Asian Journal of Scientific Research

ISSN 1992-1454

science alert
http://www.scialert.net

ANSInet
an open access publisher
http://ansinet.com

OPEN ACCESS Asian Journal of Scientific Research

ISSN 1992-1454
DOI: 10.3923/ajsr.2018.185.194

Research Article

A Five-term Hybrid Conjugate Gradient Method with Global Convergence and Descent Properties for Unconstrained Optimization Problems

[1]Olawale Joshua Adeleke and [2]Idowu Ademola Osinuga

[1]Department of Mathematics, Covenant University, Ota, Ogun State, Nigeria
[2]Department of Mathematics, Federal University of Agriculture, Abeokuta, Ogun State, Nigeria

Abstract

Background and Objective: The nonlinear conjugate gradient method is a recurrence technique for solving effectively large-scale unconstrained optimization problems. In this study, a new hybrid nonlinear conjugate gradient method that combines the features of 5 different conjugate gradient methods is proposed with the aim of combining the positive features of different non-hybrid methods. **Methodology:** The proposed method was able to generate descent directions independent of line search procedures. By making assumptions on the objective function, the global convergence of the method was established under the standard Wolfe line search conditions. **Results:** Preliminary results showed that the method is very competitive and promising when subjected to comparison with other non-hybrid methods based on numerical experiments with selected benchmark test functions. **Conclusion:** As a future study, the proposed method will be tested against recently proposed related methods.

Key words: Unconstrained optimization problems, hybrid nonlinear conjugate gradient method, descent direction, global convergence, standard Wolfe line search conditions, numerical experiment

Received: August 10, 2017 Accepted: December 12, 2017 Published: March 15, 2018

Citation: Olawale Joshua Adeleke and Idowu Ademola Osinuga, 2018. A five-term hybrid conjugate gradient method with global convergence and descent properties for unconstrained optimization problems. Asian J. Sci. Res., 11: 185-194.

Corresponding Author: Olawale Joshua Adeleke, Department of Mathematics, Covenant University, Canaan land, PMB 1023, Ota, Ogun State, Nigeria
Tel: +2348036981967

Copyright: © 2018 Olawale Joshua Adeleke and Idowu Ademola Osinuga. This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

The field of optimization is growing at a very fast pace due to it wide range of application to many industrial[1] and environmental[2] problems. In this study, authors present a new hybrid conjugate gradient method with it global convergence results for solving unconstrained optimization problems. The conjugate gradient (CG) method is an efficient gradient based method that has been successfully applied to unconstrained optimization problems of the general form:

$$\text{Min } [f(x): x\in\mathbb{R}^n] \tag{1}$$

where, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously differentiable and its gradient, g(x), exists. The CG algorithm solves (1) iteratively using the recurrence rule:

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

In Eq. 2, α_k is the step-length at the kth iteration and can be computed using a suitable line search procedure (exact or inexact), while d_k is the search direction generated by the following rules:

$$d_0 = g_0 \text{ for } k = 0,\ d_k = -g_k+\beta_k d_{k-1} \text{ for } k\geq 1 \tag{3}$$

where, β_k is the CG update parameter and $g_k = \nabla f(x_k)$. Different values of β_k correspond to different CG methods. For instance:

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2},\ \beta_k^{HS} = \frac{g_k^T y_k}{d_{k-1}^T y_k},\ \beta_k^{PRP} = \frac{g_k^T y_k}{\|g_{k-1}\|^2},$$

$$\beta_k^{LS} = \frac{g_k^T y_k}{-d_{k-1}^T g_{k-1}},\ \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_k}$$

are the Fletcher and Reeves[3], Hestenes and Stiefel[4], Polak and Ribiere[5], Polyak[6], Liu and Storey[7] and Dai and Yuan[8] methods, respectively. In these methods, $\|\cdot\|$ is the Euclidean norm and $y_k = g_k-g_{k-1}$. For strict convex objective functions, the methods are equivalent. However, for nonconvex functions, their behaviours differ. Although, these methods represent the earliest CG methods, other variants have also been proposed recently with some exhibiting nice computational and convergence properties.

The FR and DY methods have been identified as having the best convergence results Al-Baali[9] and Dai and Yuan[8] for comprehensive proofs. However, for general objective functions, the two methods perform poorly computational-wise. Conversely, the HS and PRP methods have good computational strength even though they exhibit poor convergence results. This contrasting standpoint is the main motivation behind the development of hybrid methods, which are constructed with the objective of overcoming any deficiencies in two or more methods. For instance, a well-constructed hybrid method of FR and PRP should perform well computationally as well as yield good convergence properties. In this study, a new hybrid conjugate gradient method is proposed and analysed. The method was subjected to numerical test and compared to classical non-hybrid methods. The remainder of this paper is organized as follows. In section 2, a summary of related hybrid methods is presented and the newly proposed method is stated. A corresponding algorithm together with the result of the descent property is given in section 3. The proposed method is shown to be globally convergent in section 4. Numerical results and comparison with other methods are presented in section 5. The study is concluded in section 6 with direction towards possible future research.

OVERVIEW OF RELATED METHODS AND NEW HYBRID METHOD

One of the earliest developed hybrid methods may be found in Gilbert and Nocedal[10], where the positivity of β_k^{PRP} was considered crucial in establishing the convergence of the method. The β_k of the method was given by $\beta_k^{PRP+} = \max\{0, \beta_k^{PRP}\}$ and was shown to be globally convergent with the standard Wolfe line search condition. In Xu and Kong[11], carried out a linear combination of β_k^{DY} and β_k^{HS} methods to obtain an hybrid method with β_k as follows:

$$\beta_k = a_1\beta_k^{DY} + a_2\beta_k^{HS}$$

In this case, α_1 and α_2 are non-negative numbers and atleast one is not equal to zero and they satisfy:

$$0 < a_1 < 2a_2 < \frac{1}{1+\sigma_2} < 1$$

Kaelo[12] proposed a hybrid method that was borne out of the hybrid methods of Gilbert and Nocedal[10] and Dai and Yuan[13] which are represented as $\beta_k^{GN} = \max\{-\beta_k^{FR}, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\}$ and $\beta_k^{HS-DY} = \max\{-c\beta_k^{DY}, \min\{\beta_k^{HS}, \beta_k^{DY}\}\}$, respectively, where, $c = 1-\gamma/1+\gamma>0$ and $\gamma\in[½, 1]$. This method is given as:

$$\beta_k^{KK} = \max\{\min\{-c\beta_k^{PRP}, \beta_k^{FR}\}, \min\{\beta_k^{FR}, \beta_k^{PRP}\}\}$$

with the search direction given as:

$$d_k = \begin{cases} g_k & \text{if } k = 0 \\ -\theta_k\, g_k + \beta_k^{KK}\, d_{k-1} & \text{if } k \geq 1 \end{cases}$$

Where:

$$\theta_k = 1 + \beta_k^{KK} \frac{d_{k-1}^T g_k}{\|g_k\|^2}$$

and the search direction d_k satisfies the descent condition independent of any line search.

