Journal of Computational and Applied Mathematics 330 (2018) 15–27

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Precompact convergence of the nonconvex Primal–Dual Hybrid Gradient algorithm

Tao Sun [a,b,*], Roberto Barrio [c], Lizhi Cheng [a,e], Hao Jiang [d]

[a] College of Science, National University of Defense Technology, Changsha, Hunan, 410073, China
[b] Department of Mathematics, University of California, Los Angeles, CA 90095, USA
[c] IUMA and CoDy group, Universidad de Zaragoza, Departamento de Matemática Aplicada, Pedro Cerbuna, 12, E-50009, Zaragoza, Spain
[d] College of Computer, National University of Defense Technology, Hunan, Changsha, 410073, China
[e] The State Key Laboratory for High Performance Computation, National University of Defense Technology, Hunan, Changsha, 410073, China

ARTICLE INFO

Article history:
Received 14 February 2017

Keywords:
Primal–Dual Hybrid Gradient algorithms
Kurdyka–Ł ojasiewic functions
Nonconvex optimization
Convergence analysis

ABSTRACT

The Primal–Dual Hybrid Gradient (PDHG) algorithm is a powerful algorithm used quite frequently in recent years for solving saddle-point optimization problems. The classical application considers convex functions, and it is well studied in literature. In this paper, we consider the convergence of an alternative formulation of the PDHG algorithm in the nonconvex case under the precompact assumption. The proofs are based on the Kurdyka–Ł ojasiewic functions, that cover a wide range of problems. A simple numerical experiment illustrates the convergence properties.

1. Introduction

This paper is devoted to solving the following well-studied primal problem

$$\min_x \Phi(x) := f(x) + g(Kx), \tag{1}$$

where $K \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$. If f and g are convex functions, one can represent model (1) as the following saddle-point problem

$$\min_x \max_y \Psi(x, y) := f(x) - y^\top Kx - g^*(y), \tag{2}$$

where g^* is the convex conjugate function of g. The saddle-point problem (2) is ubiquitous in different disciplines and applications, especially in the total variation regularization problem arising in imaging science [1–3].

A classical and frequently-used method for solving problem (2) is the Primal–Dual Hybrid Gradient (PDHG) algorithm [4,5]. Mathematically, PDHG algorithm can be described as the iterative process

$$\begin{cases} x^{k+1} = \arg\min_x \left\{\Psi(x, y^k) + \frac{r}{2}\|x - x^k\|_2^2\right\}, \\ y^{k+1} = \arg\max_y \left\{\Psi(x^{k+1}, y) - \frac{s}{2}\|y - y^k\|_2^2\right\}, \end{cases} \tag{3}$$

* Corresponding author.
E-mail addresses: tsun@math.ucla.edu (T. Sun), rbarrio@unizar.es (R. Barrio), clzcheng@nudt.edu.cn (L. Cheng), haojiang@nudt.edu.cn (H. Jiang).

http://dx.doi.org/10.1016/j.cam.2017.07.037
0377-0427/

where r and s are the step sizes of the method. If f and g are convex functions, another description (obtained by choosing $\theta = 0$ in [6]) of the PDHG algorithm is given by

$$\begin{cases} y^{k+1} \in \arg\min_{y} \left\{ \frac{s}{2}\|y - Kx^k\|_2^2 - \langle y, q^k\rangle + g(y) \right\}, \\ q^{k+1} = q^k + s(Kx^k - y^{k+1}), \\ x^{k+1} \in \arg\min_{x} \left\{ \frac{1}{2t}\|x - x^k\|_2^2 + \langle Kx, q^{k+1}\rangle + f(x) \right\}, \end{cases} \tag{4}$$

where now K, r and t are the parameters and step sizes of the method. We can easily see that in (4) the main steps in each iteration just lay on calculating the proximal maps of f and g. Compared with (3), the scheme (4) can be directly used to the nonconvex case. This is because in (3) the convex conjugate function g^* is used in the definition (2) of the function $\Psi(x, y)$, however, the conjugate function of a nonconvex function has not been well defined yet in literature, but the proximal maps used in (4) exist for closed nonconvex functions.

Although scheme (4) is apparently similar to the Alternating Direction Method of Multipliers (ADMM) algorithm [7–10], they are in fact quite different (the authors in [6] also point out this fact). Actually, PDHG has a deep relationship with the inexact Uzawa method [11]. In [12], the authors prove the convergence of PDHG under asymptotical assumptions on the step sizes. The sublinear convergence rate was established in [13] in an ergodic sense via variational inequalities, providing a very concise way to understand the convergence influenced by the step size. Meanwhile, paper [1] also proves the sublinear convergence rate of the PDHG, and this convergence problem is an active research subject in literature [14–17].

In all the previous papers, the convergence problem was studied for the convex case. In this paper, we study the convergence of the PDHG algorithm for the nonconvex case. More precisely, we consider the scheme (4) where f and g are both nonconvex functions. With the help of Kurdyka–Ł ojasiewic function properties, we prove that the points generated by scheme (4) converge to a critical point of Φ under the precompact assumption (that is, assuming the sequence is bounded). The proofs of our results are motivated by previous recent works [18–20]. Finally, we present a simple numerical experiment to show the convergence properties. We remark that the convergence results presented in this paper are based on the precompact assumption. Then, we do not prove the generic convergence problem for the nonconvex PDHG algorithm.

The paper is organized as follows: Section 2 presents some basic results and definitions; Section 3 contains the convergence results for the nonconvex PDHG algorithm; Section 4 reports a simple numerical example; and finally, Section 5 presents some conclusions.

2. Preliminaries

In this section we present the definitions and basic properties in variational and convex analysis and in Kurdyka–Ł ojasiewicz functions used later in the convergence analysis.

2.1. Subdifferentials

We collect several definitions as well as some useful properties in variational and convex analysis (see for more details the excellent monographes [21–23]).

Given a lower semicontinuous function $J : \mathbb{R}^N \to (-\infty, +\infty]$, its *domain* is defined by

$$\mathrm{dom}(J) := \{x \in \mathbb{R}^N : J(x) < +\infty\}.$$

The *graph* of a real extended valued function $J : \mathbb{R}^N \to (-\infty, +\infty]$ is defined by

$$\mathrm{graph}(J) := \{(x, v) \in \mathbb{R}^N \times \mathbb{R} : v = J(x)\}.$$

The notation of subdifferential plays a central role in (non)convex optimization.

Definition 2.1 (*Subdifferentials [21,22]*)**.** Let $J : \mathbb{R}^N \to (-\infty, +\infty]$ be a proper and lower semicontinuous function.

1. For a given $x \in \mathrm{dom}(J)$, the *Fréchet subdifferential* of J at x, written as $\hat{\partial}J(x)$, is the set of all vectors $u \in \mathbb{R}^N$ which satisfy

$$\liminf_{\substack{y \neq x \\ y \to x}} \frac{J(y) - J(x) - \langle u, y - x\rangle}{\|y - x\|_2} \geq 0.$$

   When $x \notin \mathrm{dom}(J)$, we set $\hat{\partial}J(x) = \emptyset$.
2. The (limiting) subdifferential, or simply the *subdifferential*, of J at $x \in \mathbb{R}^N$, written as $\partial J(x)$, is defined through the following closure process

$$\partial J(x) := \{u \in \mathbb{R}^N : \exists x^k \to x, J(x^k) \to J(x) \text{ and } u^k \in \hat{\partial}J(x^k) \to u \text{ as } k \to \infty\}.$$