Applied Mathematics and Computation 265 (2015) 708–722

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

A class of one parameter conjugate gradient methods☆

Shengwei Yao[a,*], Xiwen Lu[b], Liangshuo Ning[a], Feifei Li[a]

[a] School of Information and Statistics, Guangxi University of Finance and Economics, Nanning 530003, China
[b] School of Science, East China University of Science and Technology, Shanghai 200237, China

ARTICLE INFO

Keywords:
Unconstrained optimization
Continuous optimization
Conjugate gradient method
Global convergence
Wolfe line search

ABSTRACT

This paper proposes a class of one parameter conjugate gradient methods, which can be regarded as some kinds of convex combinations of some modified form of PRP and HS methods. The scalar β_k has the form of $\frac{\phi_k}{\phi_{k-1}}\mu_k$. The convergence of the given methods is analyzed by some unified tools which show the global convergence of the proposed methods. Numerical experiments with the CUTE collections show that the proposed methods are promising.

1. Introduction

In this paper, we focus on nonlinear conjugate gradient methods applied to the following nonlinear unconstrained optimization problem

$$\min f(x), x \in R^n, \tag{1}$$

where $f(x)$: $R^n \to R$ is a continuously differentiable function, bounded from below.

The conjugate gradient method is very efficient for solving (1), especially when n is large, and, has the following form:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

$$d_k = \begin{cases} -g_k, & \text{for } k = 1; \\ -g_k + \beta_k d_{k-1}, & \text{for } k \geq 2, \end{cases} \tag{3}$$

where $g_k = \nabla f(x_k)$ is the gradient of the objective function, α_k is a stepins size and β_k is a scalar.

Since the first nonlinear conjugate gradient method has been proposed by Fletcher and Reeves in 1964 [1], many formulae have been proposed to compute the scalar β_k (see refs. [2–11], etc.).

Among them, two typical formulae of β_k are the FR and PRP formulae (see refs. [1,4,5]), which are given by

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \tag{4}$$

and

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \tag{5}$$

☆ This research was partly supported by Guangxi Universities Foundation grant (no. KY2015YB268), Guangxi University of Finance and Economics grant (no. BS201413) and Quantitative Economics Key Laboratory Program of Guangxi University of Finance and Economics grants (no. 2014SYS04 and 2014CX06).
* Corresponding author. Tel.: +86 771 3233610.
E-mail addresses: idhot@163.com (S. Yao), xwlu@ecust.edu.cn (X. Lu), nileser@163.com (L. Ning), lorita2006@126.com (F. Li).

http://dx.doi.org/10.1016/j.amc.2015.05.115
0096-3003/

respectively, where $y_{k-1} = g_k - g_{k-1}$ is the gradient change, $\|\cdot\|$ denotes Euclidean norm. For different choice of the scalar β_k, properties of the nonlinear conjugate gradient methods may be quite different. During the past two decades, many different nonlinear conjugate gradient methods are proposed and analyzed individually (see refs. [12–18], etc.). More detail description can be found in Hager and Zhang's paper [19].

It is well known that some quasi-Newton methods can be expressed in a unified way, such as Broyden's family or Huang's family [20–22]. Therefore, a question is as follows: Does there exist a class of nonlinear conjugate gradient methods and their properties can be analyzed by a unified tool? For this question, Dai and Yuan gave a positive answer in [23]. In [23], authors observed that

$$\beta_k^{\mathrm{FR}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \tag{6}$$

and

$$\beta_k^{\mathrm{DY}} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}. \tag{7}$$

Combining (3) and (7), By direct calculations, one can deduce an equivalent form of β_k^{DY} as follows:

$$\beta_k^{\mathrm{DY}} = \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}. \tag{8}$$

Based on (4) and (8), the above two formulae are some special forms of

$$\beta_k = \frac{\phi_k}{\phi_{k-1}}, \tag{9}$$

where

$$\phi_k = \lambda\|g_k\|^2 + (1-\lambda)(-g_k^T d_k), \tag{10}$$

with $\lambda \in [0, 1]$ being a parameter. This class of formulae can be viewed as some kinds of convex combinations of the FR and DY formulae. From the expression of ϕ_k, it is formed only by the information g_k and $g_k^T d_k$ at k's iteration. So, ϕ_k and ϕ_{k-1} are independent of each other.

Recently, Wei et al. [24] proposed a new formula for computing β_k:

$$\beta_k^{\mathrm{WYL}} = \frac{g_k^T y_{k-1}^*}{\|g_{k-1}\|^2}, \tag{11}$$

$$y_{k-1}^* = g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}. \tag{12}$$

The formula β_k^{WYL} can be regarded as a modified form of β_k^{PRP}. Ref. [24] shows that the method with the formula β_k^{WYL} not only has nice numerical results but also possesses the sufficient descent condition and globally converges. In [25], Yao et al. extended a similar strategy to the HS formula as follows

$$\beta_k^{\mathrm{MHS}} = \frac{g_k^T y_{k-1}^*}{d_{k-1}^T y_{k-1}}, \tag{13}$$

and showed that the method with (13) globally converges under the following Wolfe line search. The Wolfe line search [26] is to find a positive stepsize α_k such that

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k, \tag{14}$$

and

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{15}$$

where $0 < \delta < \sigma < 1$.

In the convergence analysis and numerical implementations of conjugate gradient methods, the stepsize α_k is often computed by the strong Wolfe line search which requires α_k satisfied (14) and

$$|g(x_k + \alpha_k d_k)^T d_k| \le -\sigma g_k^T d_k, \tag{16}$$

where $0 < \delta < \sigma < 1$. This paper is also concerned about the general Wolfe line search which requires α_k satisfied (14) and

$$\sigma_1 g_k^T d_k \le g(x_k + \alpha_k d_k)^T d_k \le -\sigma_2 g_k^T d_k, \tag{17}$$

where $0 < \delta < \sigma_1 < 1$ and $\sigma_2 \ge 0$.

The basic topic of this paper is to investigate whether we can further extend Dai's formula (9) and obtain other methods which can be analyzed uniformly. The rest of this paper is organized as follows. In Section 2, by briefly reviewing some existing methods, the motivations of this paper are shown. Accordingly, a class of one parameter conjugate gradient methods are proposed. In Section 3, the convergence properties of the proposed methods are discussed. In Section 4, numerical experiments are carried out to comparing different values of the parameter. Some conclusions are given in the last section.