In Dai and Yuan[13], combined the methods in Hestenes and Stiefel[4] and Dai and Yuan[8] to obtain two hybrid methods. The β_ks in this case are:

$$\beta_k^1 = \max\left\{-c\beta_k^{DY}, \min\left\{\beta_k^{HS}, \beta_k^{DY}\right\}\right\}$$

and:

$$\beta_k^2 = \max\left\{0, \min\left\{\beta_k^{HS}, \beta_k^{DY}\right\}\right\}$$

For strictly convex quadratic functions and the use of exact line search, the method with β_k^1 reduces to the FR method. This establishes the claim of the authors that β_k^1 method is a CGM. Two results prompted the authors' suggestion of β_k^2. The first relates to the restart condition proposed by Powell[14] where restart was enabled if the following inequality holds:

$$\left|g_k^T g_{k-1}\right| > 0.2\|g_k\|^2 \tag{4}$$

Definitely $\left|g_k^T g_{k-1}\right| > 0.2\|g_k\|^2$ if $\beta_k^{HS} \leq 0$. This makes it appropriate to set β_k. In this case, the restart procedure will be along $-g_k$. The second reason given is connected to the fact that d_{k+1} may tend to the opposite of d_k if $\beta_k = 0$ and $\|d_k\| > \|g_k\|$. Thus enforcing that $\beta_k > 0$ will prevent two consecutive search direction from tending to be almost opposite.

Convex combination of CG algorithms is another common technique for obtaining hybrid CGM. Andrei[15] used the procedure to construct a hybrid method of DY and HS methods such that:

$$\beta_k = (1-\theta_k)\beta_k^{HS} + \theta_k\, \beta_k^{DY}$$

The value of θ_k is computed in a way that the direction corresponding to the CG algorithm is the Newton direction and the secont equation is also satisfied. The algorithm was implemented with the standard Wolfe line search condition. More recently, Liu and Li[16] considered the convex combination of β_k^{LS} and β_k^{DY} to obtain a hybrid CG method of the form $\beta_k^{HLSDY} = (1-\theta_k)\beta_k^{LS} + \theta_k\beta_k^{DY}$, with $\theta_k \in [0, 1]$ and satisfies the D-L conjugacy condition of Dai and Liao[17].

Wei *et al.*[18] proposed a new nonlinear CG method which is given by:

$$\beta_k^{WYL} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{\|g_{k-1}\|^2}$$

Shengwei *et al.*[19] extended the study Wei *et al.*[18] to the HS and LS methods and came up with a new version of the CG algorithm with the β_k given as:

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

and:

$$\beta_k^{MLS} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{-d_{k-1}^T g_{k-1}}$$

These two methods combined together give the YWH method. A number of hybrid methods have been constructed based on Wei *et al.*[18] and Shengwei *et al.*[19]. For instance, Li and Zhao[20] proposed a hybrid algorithm featuring β_k^{PRP} and β_k^{WYL} and given by:

$$\beta_k^{P-W} = \max\left\{\beta_k^{PRP}, \beta_k^{WYL}\right\}$$

This method possesses the tendency to move in the steepest descent direction if a small step is generated away from the solution, thus preventing the occurrence of a sequence of small steps. More recently, Jiang *et al.*[21] proposed a four term hybrid CG method with β_k obtained as:

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

Building on this idea, Jian *et al.*[22] introduced another four term hybrid method with the β_k in this case computed as:

$$\beta_k^N = \frac{\|g_k\|^2 - \max\left\{0, \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}\right\}}{\max\left\{\|g_{k-1}\|^2, d_{k-1}^T (g_k - g_{k-1})\right\}}$$

This method is simply one of DY/FR/WYL/YWH as it can be reduced to any of these methods with various assumptions.

The efficiencies of the hybrid methods constructed with the ideas in Wei *et al.*[18] and Shengwei *et al.*[19] serve as a motivation to develop a new method with 5 terms. The β_k of the proposed method is given by:

$$\beta_k^{hAO} = \frac{\|g_k\|^2 - \max\left\{0, \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}\right\}}{\max\left\{\|g_{k-1}\|^2, d_{k-1}^T(g_k - g_{k-1}), -d_{k-1}^T g_{k-1}\right\}} \quad (5)$$

Independent of any line search process, this method will always satisfy the descent direction criterion, that is, $d_k^T g_k < 0$. This will be shown in a moment from now. Observe that the proposed method is a hybrid of the FR, DY, WYL, MHS and MLS methods. Hence, the method is capable of exhibiting the positive characteristics of these methods. A CG algorithm to implement the proposed method is presented in the next section.

HYBRID CG ALGORITHM AND DESCENT PROPERTY

The algorithm to implement the proposed hybrid nonlinear CG method is as described below:

Hybrid CG algorithm:

Step 1: Initiate $x_0 \in \Re^n$, $\varepsilon > 0$. Set $d_0 = -g_0$ and k = 1

Step 2: While $\|g_k\| \leq \varepsilon$, continue to step 3, otherwise, stop

Step 3: Obtain α_k by a suitable line search technique (in this study, the standard Wolfe line search technique was used)

Step 4: Generate the sequences $\{x_k\}$, $\{g_k\}$ and $\{d_k\}$, where, $\beta_k = \beta_k^{hAO}$

Step 5: Set k = k+1 and return to step 2

In what follows, authors establish the descent property of their method.

