Numer. Math. Theor. Meth. Appl. Vol. **14**, No. 2, pp. 527-539
doi: 10.4208/nmtma.OA-2020-0056 May 2021

A Linear Hybridization of Dai-Yuan and Hestenes-Stiefel Conjugate Gradient Method for Unconstrained Optimization

Sindhu Narayanan and P. Kaelo*

Department of Mathematics, University of Botswana, Private Bag UB00704, Gaborone, Botswana

Received 4 April 2020; Accepted (in revised version) 19 October 2020

Abstract. Conjugate gradient methods are interesting iterative methods that solve large scale unconstrained optimization problems. A lot of recent research has thus focussed on developing a number of conjugate gradient methods that are more effective. In this paper, we propose another hybrid conjugate gradient method as a linear combination of Dai-Yuan (DY) method and the Hestenes-Stiefel (HS) method. The sufficient descent condition and the global convergence of this method are established using the generalized Wolfe line search conditions. Compared to the other conjugate gradient methods, the proposed method gives good numerical results and is effective.

AMS subject classifications: 90C06, 90C30, 65K05

Key words: Unconstrained optimization, conjugate gradient method, global convergence.

1. Introduction

Consider the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1.1}$$

where $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ is a continuously differentiable function that is bounded below. There are several numerical methods for solving the unconstrained optimization problem (1.1). These include conjugate gradient methods [1, 6, 9, 11, 19], Newton methods [18, 39], quasi-Newton methods [8, 12, 33, 37, 42, 43] and steepest descent methods [7, 22, 26, 49]. These methods are iterative, that is, given an initial guess $x_0 \in \mathbb{R}^n$, they generate a sequence $\{x_k\}$ using

*Corresponding author. *Email address:* `kaelop@ub.ac.bw` (P. Kaelo)

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

where $\alpha_k > 0$ is a step length and d_k is a descent direction. They also differ according to how the search direction d_k is obtained or updated. Newton and quasi-Newton methods require the second derivative information for updating the direction d_k and hence have good convergence rate. Conjugate gradient and steepest descent methods only require the first derivative information, which makes them more applicable to solving large-scale optimization problems.

The step length $\alpha_k > 0$ is chosen to satisfy certain line search conditions. Two of the usually used line searches are the strong Wolfe conditions

$$\begin{cases} f(x_k + \alpha_k d_k) \le f(x_k) + \sigma \alpha_k g_k^T d_k, \\ \left| g(x_k + \alpha_k d_k)^T d_k \right| \le \sigma_1 \left| g_k^T d_k \right|, \end{cases} \tag{1.3}$$

and the weak Wolfe conditions

$$\begin{cases} f(x_k + \alpha_k d_k) \le f(x_k) + \sigma \alpha_k g_k^T d_k, \\ g(x_k + \alpha_k d_k)^T d_k \ge \sigma_1 g_k^T d_k, \end{cases} \tag{1.4}$$

where $0 < \sigma < \sigma_1 < 1$.

In this paper, we consider solving problem (1.1) using a conjugate gradient method. Conjugate gradient methods generate the next iterate x_{k+1} by updating the direction d_k as

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k d_{k-1}, & k \ge 1, \end{cases} \tag{1.5}$$

where $g_k = \nabla f(x_k)$ is the gradient of the function f at x_k, and $\beta_k \in \mathbb{R}$ is a parameter known as the conjugate gradient coefficient. Different choices of β_k lead to different conjugate gradient methods, with the most well known methods being the Fletcher-Reeves (FR) [17], Polak-Ribière-Polyak (PRP) [36, 38], conjugate descent (CD) [16], Dai-Yuan (DY) [10], Liu-Storey (LS) [29] and Hestenes-Stiefel (HS) [21]. The PRP, LS and HS conjugate gradient methods have been shown to be numerically efficient while the others are theoretically effective. Other conjugate gradient methods have also been suggested in the literature [5,13,24,28,30,34,35,47] and a number of them are either modifications or hybridizations of the above methods. For instance, Wei *et al.* [44] proposed a conjugate gradient method

$$\beta_k^{WYL} = \frac{g_k^T \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right)}{\|g_{k-1}\|^2}, \tag{1.6}$$

which is a modification of the PRP method. It is globally convergent under weak Wolfe line search and numerically better than the PRP method. Similarly, Yao *et al.* [46] extended the above modification to HS method, that is,

$$\beta_k^{VHS} = \frac{g_k^T \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right)}{d_{k-1}^T y_{k-1}}, \tag{1.7}$$