## A Globally Convergent Hybrid FR-PRP Conjugate Gradient Method for Unconstrained Optimization Problems

OLAWALE J. ADELEKE<sup>1</sup>, IDOWU A. OSINUGA<sup>2</sup>, RAUFU A. RAJI<sup>3</sup>

<sup>1</sup>Department of Mathematical Sciences, Redeemer's University, Ede, NIGERIA <sup>2</sup>Department of Mathematics, Federal University of Agriculture, Abeokuta, NIGERIA <sup>3</sup>Department of Science and Social Sciences, Osun State Polytechnic, Ire, NIGERIA

*Abstract:* - In this paper, a new conjugate gradient (CG) parameter is proposed through the convex combination of the Fletcher-Reeves (FR) and Polak-Ribiére-Polyak (PRP) CG update parameters such that the conjugacy condition of Dai-Liao is satisfied. The computational efficiency of the PRP method and the convergence profile of the FR method motivated the choice of these two CG methods. The corresponding CG algorithm satisfies the sufficient descent property and was shown to be globally convergent under the strong Wolfe line search procedure. Numerical tests on selected benchmark test functions show that the algorithm is efficient and very competitive in comparison with some existing classical methods.

*Key-Words:* - Unconstrained optimization, convex combination, conjugate gradient method, conjugacy condition, hybridization.

Received: August 3, 2021. Revised: December 10, 2021. Accepted: December 21, 2021. Published: December 31, 2021.

### **1** Introduction

This study considers the nonlinear conjugate gradient (CG) method that is widely used to solve nonlinear unconstrained optimization problems of the form

$$\min \left[ f(x) \colon x \in \mathbb{R}^n \right] \tag{1}$$

where  $f: \mathbb{R}^n \to \mathbb{R}$  is a continuous and differentiable function which is bounded from below. The method generates a sequence of points  $x_k: k \ge 1$  starting with an initial guess, say  $x_0 \in \mathbb{R}^n$ , using the recurrence relation

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

where  $\alpha_k$  is a positive step-size usually obtained by a line search, and  $d_k$  is the search direction generated by the following rules

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \quad d_0 = -g_0, \tag{3}$$

where  $g_k = \nabla f(x_k)$  is the gradient of f at  $x_k$  so that  $g_0$  is the gradient at  $x_0$  and  $\beta_k$  is the update parameter for the method. Many variants of the method exist, each of which corresponds to a distinct value of  $\beta_k$ . Few of the well-known variants are:

Hestenes & Stiefel (HS): 
$$\beta_{k}^{HS} = \frac{g_{k+1}^{T} y_{k}}{d_{k}^{T} y_{k}},$$
  
Fletcher & Reeves (FR):  $\beta_{k}^{FR} = \frac{\|g_{k+1}\|^{2}}{\|g_{k}\|^{2}},$   
Daniel (D):  $\beta_{k}^{D} = \frac{g_{k+1}^{T} \nabla^{2} f(x_{k}) d_{k}}{d_{k}^{T} \nabla^{2} f(x_{k}) d_{k}},$   
Polak, Ribiere & Polyak (PRP):  $\beta_{k}^{PRP} = \frac{g_{k+1}^{T} y_{k}}{\|g_{k}\|^{2}},$   
Liu & Storey (LS):  $\beta_{k}^{LS} = \frac{g_{k+1}^{T} y_{k}}{-d_{k}^{T} g_{k}},$   
Fletcher (CD):  $\beta_{k}^{CD} = \frac{\|g_{k+1}\|^{2}}{-d_{k}^{T} g_{k}},$   
Dai & Yuan (DY):  $\beta_{k}^{DY} = \frac{\|g_{k+1}\|^{2}}{d_{k}^{T} y_{k}},$   
Hager & Zhang (N):  $\beta_{k}^{N} = \left(y_{k} - 2d_{k}\frac{\|y_{k}\|^{2}}{d_{k}^{T} y_{k}}\right)^{T} \frac{g_{k+1}}{d_{k}^{T} y_{k}},$ 

Bamigbola, Ali & Nwaeze (BAN):  $\beta_k^{BAN} = -\frac{g_{k+1}^T y_k}{g_k^T y_k}$ , where  $y_k = g_{k+1} - g_k$ ,  $\|\bullet\|$  is the Euclidean norm of

vectors and each parameter corresponding to the method in [20], [15], [13], [26 & 27], [16], [23], [11], [18] and [8], respectively.

