MJFAS MALAYSIAN JOURNAL OF FUNDAMENTAL AND APPLIED SCIENCES
PRINT ISSN: 2289-5981 | ONLINE ISSN: 2289-599X

FULL PAPER

Global convergence analysis of a new hybrid conjugate gradient method for unconstrained optimization problems

Ibrahim Abdullahi [a, b], Rohanin Ahmad [a, *]

[a] Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, Skudai 81310, Johor, Malaysia
[b] Department of Mathematics, Faculty of Science, Federal University Dutse, Jigawa State, Nigeria

[*] Corresponding author: rohanin@utm.my

Article history
Received 26 January 2017
Accepted 22 Mac 2017

Abstract

In this paper, we propose a new hybrid conjugate gradient method for unconstrained optimization problems. The proposed method comprises of β_k^{DY}, β_k^{YWH}, β_k^{RAMI}, and β_k^{New}. The β_k^{New} was constructed purposely for this proposed hybrid method. The method possesses sufficient descent property irrespective of the line search. Under Strong Wolfe-Powell line search, we proved that the method is globally convergent. Numerical experimentation shows the effectiveness and robustness of the proposed method when compare with some hybrid as well as some modified conjugate gradient methods.

Keywords: Unconstrained optimization, conjugate gradient, global convergence, descent properties, algorithm

INTRODUCTION

Let the function $f: R^n \to R$ be continuously differentiable. Consider the following unconstrained optimization problem

$$\min\{f(x): x_0 \in R^n\} \tag{1}$$

and its gradient is denoted by $g(x)$. We know for solving Eq. (1), starting from an initial guess $x_0 \in R^n$, conjugate gradient (CG) method generates a sequence $\{x_k\}$ as

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

and the direction d_k is defined by

$$d_{k+1} = \begin{cases} -g_k, & if\ k = 0; \\ -g_{k+1} + \beta_k d_k, & if\ \geq 1; \end{cases} \tag{3}$$

where x_k is the current iterate, β_k is an important formula called the CG coefficient and $\alpha_k > 0$ is the step-length obtained by a line search. In the line search computation, either exact or inexact line method is employed to compute the step-length α_k. In this paper, inexact line search used is given as follows:

$$f(x_k + \alpha_k d_k) \leq \delta \alpha_k g_k^T d_k \tag{4}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma |g_k^T d_k|, \tag{5}$$

where d_k is the descent direction and $0 < \delta < \sigma < 1$.

Over the years, research focused on the CG techniques which resulted to emergence of several formulas with differences in CG coefficient (β_k) in solving unconstrained optimization problems, the survey by Hager and Zhang (2006) discussed extensively on some methods with special emphasis on their global convergence. The summary of the classical CG methods are given in the Table 1 .

Table 1 The classical formulas for parameter β_k

β_k	Method Name	References
$\frac{\|\|g_{k+1}\|\|^2}{\|\|g_k\|\|^2}$	Fletcher-Reeves(FR) method	Fletcher and Reeves (1964)
$-\frac{\|\|g_{k+1}\|\|^2}{g_k^T d_k}$	Conjugate Descent(CD) method	Fletcher (2013)
$\frac{\|\|g_{k+1}\|\|^2}{d_k^T y_k}$	Dai-Yuan(DY) method	Dai and Yuan (1999)
$\frac{g_{k+1}^T y_k}{\|\|g_k\|\|^2}$	Polak-Rebiere-Polyak(PRP) method	Polyak (1969)
$-\frac{g_{k+1}^T y_k}{d_k^T g_k}$	Liu-Storey(LS) method	Hu and Storey (1991)
$\frac{g_{k+1}^T y_k}{d_k^T y_k}$	Hestenes-Stiefel(HS) method	Hestenes and Stiefel (1952)

where $||.||$ denote Euclidean norm. When the step-length $\alpha_k > 0$ is computed by exact line search condition, all the methods listed in Table 1 are equivalent if the objective function is convex quadratic but behave differently for non-convex cases. Methods such as FR, DY and CD are characterize with strong global convergence but they are not computationally powerful due to jamming phenomenon, that is, may take infinitely many steps without reaching optimum. While methods such as PRP, HS and LS may not always converge, but they often do better computational wise. Some of these attributes associated with these classical methods gave room for modification and hybridization of these existing methods to achieve a global convergence as well as better computational performances. Refer to (Hager and Zhang, 2005; Wei et al., 2006a; Wei et al., 2010; Zhang et al., 2012; Liu and Feng, 2011; Abdullahi and Ahmad, 2015; Du and Liu, 2011; Hager and Zhang, 2006; Abdullahi and Ahmad, 2016) for some modified CG methods in recent times. The works of (Dai and Wen, 2012; Yuhong, 2002) motivated Jiang and Jian (2013) to proposed modified CG method called modified Dai-Yuan (MDY) whose aim was to improve on the numerical performance of DY method and retain its good property. Also, same idea was extended to FR method called modified Fletcher-Reeves (MFR), where the parameters β_k were given by

$$\beta_k^{MDY} = \frac{||g_{k+1}||^2}{\max\{d_k{}^T y_k,\ \mu|g_{k+1}{}^T d_k|\}} \tag{6}$$

and

$$\beta_k^{MFR} = \frac{||g_{k+1}||^2}{\max\{||g_k||^2,\mu|g_{k+1}{}^T d_k|\}} \tag{7}$$

where $y_k = g_{k+1} - g_k$ and $\mu > 1$. The idea behind the work by Wei et al. (2006b) was extended to HS method by Shengwei et al. (2007) and proposed a CG method denoted by YWH where the parameter β_k is given as

$$\beta_k^{YWH} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{d_{k-1}{}^T (g_k - g_{k-1})} \tag{8}$$

under strong Wolfe line search with parameter $\alpha < \frac{1}{3}$. For general objective functions, the method always generate descent direction and it is globally convergent. Rivaie et al. (2016) proposed a new CG coefficient, RAMI for short form and defined by

$$\beta_k^{RAMI} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{d_{k-1}{}^T (d_{k-1} - g_k)} . \tag{9}$$

Global convergence of the method was established under exact line search. Numerical results showed the effectiveness of the method as compared to FR and PRP.