Theorem 1: Let d_k and g_k be generated by the hybrid CG algorithm above. Then, the search direction d_k satisfies the descent condition:

$$d_k^T g_k \leq 0 \text{ for each } k \geq 0 \quad (6)$$

Proof: For k = 0, it can be shown that $d_0^T g_0 = -\|g_0\|^2$. Suppose its assume that $d_{k-1}^T g_{k-1} < 0$ for each k-1 and k > 2 and that $\beta_k^{hAO} = 0$ it will be obvious by the inner product of Eq. 3 with g_k that $d_k^T g_k = -\|g_k\|^2 + \beta_k d_{k-1}^T g_k = -\|g_k\|^2 < 0$. Thus, it will always assume that $\beta_k^{hAO} \neq 0$. Five different cases are of interest and considered as follow:

Case I: If $g_k^T g_{k-1} \leq 0$ and $\|g_{k-1}\|^2 \geq d_{k-1}^T(g_k - g_{k-1}) \geq -d_{k-1}^T g_{k-1}$. From Eq. 5 it can be deduce that $\beta_k^{hAO} = \beta_k^{FR}$. Therefore, from Eq. 3:

$$d_k^T g_k = \begin{bmatrix} g_k\left(-g_k + \beta_k^{FR} d_{k-1}\right) = -\|g_k\|^2 + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} d_{k-1}^T g_k \\ = \left(1 - \frac{d_{k-1}^T g_k}{\|g_{k-1}\|^2}\right)\|g_k\|^2 = \left(\frac{d_{k-1}^T g_k - \|g_k\|^2}{\|g_{k-1}\|^2}\right)\|g_k\|^2 \\ \leq \frac{d_{k-1}^T g_k}{\|g_{k-1}\|^2}\|g_k\|^2 = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} d_{k-1}^T g_{k-1} < 0 \end{bmatrix} \quad (7)$$

The first inequality uses the fact that $-\|g_{k-1}\|^2 \geq d_{k-1}^T(g_k - g_{k-1})$, while the second inequality uses the assumption at the beginning of this proof and the fact that $\beta_k^{FR} > 0$.

Case II: If $g_k^T g_{k-1} \leq 0$ and $d_{k-1}^T(g_k - g_{k-1}) \geq \|g_{k-1}\|^2 \geq -d_{k-1}^T g_{k-1}$, then Eq. 5 reduces to $\beta_k^{hAO} = \beta_k^{DY}$. Therefore:

$$d_k^T g_k = \begin{bmatrix} g_k\left(-g_k + \beta_k^{FR} d_{k-1}\right) = -\|g_k\|^2 + \frac{\|g_k\|^2}{d_{k-1}^T(g_k - g_{k-1})} d_{k-1}^T g_k \\ = \frac{d_{k-1}^T g_{k-1}}{d_{k-1}^T(g_k - g_{k-1})}\|g_k\|^2 < 0 \end{bmatrix} \quad (8)$$

since, $\|g_{k-1}\|^2 \geq 0 \Rightarrow d_{k-1}^T(g_k - g_{k-1}) \geq 0$ and the assumption that $d_{k-1}^T g_{k-1} < 0$.

Case III: If $g_k^T g_{k-1} > 0$ and $\|g_{k-1}\|^2 \geq d_{k-1}^T(g_k - g_{k-1}) \geq -d_{k-1}^T g_{k-1}$, then $\beta_k^{hAO} = \beta_k^{WYL}$. The fact that 0 $g_k^T g_{k-1} > 0$ means we can get an inequality $0 < \cos\theta_k < 1$ where, θ_k is the angle between g_k and g_{k-1}. Therefore, following the same process as in cases I and II:

$$d_k^T g_k = \begin{bmatrix} g_k\left(-g_k + \beta_k^{WYL} d_{k-1}\right) = -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{\|g_{k-1}\|^2} d_{k-1}^T g_k \\ = -\|g_k\|^2 + \frac{\|g_k\|^2 d_{k-1}^T g_k - \|g_k\|^2 \cos\theta_k\, d_{k-1}^T g_k}{\|g_{k-1}\|^2} \\ = -\|g_k\|^2 + \frac{(1 - \cos\theta_k)\|g_k\|^2 d_{k-1}^T g_k}{\|g_{k-1}\|^2} \\ = -\|g_k\|^2 + \frac{(1 - \cos\theta_k)\|g_k\|^2}{\|g_{k-1}\|^2}\left(\|g_{k-1}\|^2 + d_{k-1}^T g_{k-1}\right) \\ < \frac{(1 - \cos\theta_k)\|g_k\|^2}{\|g_{k-1}\|^2}\left(\|g_{k-1}\|^2 + d_{k-1}^T g_{k-1}\right) < 0 \end{bmatrix} \quad (9)$$

Case IV: If $g_k^T g_{k-1} > 0$ and $d_{k-1}^T(g_k - g_{k-1}) \geq \|g_{k-1}\|^2 \geq -d_{k-1}^T g_{k-1}$, then Eq. 5 becomes β_k^{MHS}. Therefore, from Eq. 3, case III and the fact that $\beta_k^{hAO} \neq 0$:

$$d_k^T g_k = \begin{bmatrix} g_k\left(-g_k + \beta_k^{MHS} d_{k-1}\right) \\ = -\|g_k\|^2 + \dfrac{\|g_k\|^2 - \dfrac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{d_{k-1}^T(g_k - g_{k-1})} d_{k-1}^T g_k \\ = -\|g_k\|^2 + \dfrac{(1-\cos\theta_k)\|g_k\|^2 d_{k-1}^T g_k}{d_{k-1}^T(g_k - g_{k-1})} \\ = -\dfrac{\|g_k\|^2 \cos\theta_k}{d_{k-1}^T(g_k - g_{k-1})} + \dfrac{\|g_k\|^2 d_{k-1}^T g_{k-1}}{d_{k-1}^T(g_k - g_{k-1})} \\ < \dfrac{\|g_k\|^2 d_{k-1}^T g_{k-1}}{d_{k-1}^T(g_k - g_{k-1})} < 0 \end{bmatrix} \quad (10)$$

Case V: If $g_k^T g_{k-1} > 0, -d_{k-1}^T g_{k-1} \geq \|g_{k-1}\|^2$ and:

$$-d_{k-1}^T g_{k-1} \geq d_{k-1}^T(g_k - g_{k-1})$$

then $\beta_k^{hAO} = \beta_k^{MLS}$ and the following process holds:

$$d_k^T g_k = \begin{bmatrix} g_k\left(-g_k + \beta_k^{MLS} d_{k-1}\right) \\ = -\|g_k\|^2 + \dfrac{\|g_k\|^2 - \dfrac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{-d_{k-1}^T g_{k-1}} d_{k-1}^T g_k \\ = -\|g_k\|^2 + \dfrac{(1-\cos\theta_k)\|g_k\|^2 d_{k-1}^T g_k}{-d_{k-1}^T g_{k-1}} \\ < \dfrac{(1-\cos\theta_k)\|g_k\|^2 d_{k-1}^T g_k}{-d_{k-1}^T g_{k-1}} < 0 \end{bmatrix} \quad (11)$$

The second inequality follows the facts that $-d_{k-1}^T g_{k-1} \geq d_{k-1}^T(g_k - g_{k-1}) \Rightarrow d_{k-1}^T g_k < 0$ and $d_{k-1}^T g_{k-1} < 0 \Rightarrow -d_{k-1}^T g_{k-1} > 0$.