All the variants mentioned above are equivalent when f is a strict convex quadratic function and  $\alpha_k$  is evaluated with an exact method. However, their behaviours vary for non-quadratic cost functions [22]. Experiences from the literature have also shown that these methods can be categorized into two classes: those with strong global convergence properties and those with good computational performances. For instance, in the first category are the FR and DY methods, while in the second category are the PRP, HS and LS methods.

Zouténdijk [35] reported the first global convergence result for the FR method when the line search is exact. However, Powell in [28] claimed that the FR method was susceptible to jamming when exact line search is used. This constitutes a major drawback in the computational efficiency of the method. On the other hand, it was shown in [26] that the (CG) method with  $\beta_k^{PRP}$  and exact line search is globally convergent when the objective function is convex. A counterexample to this was given in [29] to show that there exist convex functions for which the PRP method does not converge globally with exact line search. It was suggested in the same work that it is possible to obtain the global convergence result for the method by enforcing that the value of  $\beta_{i}^{PRP}$  to be positive.

The classical CG methods highlighted above possessed contrasting features based on convergence and computational performance, thereby giving rise to two classes of methods. In particular, those in the class possessing good convergence properties are usually prone to jamming, a situation whereby as the iteration continues, shorter steps without significant progress are taken toward the minimum point. This backdrop was. however, resolved in the class of computationally-performing methods due to their in-built restart feature [19]. To improve the efficiency of the CG method, studies on the hybridization of methods from these two classes have been carried out with a number of efficient

methods being the outcomes. Although the first work on hybrid CG methods can be traced to Touati-Ahmed-Storey [30], a brief overview of only recent hybrid methods is given in this section.

Mo et al. [25] propose two kinds of CG methods based on the modifications of the hybrid methods of Touati-Ahmed-Storey [30] and Dai-Yuan [12]. Under mild conditions, the methods were shown to be globally convergent. Babaie-Kafaki [4, 6] combined the features of HS and DY methods to suggest a hybrid method by performing a quadratic relaxation of the hybrid CG parameter proposed by Dai and Yuan [12]. The hybridization parameter, in this case, was obtained based on a conjugacy condition that is independent of the line search. The global convergence property of the method was established for uniformly convex functions. Yao and Qin [33] suggested the hybrid of DL (Dai and Liao [10]) and WYL (Wei el al. [31]) methods and showed that the method satisfied the sufficient descent condition under the Wolfe line search and globally convergent for general functions.

A growing idea for constructing hybrid methods is the use of linear and convex combinations. Based on the former, Xu and Kong [32] proposed a hybridization of DY and HS methods with nonnegative combination parameters in which at least one of them is not equal to zero. The former approach was adopted by Andrei [2] the same set of methods with the combination parameter computed such that the search direction is the Newton direction and the secant equation also satisfied. Based on this approach, Babaie-Kafaki and Mahdavi-Amiri [5] constructed a hybrid method from same HS and DY methods which is globally convergent for uniformly convex functions. The method yielded global convergence result for general functions when Powell's nonnegative restriction on the CG parameter was enforced. Further convex combination of these two classical methods led Babaie-Kafaki and Ghanbari [7] to propose a hybrid method whereby the hybridization parameter was computed from a modified secant equation obtained based on the search direction of the Hager–Zhang nonlinear CG method [18].

Recently, Alhawarat et al. [1] proposed a new hybrid CG method using the positive PRP and NPRP method proposed by Zhang [34] whose properties were studied under the strong Wolfe line search and mild assumptions. Li and Yang [21] also propose a nonmonotone hybrid conjugate gradient method which combines the nonmonotone line search technique with the LY method (See Lu et al. [24]). The approach was based on the idea that nonmonotone algorithmic frameworks may accept a larger value of step-length in the quest to improve the behaviour of the LR method. Interested readers may also consult [36-38] for other recently proposed hybrid CG methods.