In this paper, our emphasis is on the hybrid CG methods. Combinations of different CG methods give rise to hybrid methods. Usually these methods are proposed to avoid jamming phenomenon and to improve the numerical experiment of CG methods in existence while the global convergence is also established. The first hybrid for the solution of unconstrained optimization problems was introduced by Touati-Ahmed and Storey (1990). Several developments have been recorded in this area which yielded a better performance as compared to the classical CG methods. Combination between PRP and FR lead to an hybrid proposed by Hu and Storey (1991) where the coefficient of CG method is denoted by HuS for convenience and the parameter β_k is given as

$$\beta_k^{HuS} = \max\{0, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\}. \tag{10}$$

The PRP method is used to address the jamming if it occurs, since it has a built-in restart feature. In the same regards, another hybrid method was proposed by Dai and Yuan (2001) which is a combination of HS and DY methods denoted by HSDY and the parameter β_k is given as

$$\beta_k^{HSDY} = \max\{0, \min\{\beta_k^{HS}, \beta_k^{DY}\}\}. \tag{11}$$

The global convergence of the method was established under standard Wolfe condition. Liu and Li (2014) proposed a new hybrid CG method on LS and DY methods. Their formula β_k is computed as a convex combination of β_k^{LS} and β_k^{DY} , that is, LSDY, where the parameter β_k is yielded by

$$\beta_k^{LSDY} = (1 - \gamma_K)\, \beta_k^{LS} + \gamma_K\, \beta_k^{DY} \tag{12}$$

where $\gamma_k \in [0,1]$. Numerical experiment showed the effectiveness of the method, the method is globally convergent under strong Wolfe line search. Jiang et al. (2012) proposed another hybrid method using the references (Jiang et al., 2011; Wei et al., 2006b; Dai and Yuan, 1999), where the parameter choice β_k denoted by JHJ is given as

$$\beta_k^{JHJ} = \frac{||g_k||^2 - \max\{0,\ \frac{||g_k||}{||d_{k-1}||} g_k{}^T d_{k-1},\ \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}\}}{d_{k-1}{}^T (g_k - g_{k-1})}. \tag{13}$$

Under Wolfe line search, they proved the global convergence of the method. The combination between LS and CD CG methods by Yang et al. (2013) gave birth to the following hybrid method denoted by LSCD

$$\beta_k^{LSCD} = \max\{0, \min\{\beta_k^{LS}, \beta_k^{CD}\}\}. \tag{14}$$

Refer to (Babaie-Kafaki and Ghanbari, 2014; Jian et al., 2015; Lu et al., 2015; Babaie-Kafaki, 2013; Kaelo, 2015; Zoutendijk, 1970) for more hybrid CG methods.

In contrast to some existing hybrid methods and in particular the hybrid method from (Jian et al., 2015; Jiang et al., 2012), we propose a new hybrid CG method based on the works by (Shengwei et al., 2007; Dai and Yuan, 1999; Rivaie et al., 2016) together with β_k^{New} for the purpose of this hybrid CG method given as

$$\beta_k^{IR} = \frac{||g_k||^2 - \max\{0,\ \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}\}}{masx\{d_{k-1}{}^T (d_{k-1} - g_k),\ d_{k-1}{}^T (g_k - g_{k-1})\}}. \tag{15}$$

β_k^{IR} comprises of β_k^{RAMI}, β_k^{DY}, β_k^{YWH}, and β_k^{New}

where

$$\beta_k^{New} = \frac{||g_k||^2}{d_{k-1}{}^T (d_{k-1} - g_k)} ,$$

therefore ,

$$\beta_k^{IR} = \beta_k^{RAMI} \text{ or } \beta_k^{DY} \text{ or } \beta_k^{YWH} \text{ or } \beta_k^{New}.$$

It has been shown, for general objective functions, method from Shengwei et al. (2007) can produce descent direction and it is globally convergent, while DY has strong convergence properties and also known to perform better than PRP and FR under exact line search (Rivaie et al., 2016). The good properties of these methods motivate us to propose a more robust CG method that possesses sufficient descent under any line search technique. The organization of the paper is as follows. In Section 2, we present the algorithm and prove that our formula can always generate descent directions. Section 3 presents the global convergence of our method. Section 4 covers the numerical experimentation of our method as well as representation of our method against other CG methods using performance profiles by Dolan and More' (2002).

ALGORITHM AND ITS SUFFICIENT DESCENT PROPERTY

In this section, we describe the CG algorithm and show that the propose formula Eq. (15) possesses the sufficient descent properties.

Algorithm 1

1. Initialization. Select $x_0 \in R^n$, *set* $k = 0$
2. Computation of parameter β_k based on Eq. (6), Eq. (7) and Eq. (10) - Eq. (15)
3. Generate d_k using Eq. (3). If $g_k = 0$, then stop.
4. Compute α_k based on inexact line search Eq. (4) and Eq. (5)
5. Variable update, $x_{k+1} = x_k + \alpha_k d_k$. Compute $f(x_{k+1})$, g_{k+1}.
6. Test for convergence. If $||g_k|| \leq \varepsilon$ then stop. Otherwise, set $k = k+1$ and go to 2.

Lemma 1. *Let the sequences* $\{x_k\}$ *and* $\{d_k\}$ *be generated by the algorithm. Then,* $g_k{}^T d_k < 0$ *hold true.*

Proof. We proceed by induction to arrive at the conclusion. Obviously, if $k = 1$, we have $g_1{}^T d_1 = -||g_1||^2 < 0$. Assume that $g_{k-1}{}^T d_{k-1} < 0$ holds true. To obtain $g_k{}^T d_k < 0$ particularly for our method, we divide the sufficient descent analysis into four parts. We assume that $\beta_k^{IR} \neq 0$. Clearly, for $\beta_k^{IR} = 0$ it follows from Eq. (3), $g_k{}^T d_k = -||g_k||^2 < 0$.