Results in Eq. 7-11 confirm that the new method satisfies the descent property in its approach to obtaining the optimal value of the objective function.

Another important property of our hybrid method is that it satisfies the inequality $0 \leq \beta_k^{hAO} \leq \dfrac{d_k^T g_k}{d_{k-1}^T g_{k-1}}$ for every k $\geq$1. The proof to this may be established by following similar result in Kaelo[12].

GLOBAL CONVERGENCE

In this section, the proof of the global convergence result of the five-term hybrid method is established under the standard Wolfe line search conditions given by the following pair of inequalities:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta\alpha_k g_k^T d_k \quad (12)$$

and:

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \quad (13)$$

with $0<\delta\leq \sigma$.

To do this, the following assumptions on the objective function are stated.

Assumption:

- Level set $\Omega = \{x\in\mathbb{R}^n | f(x) \leq f(x_0)\}$ is bounded
- In some neighborhood $N\in\Omega$, f(x) is Lipschitz continuously differentiable, i.e., there exists a positive constant L such that:

$$\|g(x)-g(y)\| \leq L\|x-y\|,\ \forall x, y\in N \quad (14)$$

Assumptions (i)-(ii) imply that a positive constant m exists such that $\|g(x)\| \leq m\ \forall x\in N$.

An important result for proving the global convergence properties of nonlinear CG algorithms is the famous Zoutendijk[23] condition which is stated as a lemma here.

Lemma 1: Suppose assumptions (i)-(ii) hold. Given any iteration of the form (2), where d_k is a descent direction and α_k satisfies Eq. 12,13, then the following result holds:

$$\sum_{k=0}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < +\infty \quad (15)$$

A straightforward proof of this lemma can be found in Dai and Yuan[8]. Next its state and proof the global convergence result for the new hybrid method.

Theorem 2: If assumptions (i)-(ii) hold and $\{x_k\}$ is generated by the Hybrid CG Algorithm, then:

$$\liminf_{k\to\infty} \|g_k\| = 0 \quad (16)$$

Proof: The proof is given by contradiction. Let $\liminf_{k\to\infty} \|g_k\| \neq 0$. Then, since $\|g_k\|>0$, there exists a constant n>0 such that $\|g_k\| \geq n$, $\forall$k. Squaring both sides of Eq. 3 gives:

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

Dividing both sides of the above equation by $\left(d_k^T g_k\right)^2$ and using the fact that $0 \le \beta_k^{hAO} \le \frac{d_k^T g_k}{d_{k-1}^T g_{k-1}}$ for every k≥1, the following was obtained:

$$\frac{\|d_k\|^2}{\left(d_k^T g_k\right)^2} = \left[\begin{array}{l} \frac{\|d_{k-1}\|^2}{\left(d_{k-1}^T g_{k-1}\right)^2} - \frac{2}{d_k^T g_k} - \frac{\|g_k\|^2}{\left(d_k^T g_k\right)^2} \\ = \frac{\|d_{k-1}\|^2}{\left(d_{k-1}^T g_{k-1}\right)^2} - \left(\frac{1}{\|g_k\|} + \frac{\|g_k\|}{d_k^T g_k}\right)^2 + \frac{1}{\|g_k\|^2} \\ \le \frac{\|d_{k-1}\|^2}{\left(d_{k-1}^T g_{k-1}\right)^2} + \frac{1}{\|g_k\|^2} \end{array}\right] \quad (17)$$

Since $\frac{\|d_1\|^2}{\left(d_1^T g_1\right)^2} = \frac{1}{\|g_1\|^2}$, Eq. 16 gives:

$$\frac{\|d_{k-1}\|^2}{\left(d_{k-1}^T g_{k-1}\right)^2} \le \sum_{i=1}^{k} \frac{1}{\|g_i\|^2} \le \frac{k}{n}, \quad \forall\, k$$

This implies:

$$\sum_{k \ge 1} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} = \infty$$

which contradicts Eq. 15. Hence, the result is proved.

NUMERICAL TEST

In this section, a mild numerical experiment is presented to investigate the efficiency of the new hybrid method described above. The method with (7) as the CG parameter was tested numerically on Windows 7 with installed RAM of 2GB. The code for the algorithm was written with MatLab 7.10 using the Wolfe line search conditions (12)-(13) where, $\delta = 0.0001$ and $\sigma = 0.9$. The initial value of α_0 is 0 and the stopping condition is $\|g_k\| \le 10^{-6}$. Since the hybrid method combines the structures of the FR, DY, WYL, MHS and MLS methods, it is computationally convenient to compare the hAO method with any of the classical methods of FR, DY, WYL, HS and LS. For this particular study, the comparison was done in relation to FR and DY methods. The test functions were drawn from the collection of Andrei[24]. The test was conducted on 38 problems with different dimensions (n). In Table 1, the numerical results were presented in the form G/F/Itr/T, where G is the gradient evaluation, F is the function evaluation, Itr is the number of iterations and T is the CPU time. Whenever, a method fails to solve a problem, the cell is designated with -/-/?/? meaning the gradient and objective function values could not be obtained for the specified tolerance even though the algorithm was able to run with the problem producing the number of iterations and the CPU time.