In this present work, the computational efficiency and good convergence property of the PRP and FR methods, respectively, motivate the construction of a new hybrid CG parameter, based on the idea of convex combination of CG parameters introduced in Andrei [2] and further utilized in [22]. The proposed  $\beta_k$  is computed as a convex combination of  $\beta_k^{FR}$ and  $\beta_k^{PRP}$ , that is

$$\beta_{k} = (1 - \varphi_{k}) \beta_{k}^{PRP} + \varphi_{k} \beta_{k}^{FR}$$

$$\tag{4}$$

where  $\beta_k^{PRP}$  and  $\beta_k^{FR}$  are as given above and  $\varphi_k \in (0,1)$ . An interesting feature of this method is that the search direction satisfies the Dai-Liao conjugacy condition [10]

$$d_{k+1}^{T} y_{k} = -t s_{k}^{T} g_{k+1}, \quad t \ge 0,$$
(5)

where  $s_k = x_{k+1} - x_k$ . This condition, unlike the traditional conjugacy condition,  $d_{k+1}^T y_k = 0$ , depends on approximate (inexact) line search. In this work, we assume the step-size  $\alpha_k$  satisfies the strong Wolfe line search criterion.

The remaining parts of this paper consist of the following sections. In Section 2, the new hybrid method is developed, and the corresponding conjugate gradient algorithm is outlined. The search direction,  $d_k$ , is shown to be a descent direction in Section 3 while Section 4 details the global convergence result of our CG algorithm. Numerical results obtained through computational experiments are presented and discussed in Section 5. The concluding remarks are given Section 6.

### 2 The New Hybrid Method

To construct the value of  $\beta_k$  for the proposed method and as stated above, the parameter  $\varphi_k$  in Eq. (4) is chosen so that Eq. (5) is satisfied for every member of the sequence  $\{d_k\}$ . Thus, by taking the inner product of Eq. (3) with the vector  $y_k$ , we obtain

$$d_{k+1}^{T} y_{k} = -g_{k+1}^{T} y_{k} + \beta_{k} d_{k}^{T} y_{k}$$
(6)

Computing the value of  $\beta_k$  in Eq. (6) according to Eq. (4) gives

$$d_{k+1}^{T} y_{k} = -g_{k+1}^{T} y_{k} + (1 - \varphi_{k}) \beta_{k}^{PRP} d_{k}^{T} y_{k} + \varphi_{k} \beta_{k}^{FR} d_{k}^{T} y_{k}$$
(7)

Inserting the values of  $\beta_k^{PRP}$  and  $\beta_k^{FR}$  in Eq. (7), imposing condition (5) and after some algebra, we have

$$\varphi_{k}^{HFP} = \frac{g_{k+1}^{T} y_{k} \left( \left\| g_{k} \right\|^{2} - d_{k}^{T} y_{k} \right) - t s_{k}^{T} g_{k+1} \cdot \left\| g_{k} \right\|^{2}}{\left( \left\| g_{k+1} \right\|^{2} - g_{k+1}^{T} y_{k} \right) d_{k}^{T} y_{k}}.$$
 (8)

By substituting Eq. (8) into Eq. (4) and after some simplifications, we obtain the new  $\beta_k$  as

$$\beta_{k}^{HFP} = \frac{g_{k+1}^{T} y_{k}}{\left\|g_{k}\right\|^{2}} + \left(\frac{g_{k+1}^{T} y_{k} \cdot \left\|g_{k}\right\|^{2} - g_{k+1}^{T} y_{k} \cdot d_{k}^{T} y_{k}}{\left(\left\|g_{k+1}\right\|^{2} - g_{k+1}^{T} y_{k}\right) d_{k}^{T} y_{k}}\right)$$
$$\frac{-t s_{k}^{T} g_{k+1} \cdot \left\|g_{k}\right\|^{2}}{\left(\left\|g_{k+1}\right\|^{2} - g_{k+1}^{T} y_{k}\right) d_{k}^{T} y_{k}}\right) \cdot \frac{g_{k+1}^{T} g_{k}}{\left\|g_{k}\right\|^{2}}.$$
(9)

It is easy to note from Eq. (4) that  $\beta_k^{HFP} = \beta_k^{PRP}$ whenever  $\varphi_k = 0$  and  $\beta_k^{HFP} = \beta_k^{FR}$  whenever  $\varphi_k = 1$ . In what follows, we give a corresponding algorithm for analysing and implementing the new hybrid method. Note that the strong Wolfe line search criteria for estimating  $\alpha_k$  are given by