Case I

If $g_k{}^T g_{k-1} > 0$ and

$d_{k-1}{}^T(g_k - g_{k-1}) < d_{k-1}{}^T(d_{k-1} - g_k)$, then it follows from

$\beta_k^{IR} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)} = \beta_k^{RAMI}$. We know that

$d_{k-1}{}^T(g_k - g_{k-1}) > 0$, therefore, $d_{k-1}{}^T(d_{k-1} - g_k) > 0$

holds true and for $g_k{}^T g_{k-1} > 0$, we have $0 < \cos\theta_k < 1$, where θ_k is the angle between g_k and g_{k-1}. From Eq. (3) and Eq. (15), we have

$$g_k{}^T d_k = g_k(-g_k + \beta_k^{IR} d_{k-1})$$

$$= -||g_k||^2 + \frac{(||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}) g_k{}^T d_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$= -||g_k||^2 + \frac{||g_k||^2 g_k{}^T d_{k-1} - ||g_k||^2 \cos\theta_k \, g_k{}^T d_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$= \frac{-||g_k||^2 ||d_{k-1}||^2 + 2||g_k||^2 \, g_k{}^T d_{k-1} - ||g_k||^2 \cos\theta_k \, g_k{}^T d_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$= \frac{-||g_k||^2 ||d_{k-1}||^2 + (2 - \cos\theta_k)||g_k||^2 \, g_k{}^T d_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)} \quad (16)$$

From $d_{k-1}{}^T(g_k - g_{k-1}) < d_{k-1}{}^T(d_{k-1} - g_k)$, where

$g_{k-1}{}^T g_k < \frac{1}{2}(||d_{k-1}||^2 + g_{k-1}{}^T d_{k-1})$. Eq. (16) becomes

$$g_k{}^T g_k < \frac{-||g_k||^2 ||d_{k-1}||^2 + (2 - \cos\theta_k)||g_k||^2 \frac{1}{2}(||d_{k-1}||^2 + g_{k-1}{}^T d_{k-1})}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$= \frac{-\frac{1}{2}||g_k||^2 ||d_{k-1}||^2 \cos\theta_k + (2 - \frac{1}{2}\cos\theta_k)||g_k||^2 g_{k-1}{}^T d_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$< \frac{(2 - \frac{1}{2}\cos\theta_k)||g_k||^2 g_{k-1}{}^T d_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)} < 0 \quad . \quad (17)$$

Case II

If $g_k{}^T g_{k-1} \leq 0$ and $d_{k-1}{}^T(g_k - g_{k-1}) < d_{k-1}{}^T(d_{k-1} - g_k)$,

then from Eq. (15) we have $\beta_k^{IR} = \frac{||g_k||^2}{d_{k-1}{}^T(d_{k-1} - g_k)} = \beta_k^{New}$. Since $\beta_k^{IR} \neq 0$ and $d_{k-1}{}^T(g_k - g_{k-1}) > 0$ implies $d_{k-1}{}^T(d_{k-1} - g_k) > 0$. Also from $d_{k-1}{}^T(g_k - g_{k-1}) < d_{k-1}{}^T(d_{k-1} - g_k)$, we have

$$g_k{}^T d_{k-1} < \frac{1}{2}(||d_{k-1}||^2 + g_{k-1}{}^T d_{k-1}). \quad (18)$$

It follows from Eq. (3), Eq. (15) and Eq. (18)

$$g_k{}^T d_k = -||g_k||^2 + \frac{||g_k||^2 g_k{}^T d_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$= \frac{-||g_k||^2 ||d_{k-1}||^2 + 2||g_k||^2 g_k{}^T d_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$< \frac{-||g_k||^2 ||d_{k-1}||^2 + ||g_k||^2(||d_{k-1}||^2 + g_{k-1}{}^T d_{k-1})}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$= \frac{||g_k||^2 g_{k-1}{}^T d_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)} < 0 \quad (19)$$

Case III

If $g_k{}^T g_{k-1} > 0$ and $d_{k-1}{}^T(g_k - g_{k-1}) \geq d_{k-1}{}^T(d_{k-1} - g_k)$, then from Eq. (15) one gets $\beta_k^{IR} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{d_{k-1}{}^T(g_k - g_{k-1})} = \beta_k^{YWH}$.

Since the case in consideration is for $\beta_k^{IR} \neq 0$, then for $g_k{}^T g_{k-1} > 0$, we have $0 < \cos\theta_k < 1$ and it follows from Eq. (3) that

$$g_k{}^T d_k = g_k(-g_k + \beta_k^{IR} d_{k-1})$$

$$= -||g_k||^2 + \frac{(||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}) g_k{}^T d_{k-1}}{d_{k-1}{}^T(g_k - g_{k-1})}$$

$$= -||g_k||^2 + \frac{||g_k||^2 g_k{}^T d_{k-1} - ||g_k||^2 \cos\theta_k \, g_k{}^T d_{k-1}}{d_{k-1}{}^T(g_k - g_{k-1})}$$

$$= \frac{||g_k||^2 g_{k-1}{}^T d_{k-1} - ||g_k||^2 \cos\theta_k \, g_k{}^T d_{k-1}}{d_{k-1}{}^T(g_k - g_{k-1})}$$

$$< \frac{||g_k||^2 g_{k-1}{}^T d_{k-1} - ||g_k||^2 \cos\theta_k \, g_{k-1}{}^T d_{k-1}}{d_{k-1}{}^T(g_k - g_{k-1})}$$

$$= \frac{(1 - \cos\theta_k)||g_k||^2 g_{k-1}{}^T d_{k-1}}{d_{k-1}{}^T(g_k - g_{k-1})} < 0 \quad (20)$$

Case IV

If $g_k{}^T g_{k-1} \leq 0$ and $d_{k-1}{}^T(g_k - g_{k-1}) \geq d_{k-1}{}^T(d_{k-1} - g_k)$, then from Eq. (15) we get $\beta_k^{IR} = \frac{||g_k||^2}{d_{k-1}{}^T(g_k - g_{k-1})} = \beta_k^{DY}$. From Eq. (3) and Eq. (15), it follows that

$$g_k{}^T d_k = g_k(-g_k + \beta_k^{IR} d_{k-1}) = -||g_k||^2 + \frac{||g_k||^2 g_k{}^T d_{k-1}}{d_{k-1}{}^T(g_k - g_{k-1})}$$

$$= \frac{||g_k||^2 g_{k-1}{}^T d_{k-1}}{d_{k-1}{}^T(g_k - g_{k-1})} < 0 \tag{21}$$

Therefore, $g_k{}^T g_k < 0$ always holds $\forall k \geq 1$. Thus, the sufficient descent property is satisfied and the proof is completed.

