A1) $f(x)$ is strongly convex with the modulus $\tau > 0$, i.e., there exists a positive constant τ such that for any $\xi \in \partial f(x)$, it holds

$$f(\widetilde{x}) - f(x) \geq \langle \widetilde{x} - x, \xi \rangle + \frac{\tau}{2} \|\widetilde{x} - x\|_2^2, \qquad \forall x, \widetilde{x} \in \mathcal{X};$$

- (A2) For given matrix A and $\tau > 0$, the parameter s in (1.2) satisfies $s > \frac{\rho(A^T A)}{\tau}$ where $\rho(\cdot)$ denotes the spectral radius of a matrix.

Clearly, (A1)-(A2) are strong and not always satisfied for some cases in real applications. For example, the TV regularized linear inversion problem, widely used as a model of salt-pepper noisy image deblurring [12], is of the following form

$$\min_{x \in \mathcal{X}} \left\{ \frac{1}{2} \|Kx - b\|_1 + \lambda \|Ax\|_{2,1} \right\}.$$