In order to compare these methods for efficiency, the performance profile evaluation approach of Dolan and Moré[25] was adopted. Specifically, this technique helps to compare the performance of the hAO method against the FR and DY methods according to the number of iterations (Itr), the value of the objective function (F), the gradient norm (G) and the CPU time of computation (T), respectively. The theory of the performance profile is given as follows: Let M be the set of all methods to be compared and P, the set of all benchmark problems. Assume that M contains n_m methods and P contains n_p problems. For each problem $p \in P$ and method $m \in M$, suppose $I_{p,m}$ is the number of iterations (or the objective function value, etc) required for solving problem $p \in P$ by method $m \in M$, then the comparison between the different methods is based on the ratio given by:

$$r_{p,m} = \frac{I_{p,m}}{\text{Min}\left\{I_{p,m} : m \in M\right\}} \quad (18)$$

Equation 18 gives the required number of iterations for solving problem $p \in P$ with method $m \in M$. As a consequence, if there exists a parameter, r_k, which is large enough so that $r_k \ge r_{p,m}$ for all p, m. Equality holds, that is, $r_k = r_{p,m}$ only when the chosen method m does not solve the problem p. Based on Eq. 18, the cumulative distribution function for the performance ratio, $r_{p,m}$, is defined by:

$$\rho_m(\tau) = \frac{1}{n_p}\left|p \in P : \log r_{p,m} \le \tau\right| \quad (19)$$

where, $|\cdot|$ represents cardinality, $\rho_m(\tau)$ is the probability, in relation to method m, that $r_{p,m}$ is within a factor $\tau \in \Re^n$. At $\tau = 1$, the value of $\rho_m(\tau)$ is the probability that the method will out-perform the other methods.

The following performance profile figures (Fig. 1-4) were obtained from the numerical values in Table 1. Evidently from Fig. 1, the proposed method outperformed the classical methods based on the rate of convergence measured by the number of iterations. In term of computational time (Fig. 2), the proposed method also performed excellently well and very competitive with other methods for function and gradient evaluations (Fig. 3 and 4).

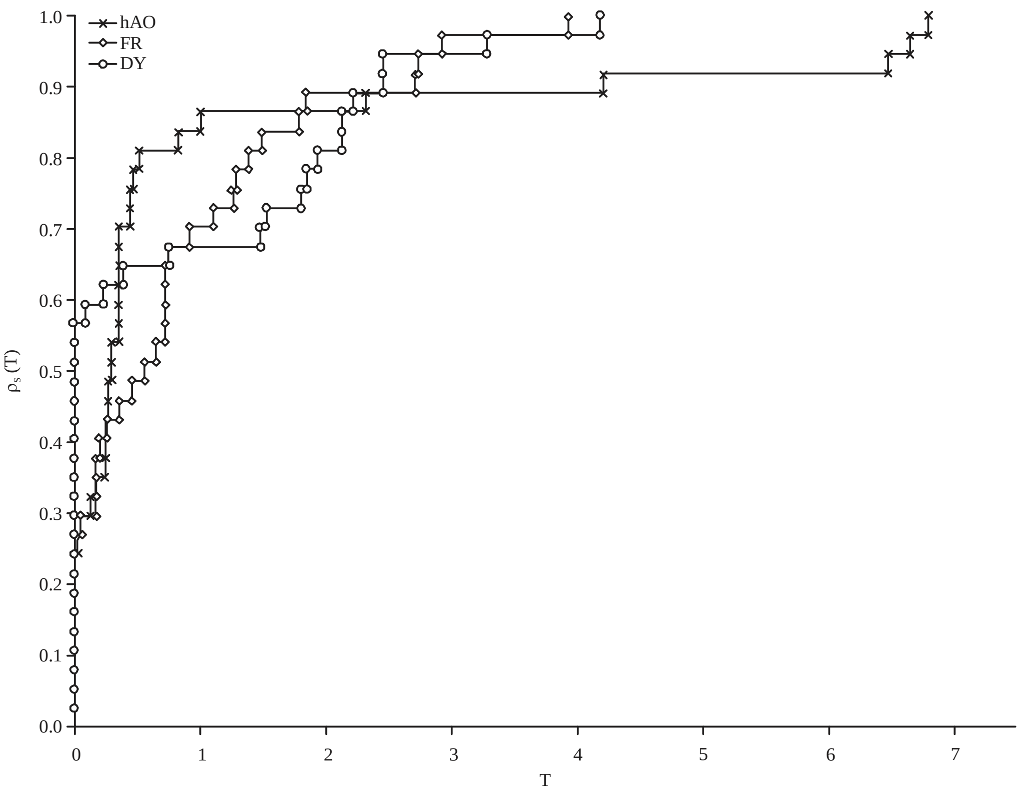

Fig. 1: Performance Profile according to the number of iterations

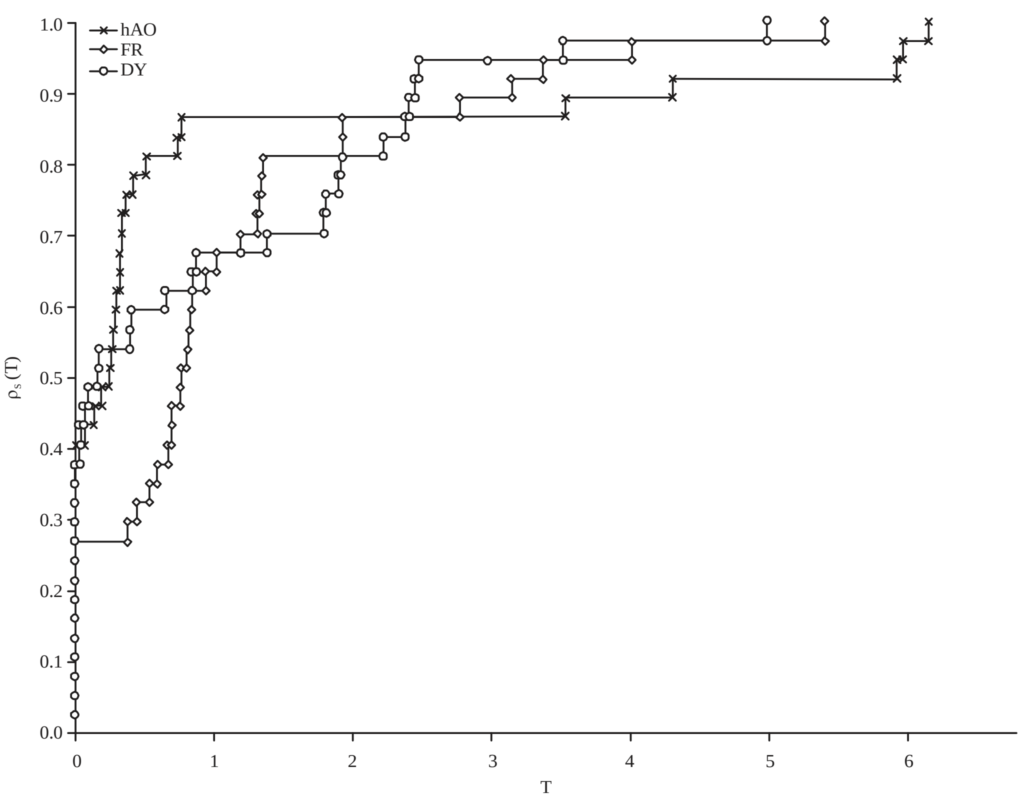

Fig. 2: Performance profile according to CPU time (sec)