Lemma 2. *The relation* $0 \leq \beta_k^{IR} \leq \frac{g_k{}^T d_k}{g_{k-1}{}^T d_{k-1}}$ *is always true for any* $k \geq 1$.

Proof. From Eq. (15), we know $\beta_k^{IR} \geq 0$. If $\beta_k^{IR} = 0$ and $g_k \neq 0$, then by Lemma 1, we have $\frac{g_k{}^T d_k}{g_{k-1}{}^T d_{k-1}} > 0$. Now putting into consideration, the four cases to show

$$\beta_k^{IR} \leq \frac{g_k{}^T d_k}{g_{k-1}{}^T d_{k-1}}.$$

Case I

If $g_k{}^T g_{k-1} > 0$ and $d_{k-1}{}^T(g_k - g_{k-1}) < d_{k-1}{}^T(d_{k-1} - g_k)$, then $\beta_k^{IR} = \beta_k^{RAMI}$. By Lemma 1 as well as inequality (17), we have

$$\beta_k^{IR} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)} = \frac{(2 - \frac{1}{2}\cos\theta_k)||g_k||^2}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$< \frac{g_k{}^T d_k}{g_{k-1}{}^T d_{k-1}}. \tag{22}$$

Case II

If $g_k{}^T g_{k-1} \leq 0$ and $d_{k-1}{}^T(g_k - g_{k-1}) < d_{k-1}{}^T(d_{k-1} - g_k)$, then $\beta_k^{IR} = \beta_k^{New}$. From Lemma 1 and inequality (19), we get

$$\beta_k^{IR} = \frac{||g_k||^2}{d_{k-1}{}^T(d_{k-1} - g_k)} < \frac{g_k{}^T d_k}{g_{k-1}{}^T d_{k-1}} \quad . \tag{23}$$

Case III

If $g_k{}^T g_{k-1} > 0$ and $d_{k-1}{}^T(g_k - g_{k-1}) \geq d_{k-1}{}^T(d_{k-1} - g_k)$, then $\beta_k^{IR} = \beta_k^{YWH}$. From Lemma 1 and inequality (20), we have

$$\beta_k^{IR} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{d_{k-1}{}^T(g_k - g_{k-1})} < \frac{g_k{}^T d_k}{g_{k-1}{}^T d_{k-1}}. \tag{24}$$

Case IV

If $g_k{}^T g_{k-1} \leq 0$ and $d_{k-1}{}^T(g_k - g_{k-1}) \geq d_{k-1}{}^T(d_{k-1} - g_k)$, then $\beta_k^{IR} = \beta_k^{DY}$. Using both Lemma 1 and Eq. (21) to get

$$\beta_k^{IR} = \frac{||g_k||^2}{d_{k-1}{}^T(g_k - g_{k-1})} = \frac{g_k{}^T d_k}{g_{k-1}{}^T d_{k-1}}. \tag{25}$$

Thus, the proof is completed.

GLOBAL CONVERGENCE

Throughout this section, we assume the following assumptions to be able to establish the global convergence of the proposed method,

Assumption(1) :

I. The level set $M = \{x \in R^n : f(x) \leq f(x_0)\}$ is bounded.

II. In some neighborhood N of M, the function is continuously differentiable and its gradient is Lipchitz continuous. i.e, there exist a constant $L > 0$ such that

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \forall x, y \in N. \tag{26}$$

The implication of this assumption on the function *f*, there exist a constant $\gamma \geq 0$ such that

$$||\nabla f(x)|| \leq \gamma, \forall x \in N. \tag{27}$$

The result of the following lemma, usually called Zoutendijk condition, is use to prove the global convergence of the proposed method. Refer to (Zoutendijk, 1970, Dai et al., 2000) for the proof.

Lemma 3. *Let* **Assumption(1)** *holds and consider any CG method of the form* $x_{k+1} = x_k + \alpha_k d_k$ *and the direction* $d_{k+1} = -g_{k+1} + \beta_k^{IR} d_k$ $d_0 = -g_0$ *where* α_k *satisfies* Eq. (4) and Eq. (5). *Then,*

$$\sum_{k\geq 1} \frac{(g_k{}^T d_k)^2}{||g_k||^2} < +\infty. \tag{28}$$

From Lemma 3,we have the following theorem which present the global convergence of the proposed method.

Theorem 1.

Let **Assumption(1)** *holds and the sequence* $\{x_k\}$ *and* $\{d_k\}$ *be generated by Algorithm 1 with* β_k^{IR}, α_k *is obtained by* Eq. (4) and Eq. (5). *Then*

$$\liminf_{k\to\infty} ||g_k|| = 0 \tag{29}$$

Proof. We proceed by contradiction to arrive at the conclusion. Suppose that $\liminf_{k\to\infty} ||g_k|| \neq 0$, it implies that there exists $m > 0$ such that

$$||g_k|| \geq m, \forall k \geq 0. \tag{30}$$

From Eq. (3), we have

$$(\beta_k^{IR} d_{k-1})^2 = (d_k + g_k)^2 \tag{31}$$

It follows from Eq. (31) and Lemma 2.

$$||d_k||^2 = (\beta_k^{IR})^2 ||d_{k-1}||^2 - 2g_k{}^T d_k - ||g_k||^2$$

$$\leq \left(\frac{g_k{}^T d_k}{g_{k-1}{}^T d_{k-1}}\right)^2 ||d_{k-1}||^2 - 2g_k{}^T d_k - ||g_k||^2. \tag{32}$$

Dividing both side of Eq. (32) by $(g_k{}^T d_k)^2$ to get

$$\frac{||d_k||^2}{(g_k{}^T d_k)^2} \leq \frac{||d_{k-1}||^2}{(g_k{}^T d_k)^2} - \frac{2}{g_k{}^T d_k} - \frac{||g_k||^2}{g_k{}^T d_k}$$

$$= \frac{||d_{k-1}||^2}{(g_k{}^T d_k)^2} - \left(\frac{1}{||g_k||} + \frac{||g_k||}{g_k{}^T d_k}\right)^2 + \frac{1}{||g_k||^2}$$

$$\leq \frac{||d_{k-1}||^2}{(g_{k-1}{}^T d_{k-1})^2} + \frac{1}{||g_k||^2}. \tag{33}$$

Hence

$$\frac{||d_k||^2}{(g_k{}^T d_k)^2} \le \sum_{i=1}^{k} \frac{1}{||g_i||^2} \le \frac{k}{m^2}, \quad (34)$$

Furthermore,

$$\frac{(g_k{}^T d_k)^2}{||d_k||^2} \ge m^2 \sum_{i=1}^{k} \frac{1}{k} = +\infty, \quad (35)$$

This contradict Zoutendijk condition in Eq. (28). Hence proof completed.

