East Asian Journal on Applied Mathematics
doi: 10.4208/eajam.140720.251220

Vol. **11**, No. 2, pp. 421-434
May 2021

An Efficient Mixed Conjugate Gradient Method for Solving Unconstrained Optimisation Problems

P. Mtagulwa and P. Kaelo*

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

Received 14 July 2020; Accepted (in revised version) 25 December 2020.

Abstract. Conjugate gradient algorithms are most commonly used to solve large scale unconstrained optimisation problems. They are simple and do not require the computation and/or storage of the second derivative information about the objective function. We propose a new conjugate gradient method and establish its global convergence under suitable assumptions. Numerical examples demonstrate the efficiency and effectiveness of our method.

AMS subject classifications: 90C06, 90C30, 65K05, 65K10

Key words: Global convergence, conjugate gradient method, sufficient descent, strong Wolfe conditions.

1. Introduction

There are various methods to solve the unconstrained optimisation problem

$$\min\ f(x), \tag{1.1}$$

where $x \in \mathbb{R}^n$ is an n dimensional real vector and $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function bounded from below. However, when n is very large, the conjugate gradient methods are usually preferable for solving problem (1.1) because of low memory requirements.

Nonlinear conjugate gradient methods for solving (1.1) start with an initial guess $x_0 \in \mathbb{R}^n$ and generate a sequence $\{x_k : k \geq 0\}$ of iterates by

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

where $\alpha_k > 0$ is the step length and d_k is the search direction

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if} \quad k > 0. \end{cases} \tag{1.3}$$

*Corresponding author. *Email addresses:* pmtagulwa@yahoo.com (P. Mtagulwa), kaelop@ub.ac.bw (P. Kaelo)

The scalar β_k, referred to as the conjugate gradient (CG) parameter, is uniquely defined for every conjugate gradient method, and $g_k = \nabla f(x_k)$ is the gradient of f at x_k. The step length α_k can be found by trust region methods, or by line search methods, which are categorised as either exact or inexact. Inexact line search methods are more often used because they require less computations. Among others, these methods include the weak Wolfe line search

$$\begin{aligned} &f(x_k+\alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k, \\ &g_{k+1}^T d_k \ge \sigma g_k^T d_k, \end{aligned} \tag{1.4}$$

the strong Wolfe line search

$$\begin{aligned} &f(x_k+\alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k, \\ &\left|g_{k+1}^T d_k\right| \le \sigma \left|g_k^T d_k\right|, \end{aligned} \tag{1.5}$$

and the generalised Wolfe conditions [25]

$$\begin{aligned} &f(x_k+\alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k, \\ &\sigma g_k^T d_k \le g_{k+1}^T d_k \le -\sigma_1 g_k^T d_k, \end{aligned}$$

where $0 < \delta < \sigma < 1$ and $\sigma_1 \ge 0$ are constants.

Following the introduction of nonlinear conjugate gradient method by Fletcher and Reeves (FR) [12] in 1964, various approaches have been developed to determine the conjugate parameter. The most popular include Polak-Ribière-Polyak [26,27], Hestenes-Stiefel (HS) [15], Dai-Yuan (DY) [5], Liu-Storey (LS) [22] and conjugate descent (CD) [11] methods. These traditional methods can be divided into two groups:

Group I.

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T(g_k - g_{k-1})}, \quad \beta_k^{CD} = -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}.$$

Group II.

$$\beta_k^{PRP} = \frac{\|g_k\|^2 - g_k^T g_{k-1}}{\|g_{k-1}\|^2}, \quad \beta_k^{HS} = \frac{\|g_k\|^2 - g_k^T g_{k-1}}{d_{k-1}^T(g_k - g_{k-1})}, \quad \beta_k^{LS} = -\frac{\|g_k\|^2 - g_k^T g_{k-1}}{d_{k-1}^T g_{k-1}},$$

where $\|\cdot\|$ is the Euclidean norm. Methods in Group I have strong convergence properties but their numerical performance is not always satisfactory. On the other hand, methods in Group II perform better numerically but do not always guarantee convergence. They perform better numerically because for short steps, the methods in Group II tend to switch to approximately the steepest descent direction, $d_k = -g_k$.

Therefore, over the years a lot of effort has been invested in improving the methods mentioned, either by modifying them or by considering three-term conjugate gradient methods. Moreover, in order to maintain good computational performance and strong