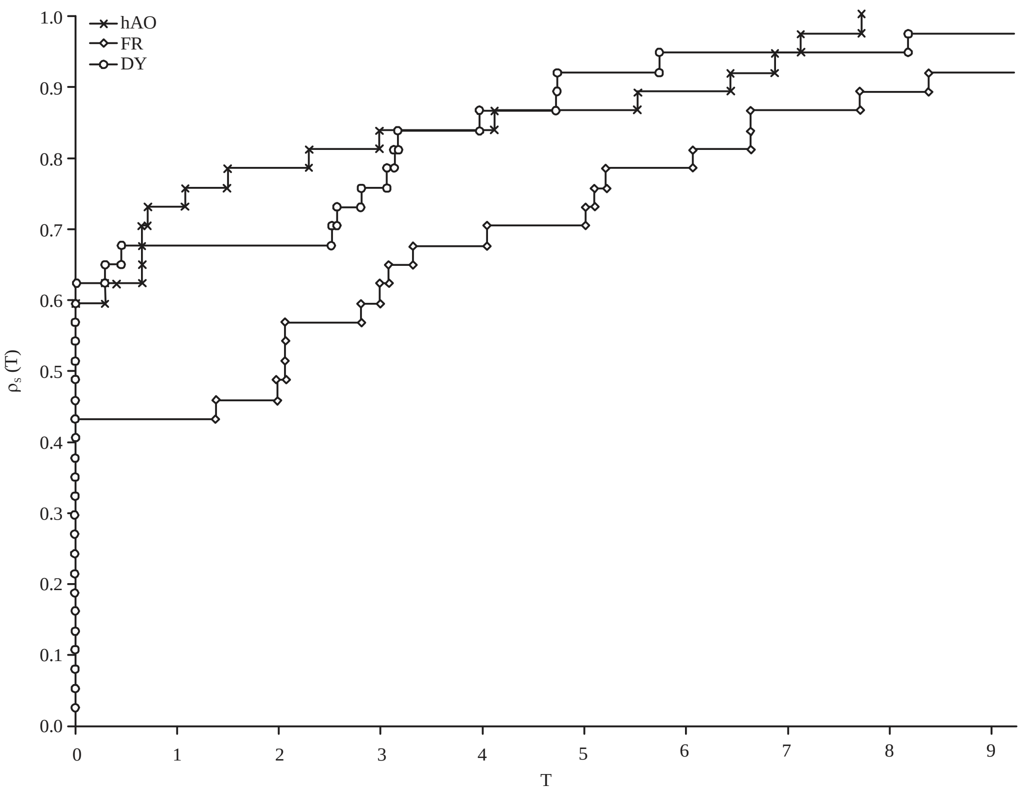

Fig. 3: Performance profile according to the value of the objective function

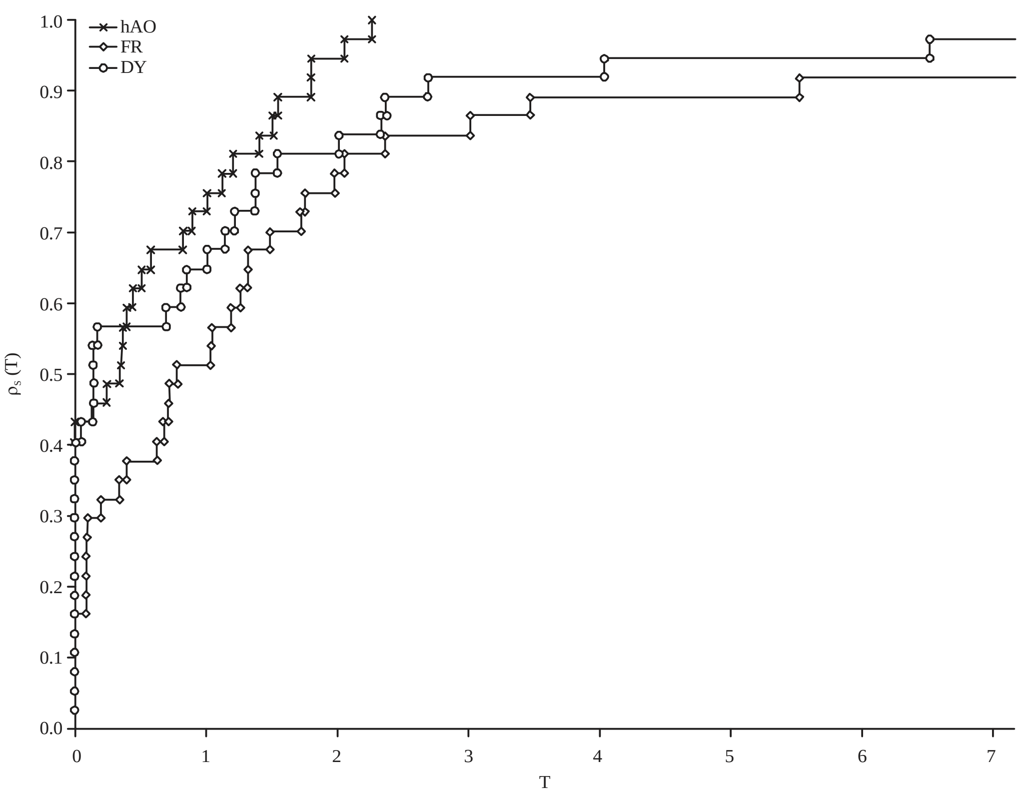

Fig. 4: Performance profile according to the value of the gradient norm

Table 1: Result from numerical experiments

Problem-n	Proposed hybrid method (hAO)	Fletcher-reeves (FR)	Dai-yuan (DY)	Hestenes-stiefel (HS)
Ext. Rosenbrock-500	9.2e-7/8.2e-14/75/0.232	9.4e-7/2.4e-14/161/0.486	6.1e-7/7.0e-16/66/0.195	7.5e-7/7.0e-16/81/0.230
Ext. Rosenbrock-1000	3.9e-7/6.9e-15/56/0.185	9.8e-7/4.7e-13/136/0.473	8.7e-7/1.4e-15/66/0.246	6.0e-7/4.5e-16/83/0.270
Ext. Rosenbrock-5000	9.3e-8/4.0e-15/70/0.525	3.5e-7/8.6e-15/107/0.890	6.8e-8/2.3e-15/68/0.524	7.6e-7/7.2e-16/85/0.527
Ext. Rosenbrock-10000	4.7e-7/1.3e-14/93/1.149	7.9e-7/3.1e-13/520/8.106	9.7e-8/4.6e-15/68/0.914	6.1e-7/4.6e-16/87/0.802
Diagonal-4-500	5.8e-8/3.8e-18/18/0.085	7.4e-8/1.0e-17/17/0.077	9.5e-7/1.1e-15/14/0.051	-/-/4/0.053
Diagonal-4-1000	8.2e-8/7.6e-18/18/0.095	1.0e-7/2.0e-17/17/0.131	6.2e-8/4.8e-18/15/0.090	-/-/4/0.047
Diagonal-4-5000	1.8e-7/3.8e-17/18/0.195	2.3e-7/1.0e-16/17/0.201	1.4e-7/2.4e-17/15/0.155	-/-/4/0.141
Diagonal-4-10000	2.6e-7/7.6e-17/18/0.243	3.3e-7/2.0e-16/17/0.374	2.0e-7/4.8e-17/15/0.221	-/-/4/0.180
Ext. Himelblau-500	8.3e-8/7.6e-17/63/0.239	9.2e-7/1.6e-14/52/0.178	4.3e-7/2.0e-15/51/0.183	-/-/8/0.079
Ext. Himelblau-1000	1.2e-7/1.5e-16/63/0.276	2.5e-7/1.2e-15/53/0.229	6.1e-7/4.0e-15/51/0.258	-/-/8/0.105
Ext. Himelblau-5000	2.5e-7/7.1e-16/63/0.753	5.7e-7/6.0e-15/53/0.529	6.5e-7/6.4e-15/52/0.566	6.1e-7/7.4e-15/61/0.610
Ext. Himelblau-10000	3.9e-7/1.7e-15/63/1.111	8.0e-7/1.2e-14/53/0.863	9.1e-7/1.2e-14/52/0.978	-/-/16/0.591
Quadratic QF1-2	2.3e-7/-2.5e-1/55/0.530	4.4e-7/-2.5e-1/29/0.093	1.3e-7/-2.5e-1/11/0.046	7.6e-7/-2.5e-1/69/0.236
Ext. Beale-500	4.8e-7/3.6e-14/55/0.530	7.5e-7/2.2e-14/93/1.018	1.8e-7/2.7e-14/242/2.817	-/-/5/0.116
Ext. Beale-1000	5.8e-7/4.0e-14/55/0.791	9.9e-7/1.3e-12/119/1.981	2.5e-7/5.5e-14/242/4.384	-/-/5/0.176
Ext. Beale-5000	7.9e-7/4.5e-14/69/3.808	5.5e-7/1.8e-13/102/6.722	5.7e-7/2.7e-13/242/17.849	-/-/20/3.571
Ext. Beale-10000	4.4e-7/2.8e-13/63/6.658	5.8e-7/3.5e-14/178/25.461	8.0e-7/5.5e-13/242/34.652	3.2e-7/1.2e-15/59/13.769