NUMERICAL RESULTS

In this section, experimentation of our proposed method has been carried out against some hybrid as well as some modified methods in the literature, to weigh the robustness of the algorithm with $\beta_k = \beta_k^{IR}$. To effect this, some test functions from (Andrei, 2008) and (Andrei, 2004) were considered. The inexact line search condition Eq. (4) and Eq. (5) were used in the computation for all formulas for easy comparison.

In carrying out the simulations, the number of iterations (IT), the number of function evaluations (NF) and CPU time (t) were put into consideration as parameters to determine the numerical strength of the proposed formula IR as compared with some hybrid methods and modified methods. The value of $\mu > 1$ in the cases of MDY and MFR as presented in (Jiang and Jian, 2013). For this experiment, $\mu = 1.2$ to avoid ambiguity is taken for both cases. For all the formulas under consideration, we take $\delta = 0.0001$ and $\sigma = 0.01$ for the purpose of this experiments and $||g_k|| \le \varepsilon$, where $\varepsilon = 10^{-5}$ is considered as the stopping criterion. We implemented the method using MATLAB R2015b (8.6.0.267246) in double precision arithmetic on a DELL computer, intel(R) Core (TM) i7-4790 CPU 3.60 GHz, 2TB HDD and 16.00GB RAM. A total of sixty (60) test functions gave birth to about seven hundred and fifty (750) problems resulted from different dimensions and initial points. The symbol (-) implies failure in numerical computation while (*) means that number of iterations or function evaluations exceeded the maximum limit set. For iteration, we set 5000 as the maximum while 20000 is the maximum for number of function evaluations.

The simulation results obtained from test functions in Table 2 and 3 were used to compared the numerical strength of the method in this paper as compared to the hybrid CG methods such as $\beta_k = \beta_k^{HSDY}$ by Dai and Yuan (2001), $\beta_k = \beta_k^{LSDY}$ by Liu and Li (2014), $\beta_k = \beta_k^{JHJ}$ by Jiang et al. (2012), $\beta_k = \beta_k^{LSCD}$ by Yang et al. (2013) and $\beta_k = \beta_k^{N}$ by Jian et al. (2015), also some modified CG methods such as $\beta_k = \beta_k^{WYL}$ by Wei et al. (2006b) , $\beta_k = \beta_k^{MFR}$ and $\beta_k = \beta_k^{MDY}$ by Jiang and Jian (2013).

Graphically, the performance of the proposed hybrid method versus HSDY, LSCD, JHJ, LSDY, JHJ and N hybrid methods are produced in Fig. 1 through Fig. 9 based on the umber of iteration (IT), the number of function valuations (NF) and CPU time (t) using performance profiles by Dolan and More' (Dolan and Moré, 2002). Performance rofile is non-arguably one of the best software for performance comparison between CG based methods for unconstrained optimization problems of the set of solvers S on a test set P of problems. For every n_p problems, we assume there exist n_s solvers, denotes $t_{p,s}$ as the executing time (or number of iteration,number of function evaluation or others) required to solve problem p by solver s. The ratio defined by

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s}: s \in S\}}$$

is the base of the comparison, the performance on problem p by solver s is compare with the best performance by any solver on this problem. Assuming that a parameter $r_M \ge r_{p,s}$ for all p,s is chosen and if and only if solver s does not solve problem p. We would like to obtain an overall assessment though performance of solver s on any given problem might be of interest. Then, we have

$$\rho_s(t) = \frac{1}{n_p} \, size\{p \in P: \log_{p,s} r \le t\}.$$

$\rho_s(t)$ is the probability for the solver $s \in S$ that a performance ratio $r_{p,s}$ is within a factor $t \in R^n$ of the best ratio. The ρ_s is the cumulative distribution function for the performance ratio, since a plot of the performance reveals all of the major performance characteristics. The value of $\rho_s(1)$ is given as the probability a solver will win over other solvers. Based on the above performance profile, a solver with large probability $\rho_s(t)$ are preferred. In other words, the solver at the top most right of the figure represent the best solver.

Fig. 1 – Fig. 6 are taken out from Fig. 7- Fig. 9 to clearly show the strength of the proposed method versus other hybrid methods in consideration. It shows clearly from the performance profile, the proposed hybrid method is the best since the solver is at the most top right on the Fig. 1- Fig. 6 based on IT, NF and t. The proposed method has high speed of convergence and in the end its able solve most of tested problems above others.

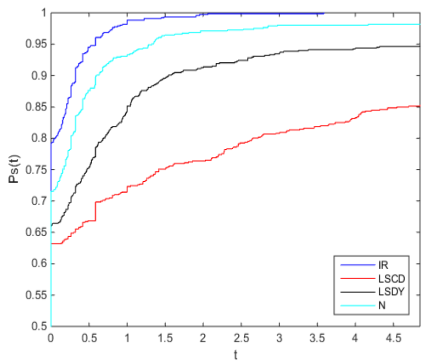

Fig. 1 Performance profile based on Iteration for IR versus hybrid methods(LSCD, LSDY and N)

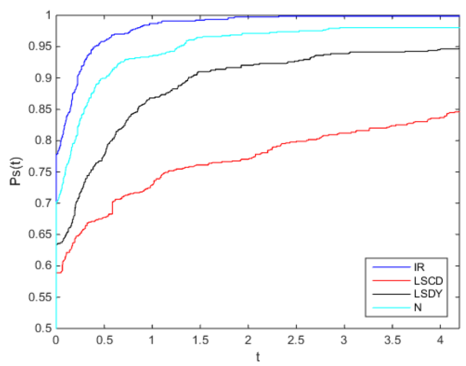

Fig. 2 Performance profile based on Function evaluation. IR versus hybrid methods(LSCD, LSDY and N)

Table 2. List of test functions

No	Function	Dim	Initial points
1	Extended Rosenbrock	2	(-1.2,1),2,5,8,10
2	Extended White Holst	2	(-1.2,1),3,15,17
3	Extended Beale	2	(1,0.8,...,0.8)
4	Extended Penalty	10,50,10000,20000	(1,2,...n),9,17
5	Perturbed Quadratic	2,4,32,500,1500,100000	-2, 0.5,5,8,13
6	Raydan 1	2,10,100,10000,50000	1,11
7	Raydan 2	2,20,200,20000,70000	1,3,6,9
8	Diagonal 1	2,8,80,800,8000	1,4,7,10,17
9	Diagonal 2	2,4,12,200,2000	1,10,20
10	Diagonal 3	2,20,500,20000	1,12
11	Hager	2,8,200,500	1,1.5,3.5,4.5,6.5
12	Extended Tridiagonal 1	2,7000,70000	2,4,6,8,10
13	Extended 3 Exponential Terms	20, 5000,100000	0.1,0.5
14	Generalized Tridiagonal 2	4,12,12000	-1,1,5,10,15
15	Diagal 4	2,200,500,5000	1,3,7,11,18
16	Diagonal 5	2,4,200,1000, 10000	1.1,1.2,3.1,4.1,5.1
17	Generalized PSC 1	120000	(3,0.1...0.1),6,12,18
18	Extended PSC 1	12000	(3,0.1...0.1),8,16
19	Extended Block Diagonal BD 1	12000,40000	1.1,3.1,7.1
20	Extended Cliff	2,12,120,200	-3,3,6
21	Quadratic Diagonal Perturbed	2,4,120,200,10000	0.5,2.5,4.5,6.5,8.5
22	Extended Hiebert	2	0,1,7,13,21
23	Quadratic	2,10,200,20000	1,13,15,17
24	Extended Quadratic Penalty	400,4000,7000	1,2,3,4,5
25	Extended Quadratic Penalty 2	4000	1
26	A Quadratic 2	4,12,40,4000,8000	0.5,3.5,7.5,9.5,12.5
27	Extended EP1	4,40,800,1000	1.5,3.5,5.5,6.5,8.5
28	Extended Tridiagonal 2	10,200, 2000,100000	1,4,12,20
29	TRIDIA	100,400,4000	1,2,3,4,5
31	NONDIA	4,12,40,120,12000	-1,-2,2,3
32	**NONDQUAR**	**5,5000**	**(1,-1,...1),-1,1,3**
33	DQDRTIC	5,10,50,500,5000	(1,-1,...1),3,5,6
34	DIXMAANA	4,12,32,400,4000,10000	2,3,4,8,13
35	DIXMAANB	4,12,32,400,4000,10000	2,3,4,8,13
36	DIXMAANC	4,12,32,400,4000,10000	2,3,4,8,13
37	DIXMAAND	4,12,32,400,4000,10000	2,3,4,8,13
38	DIXMAANL	4,12,32,400,4000,10000	2,3,4,8,13
39	Partial Perturbed Quadratic	12,120,1200	0.5,1.5,3.5,5.5,7.5
40	Broyden Tridiagonal	12,400,4000,40000	-3,-1,1,3
41	Almost Perturbed Quadratic	2,2000,8000	0.5,2.5,4.5,6.5
42	Tridiagonal Perturbed Quadratic	600,60000	0.5,2.5,6.5,9.5
43	HIMMELBHA	4,12,120,1200	(0,2,...,2)
44	STAIRCASE	4,32,400,4000,40000	1,2,4,7
45	LIARWHD	4,40,400,4000,40000	4,5,6,7
46	DIAGONAL 6	2,10,100,1000,10000	1,3,5,9
47	DIXON3DQ	400	-1,1,2,3
48	DENSCHNB	40,400,4000	1,3,4
49	SINQUAD	40,400,4000	0.1
50	BIGGSB 1	4,400,4000	0,2,3,5
51	Generalized quartic GQ1	4,40,400,4000	1,2,3,5
52	Diagonal 7	2,10,200,2000	1
53	Diagonal 8	2,20,200,2000	1
54	Full Hessian	2	1
55	Generalized quartic GQ2	4,32,40,400,4000	1,2,3,4
56	EXTROSN B	40000	3
57	ARGLINB	4,8,40,400,4000	(0.01,0.001,...,0.001),1.5,2.5,3.5
58	FLETCHCR	4,32,40,400,4000	0.5,1.5,2.5,3.5
59	HIMMELB G	4,16,40,400,4000	1.5,2.5,5.5,7.5
60	DIAGONAL 9	2,10,200,500,1000	1,2,4,6

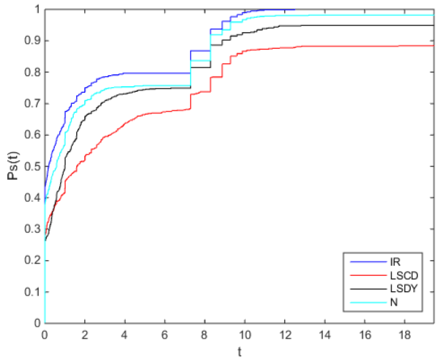

Fig. 3 Performance profile based on CPU time. IR versus hybrid methods (LSCD, LSDY and N)

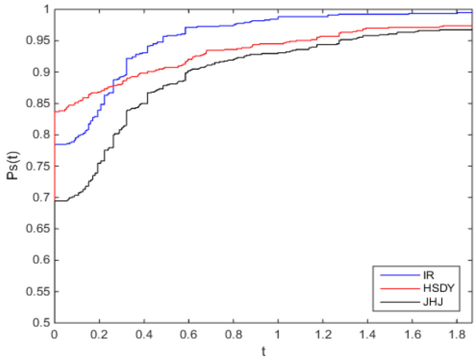

Fig. 4 Performance profile based on Iteration. IR versus hybrid methods(HSDY an d JHJ)

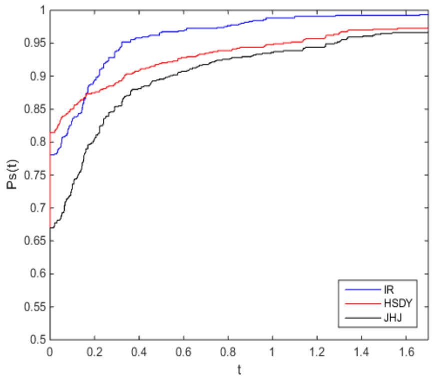

Fig. 5 Performance profile based on Function Evaluation. IR versus hybrid methods (HSDY and JHJ)