Mod. Ext. Beale-500	7.7e-7/1.5e+1/219/2.822	2.2e-7/1.5e+1/152/2.315	8.9e-7/1.5e+1/162/2.312	1.4e-7/1.5e+1/55/1.116
Mod. Ext. Beale -1000	8.7e-7/3.1e+1/199/3.953	6.0e-7/3.1e+1/146/3.461	2.5e-7/3.1e+1/407/9.042	-/-/54/1.688
Mod. Ext. Beale -5000	8.9e-7/1.5e+2/380/31.922	3.1e-7/1.5e+2/189/18.828	9.1e-7/1.5e+2/1841/216.002	6.8e-7/1.5e+2/115/11.459
Mod. Ext. Beale -10000	8.8e-7/3.1e+2/767/126.476	8.1e-7/3.1e+2/890/225.586	4.7e-7/3.1e+2/693/199.019	8.4e-7/3.1e+2/66/18.559
Ext. Block Diagonal -500	7.0e-7/3.1e-14/50/0.303	-/-/97/1.153	-/-/28/0.314	8.8e-7/3.8e-14/80/0.535
Ext. Block Diagonal -1000	1.5e-7/1.6e-15/46/0.345	-/-/303/5.548	9.7e-7/8.6e-14/78/0.634	-/-/76/0.639
Ext. Block Diagonal -5000	5.6e-7/5.8e-14/45/0.968	-/-/688/40.809	9.8e-7/4.7e-14/817/30.772	8.8e-7/2.6e-14/85/2.569
Ext Block Diagonal -10000	3.7e-7/4.3e-15/35/1.395	9.3e-7/4.3e-14/234/14.392	9.6e-7/3.6e-14/163/7.622	-/-/7/0.253
Gen. Tridiagonal-1-2	1.0e-8/5.7e-10/683/2.099	4.6e-7/6.6e-12/43/0.186	9.2e-7/3.8e-11/37/0.106	-/-/28/0.098
Gen. Rosenbrock-500	8.9e-7/4.0/63/0.431	9.5e-7/4.0/81/0.555	9.9e-7/4.0/49/0.342	2.5e-7/4.0/77/0.542
Gen. Rosenbrock-1000	8.9e-7/4.0/63/0.605	9.5e-7/4.0/81/0.857	9.9e-7/4.0/49/0.478	2.5e-7/4.0/77/0.748
Gen. Rosenbrock-5000	8.9e-7/4.0/63/2.235	9.5e-7/4.0/81/2.961	9.9e-7/4.0/49/1.815	2.5e-7/4.0/77/2.757
Gen. Rosenbrock-10000	8.9e-7/4.0/63/4.356	9.5e-7/4.0/81/5.764	9.9e-7/4.0/49/3.624	2.5e-7/4.0/77/5.400
Gen. White & Holst-2	7.9e-7/3.6e-13/78/0.333	3.9e-7/1.7e-13/108/0.453	7.9e-7/1.5e-12/281/1.168	3.3e-7/2.4e-13/87/0.483
Generalized PSC1-500	9.1e-7/5.0e+2/1386/6.168	8.3e-7/5.0e+2/1876/14.131	9.4e-7/5.0e+2/993/6.460	5.1e-7/5.0e+2/557/2.669
Generalized PSC1-1000	9.5e-7/1.0e+3/1570/9.248	8.6e-7/1.0e+3/4374/63.256	8.0e-7/1.0e+3/1216/12.194	9.7e-7/1.0e+3/721/4.829
Generalized PSC1-5000	9.4e-7/5.0e+3/2843/50.265	8.5e-7/5.0e+3/2709/127.930	7.4e-7/5.0e+3/3550/188.882	9.2e-7/5.0e+3/608/11.789
Ext. Tridiagonal-1-500	1.0e-6/3.6e-9/4337/18.802	8.0e-7/1.7e-9/49/0.311	2.4e-7/1.7e-11/269/1.166	-/-/25/0.228
Ext. Tridiagonal-1-1000	1.0e-6/4.6e-9/5467/31.015	9.6e-7/3.3e-9/49/0.436	3.4e-7/3.3e-11/269/1.517	-/-/28/0.322
Ext. Tridiagonal-1-5000	1.0e-6/7.8e-9/9353/160.967	4.6e-7/2.8e-9/93/2.596	7.5e-7/1.7e-10/269/4.667	-/-/14/0.718