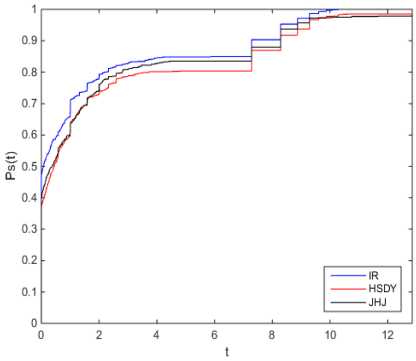

Fig. 6 Performance profile based on CPU time. IR versus hybrid methods(HSDY and JHJ)

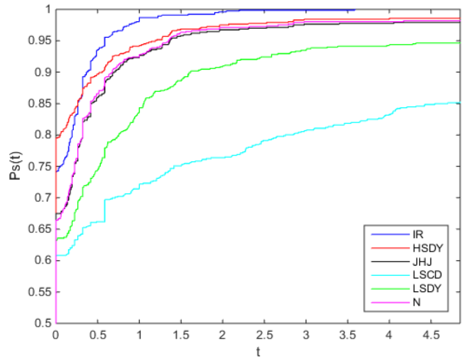

Fig. 7 Performance profile based on Iteration. IR versus hybrid methods(HSDY, JHJ, LSCD, LSCD and N)

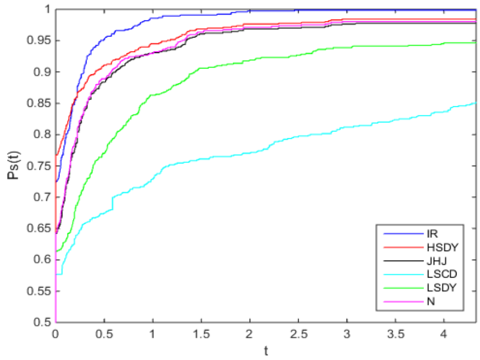

Fig. 8 Performance profile based on Function Evaluation. IR versus hybrid methods(HSDY, JHJ, LSCD, LSCD and N)

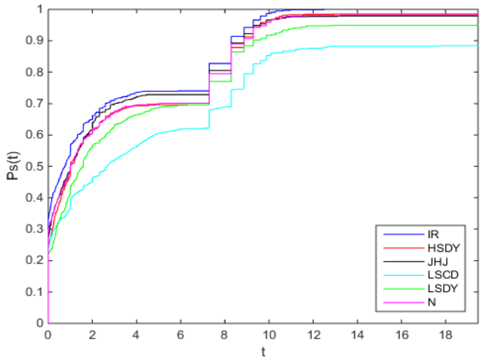

Fig. 9 Performance profile based on CPU time. IR versus hybrid methods(HSDY, JHJ, LSCD, LSCD and N)

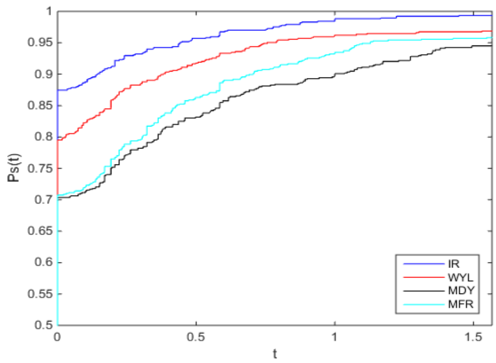

Fig. 10 Performance profile based on Iteration. IR versus WYL, MDY and MFR

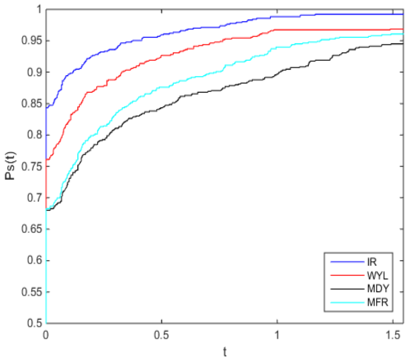

Fig. 11 Performance profile based on Function Evaluation. IR versus WYL, MDY and MFR

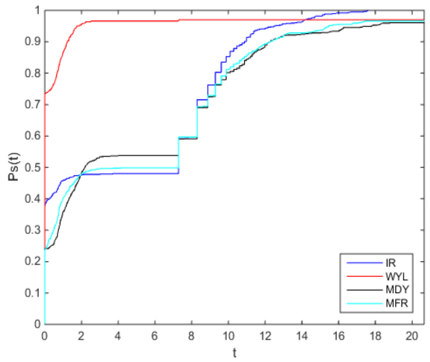

Fig. 12 Performance profile based on CPU time. IR versus WYL, MDY and MFR

From Fig. 9 – Fig. 12. The proposed hybrid method shown to be top performer when compared with modified methods (WYL, MFR and MDY) on the bases of number of iteration, number of function evaluation and CPU time.

CONCLUSION

Absorbing the advantages of some classical CG methods gave rise to hybrid methods in order to avoid the jamming phenomenon associated with them and to improve on their numerical strengths. We proposed a new type of hybrid CG method for solving unconstrained optimization problems. The proposed method satisfied sufficient descent condition irrespective of the line searches condition. The global convergence of the proposed method has been established under line search conditions Eq. (4) and Eq. (5). The parameter β_k^{IR} contains β_k^{DY} , β_k^{RAMI}, β_k^{YWH} and β_k^{New}. The β_k^{New}. constructed purposely for the proposed method. The simulation results of the proposed method showed to be efficient and robust as compared with hybrid CG methods (HSDY, JHJ, LSCD, LSCD and N). Furthermore, the proposed method is more robust than modified CG methods (WYL, MDY and MFR) for $\mu > 1$. Simultaneously, for clarity of purpose, the performance profiles by Dalon and More' are applied to show the effectiveness and robustness of the proposed method.

ACKNOWLEDGEMENT

This work was financially supported by the Universiti Teknologi Malaysia under the Research University Grant (RUG14H58), International Doctoral Fellowship (IDF) and Malaysian Ministry of Higher Education for the opportunity created to do research.

REFERENCES

Abdullahi, I., Ahmad, R. 2015. Convergence analysis of a new conjugate gradient method for unconstrained optimization. *Applied Mathematical Sciences, 9*, 6969-6984.

Abdullahi, I., Ahmad, R. 2016. Global convergence analysis of a nonlinear conjugate gradient method for unconstrained optimization problems. *Indian Journal of Science and Technology, 9*, 1-9. DOI: 10.17485/ijst/2016/v9i41/90175.

Andrei, N. 2004. Test functions for unconstrained optimization. *Research Institute for informatics. Center for Advanced Modeling and Optimization*, 8-10, Averescu Avenue, Sector 1, Bucharest, Romania, 1-15.

Andrei, N. 2008. An unconstrained optimization test functions collection. *Advanced Modeling and Optimization, 10*, 147-161.

Babaie-Kafaki, S. 2013. A hybrid conjugate gradient method based on a quadratic relaxation of the Dai–Yuan hybrid conjugate gradient parameter. *Optimization, 62*, 929-941.

Babaie-Kafaki, S., Ghanbari, R. 2014. Two hybrid nonlinear conjugate gradient methods based on a modified secant equation. *Optimization, 63*, 1027-1042.

Dai, Y. H., Yuan, Y. 1999. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization, 10*, 177-182.

Dai, Y., Han, J., Liu, G., Sun, D., Yin, H., Yuan, Y. X. 2000. Convergence properties of nonlinear conjugate gradient methods. *SIAM Journal on Optimization, 10*, 345-358.

Dai, Y., Yuan, Y. 2001. An efficient hybrid conjugate gradient method for unconstrained optimization. *Annals of Operations Research, 103*, 33-47.

Dai, Z., Wen, F. 2012. Another improved Wei–Yao–Liu nonlinear conjugate gradient method with sufficient descent property. *Applied Mathematics and Computation, 218*, 7421-7430.

Dolan, E. D., More, J. J. 2002. Benchmarking optimization software with performance profiles. *Mathematical Programming, 91*, 201-213.

Du, X., Liu, J. 2011. Global convergence of a spectral hs conjugate gradient method. *Procedia Engineering, 15*, 1487-1492.

Fletcher, R. 2013. *Practical Methods of Optimization*. Available from http://onlinelibrary.wiley.com/doi/10.1002/9781118723203.ch2

Fletcher, R., Reeves, C. M. 1964. Function minimization by conjugate gradients. *The Computer Journal, 7*, 149-154.

Hager, W. W., Zhang, H. 2005. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM Journal on Optimization, 16*, 170-192.

Hager, W. W., Zhang, H. 2006. A survey of nonlinear conjugate gradient methods. *Pacific Journal of Optimization, 2*, 35-58.

Hestenes, M. R., Stiefel, E. 1952. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards, 49*, 409-436.

Hu, Y., Storey, C. 1991. Global convergence result for conjugate gradient methods. *Journal of Optimization Theory and Applications, 71*, 399-405.

Jian, J., Han, L., Jiang, X. 2015. A hybrid conjugate gradient method with descent property for unconstrained optimization. *Applied Mathematical Modelling, 39*, 1281-1290.

Jiang, X. Z., Jian, J. B. 2013. A sufficient descent Dai–Yuan type nonlinear conjugate gradient method for unconstrained optimization problems. *Nonlinear Dynamics, 72*, 101-112.

Jiang, X., Han, L., Jian, J. 2012. A globally convergent mixed conjugate gradient method with Wolfe line search. *Mathematica Numerica Sinica, 34*, 103-112.

Jiang, X., Ma, G., Jian, J. 2011. A new global convergent conjugate gradient method with Wolfe line search. *Chinese Journal of Engineering Mathematics, 6*.

Kaelo, P. 2015. Another hybrid conjugate gradient method for unconstrained optimization problems. *Journal of Nonlinear Analysis and Optimization: Theory & Applications, 5*, 127-137.

Liu, J., Feng, Y. 2011. Global convergence of a modified LS nonlinear conjugate gradient method. *Procedia Engineering, 15*, 4357-4361.

Liu, J., Li, S. 2014. New hybrid conjugate gradient method for unconstrained optimization. *Applied Mathematics and Computation, 245*, 36-43.

Lu, Y., Li, W., Zhang, C., Yang, Y. 2015. A class new hybrid conjugate gradient method for unconstrained optimization. *Journal of Information and Computational Science, 12*, 1941-1949.

Polyak, B. T. 1969. The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics, 9*, 94-112.

Rivaie, M., Abashar, A., Mamat, M., Mohd, I. 2016. The convergence properties of a new type of conjugate gradient methods. Available from http://repository.rsu.edu.sd:8080/jspui/.

Shengwei, Y., Wei, Z., Huang, H. 2007. A note about WYL's conjugate gradient method and its applications. *Applied Mathematics and computation, 191*, 381-388.

Touti-Ahmed, D., Storey, C. 1990. Efficient hybrid conjugate gradient techniques. *Journal of Optimization Theory and Applications, 64*, 379-397.

Wei, Z., Li, G., Qi, L. 2006a. New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems. *Applied Mathematics and Computation, 179***,** 407-430.

Wei, Z., Yao, S., Liu, L. 2006b. The convergence properties of some new conjugate gradient methods. *Applied Mathematics and Computation, 183***,** 1341-1350.

Wei, Z. X., Huang, H. D., Tao, Y. R. 2010. A modified Hestenes-Stiefel conjugate gradient method and its convergence. *Journal of Mathematical Research and Exposition*, *30*. 297-308.

Yang, X., Luo, Z., Dai, X. 2013. A global convergence of LS-CD hybrid conjugate gradient method. *Advances in Numerical Analysis,* 2013. Available from http://dx.doi.org/10.1155/2013/517452.

Yuhong, D. 2002. A nonmonotone conjugate gradient algorithm for unconstrained optimization. *Journal of Systems Science and Complexity,* 15**,** 139-145.

Zhang, Y., Zheng, H., Zhang, C. 2012. Global convergence of a modified PRP conjugate gradient method. *Procedia Engineering*, *31***,** 986-995.

Zoutendijk, G. 1970. Nonlinear programming, computational methods. *Integer and Nonlinear Programming, 143***,** 37-86.