CONCLUSION

Based on two previous works on the construction of hybrid CG algorithm, a new hybrid method was proposed in this paper. An interesting feature of the method is that it comprises of five different terms corresponding to five different CG methods. The descent and global convergence properties of the method were established under the standard Wolfe line search. Numerical tests with the five-terms hybrid method revealed that the method can really compete with the well-established methods. As part of future study, the method will be compared comprehensively with the WYL, MHS and MLS methods which are recently proposed. Doing this will afford us the opportunity to compare the performance of the method according to certain features.

SIGNIFICANCE STATEMENT

This study proposes a new conjugate gradient (CG) method as a hybrid of 5 existing methods thereby improving a method known to be efficient for large-scale unconstrained optimization problems. This study will help the researcher to uncover the descent property as well as the global convergence property of a new CG method. Thus, a new theory on this method may be arrived at and added to what is already known among CG analysts.

ACKNOWLEDGMENT

The authors wish to appreciate the Covenant University Centre for Research, Innovation and Discovery (CUCRID) for providing enabling facilities and funds to conduct this study.

REFERENCES

1. Agarana, M.C., E.A. Owoloko and O.J. Adeleke, 2016. Optimizing public transport systems in sub-Saharan Africa using operational research technique: A focus on Nigeria. Proc. Manuf., 7: 590-595.
2. Adewumi, A.O. and O.J. Adeleke, 2016. A survey of recent advances in vehicle routing problems. Int. J. Syst. Assurance Eng. Manage., 10.1007/s13198-016-0493-4.
3. Fletcher, R. and C.M. Reeves, 1964. Function minimization by conjugate gradients. Comput. J., 7: 149-154.
4. Hestenes, M.R. and E. Stiefel, 1952. Methods of conjugate gradients for solving linear systems. J. Res. Nat. Bureau Stand., 49: 409-435.
5. Polak, E. and G. Ribiere, 1969. Note sur la convergence de directions conjugees. Rev. Francaise Informat Recherche Operationelle, 16: 35-43.
6. Polyak, B.T., 1969. The conjugate gradient method in extremal problems. USSR Comput. Math. Math. Phys., 9: 94-112.
7. Liu, Y. and C. Storey, 1991. Efficient generalized conjugate gradient algorithms, part 1: Theory. J. Optimiz. Theory Applic., 69: 129-137.
8. Dai, Y.H. and Y. Yuan, 1999. A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optim., 10: 177-182.
9. Al-Baali, M., 1985. Descent property and global convergence of the Fletcher-Reeves method with inexact line search. IMA J. Numer. Anal., 5: 121-124.
10. Gilbert, J.C. and J. Nocedal, 1992. Global convergence properties of conjugate gradient methods for optimization. SIAM J. Optim., 2: 21-42.
11. Xu, X. and F. Kong, 2016. New hybrid conjugate gradient methods with the generalized Wolfe line search. SpringerPlus, Vol. 5. 10.1186/s40064-016-2522-9.
12. Kaelo, P., 2014. Another hybrid conjugate gradient method for unconstrained optimization problems. J. Nonlinear Anal. Optimiz., 5: 127-137.
13. Dai, Y.H. and Y. Yuan, 2001. An efficient hybrid conjugate gradient method for unconstrained optimization. Ann. Operat. Res., 103: 33-47.
14. Powell, M.J.D., 1977. Restart procedures for the conjugate gradient method. Math. Program., 2: 241-254.
15. Andrei, N., 2008. A hybrid conjugate gradient algorithm for unconstrained optimization as a convex combination of Hestenes-Stiefel and Dai-Yuan. Stud. Inform. Control, 17: 55-70.
16. Liu, J.K. and S.J. Li, 2014. New hybrid conjugate gradient method for unconstrained optimization. Applied Math. Comput., 245: 36-43.
17. Dai, Y.H. and L.Z. Liao, 2001. New conjugacy conditions and related nonlinear conjugate gradient methods. Applied Math. Optim., 43: 87-101.
18. Wei, Z., S. Yao and L. Liu, 2006. The convergence properties of some new conjugate gradient methods. Applied Math. Comput., 183: 1341-1350.
19. Shengwei, Y., Z. Wei and H. Huang, 2007. A note about Wyl's conjugate gradient method and its applications. Applied Math. Comput., 191: 381-388.
20. Li, X. and X. Zhao, 2011. A hybrid conjugate gradient method for optimization problems. Nat. Sci., 3: 85-90.
21. Jiang, X.Z., L. Han and J.B. Jian, 2012. A globally convergent mixed conjugate gradient method with Wolfe line search. Math. Numer. Sin., 34: 103-112.
22. Jian, J., L. Han and X. Jiang, 2015. A hybrid conjugate gradient method with descent property for unconstrained optimization. Applied Math. Model., 39: 1281-1290.
23. Zoutendijk, G., 1970. Nonlinear Programming, Computational Methods. In: Integer and Nonlinear Programming, Abadie, J. (Ed.). Elsevier, Amsterdam, ISBN-13: 978-07204 20364, pp: 37-86.
24. Andrei, N., 2008. An unconstrained optimization test functions collection. Adv. Mod. Optimiz., 10: 147-161.
25. Dolan, E.D. and J.J. More, 2002. Benchmarking optimization software with performance profiles. Math. Program., 91: 201-213.