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k$$
(10)  
and

$$\left|g_{k+1}^{T}d_{k}\right| \leq -\sigma g_{k}^{T}d_{k} \tag{11}$$

### Algorithm 1: Hybrid CG Procedure

1: input  $x_0 \in \mathbb{R}^n$ ,  $\varepsilon \ge 0$ , set  $d_0 = -g_0$ , k = 0; 2: while  $||g_k|| < \varepsilon$  do 3:  $y_{k+1} = g_{k+1} + g_k$ ,  $s_{k+1} = x_{k+1} + x_k$ ; 4: if  $(||g_{k+1}||^2 - g_{k+1}^T y_k) d_k^T y_k = 0$  then  $\varphi_k = 0$ ; 5: else  $\varphi_k = \varphi_k^{HFP}$ ; 6: if  $0 < \varphi_k < 1$  then  $\beta_k^{HFP} = (1 - \varphi_k) \beta_k^{PRP} + \varphi_k \beta_k^{FR}$ ; 7: else if  $\varphi_k \ge 1$  then  $\beta_k^{HFP} = \beta_k^{FR}$ ; 8: else if  $\varphi_k \le 0$  then  $\beta_k^{HFP} = \beta_k^{PRP}$ ; 9:  $d_{k+1} = -g_{k+1} + \beta_k^{PRP} d_k$ ; 10: obtain  $\alpha_k$  using (10) and (11); 11: evaluate  $x_{k+1} = x_k + \alpha_k d_k$ ; 12: k = k + 1; 13: end

## **3** The Descent Property of Algorithm **1**

In this section, the search direction  $d_k$  generated by Algorithm 1 is shown to satisfy the sufficient descent property.

**Theorem 1:** Suppose the sequences  $\{d_k\}$  and  $\{g_k\}$  generated by Algorithm 1, then the search direction  $d_k$  satisfies the sufficient descent condition

$$g_k^T d_k \le -c \left\| g_k \right\|^2 \quad \forall \ k \ge 0, \ c \le 1.$$
(12)

**<u>Proof:</u>** If we consider the restart criterion of Powell [18] given by

$$\left|g_{k+1}^{T}g_{k}\right| > \frac{1}{5}\left\|g_{k+1}\right\|^{2},$$
(13)

then, it is obvious from Algorithm 1 that (13) holds. However, if (13) does not hold, i.e.,  $\left|g_{k+1}^{T}g_{k}\right| \leq \frac{1}{5}\left\|g_{k+1}\right\|^{2}$ , the following results are established by induction.

Suppose k = 0, then from Line 1 of Algorithm 1 together with the fact that  $c \le 1$ , we obtain  $g_0^T d_0 = -||g_0||^2$ , which shows that (12) is satisfied.

Now, we assume (12) hold for some  $k \ge 1$ . Using (11) in the expression  $d_k^T y_k = d_k^T g_{k+1} - d_k^T g_k$ , we obtain by virtue of the fact that  $\sigma < 1$ 

$$d_k^T y_k \ge -(1-\sigma) d_k^T g_k.$$
<sup>(14)</sup>

By taking the inner product of (3) with  $g_{k+1}$ , we obtain

$$d_{k+1}^{T}g_{k+1} = -\|g_{k+1}\|^{2} + \beta_{k}d_{k}^{T}g_{k+1}.$$
 (15)

According to Line 11-12 of Algorithm 1, if  $\varphi_k \ge 1$ , then  $\beta_k^{HFP} = \beta_k^{FR}$ . It follows that

$$\begin{aligned} d_{k+1}^{T} g_{k+1} &= - \left\| g_{k+1} \right\|^{2} + \frac{\left\| g_{k+1} \right\|^{2}}{\left\| g_{k} \right\|^{2}} d_{k}^{T} g_{k+1} \\ &\leq - \left\| g_{k+1} \right\|^{2} + \frac{\left\| g_{k+1} \right\|^{2}}{\left\| g_{k} \right\|^{2}} \cdot \left| d_{k}^{T} g_{k+1} \right| \\ &\leq - \left\| g_{k+1} \right\|^{2} + \frac{\left\| g_{k+1} \right\|^{2}}{\left\| g_{k} \right\|^{2}} \cdot -\sigma d_{k}^{T} g_{k} \\ &= - \left\| g_{k+1} \right\|^{2} - \frac{\sigma d_{k}^{T} g_{k}}{\left\| g_{k} \right\|^{2}} \cdot \left\| g_{k+1} \right\|^{2} \\ &= - \left( 1 + \frac{\sigma d_{k}^{T} g_{k}}{\left\| g_{k} \right\|^{2}} \right) \cdot \left\| g_{k+1} \right\|^{2} \\ &\leq - \left( 1 + \frac{\sigma \left| d_{k}^{T} g_{k} \right|}{\left\| g_{k} \right\|^{2}} \right) \cdot \left\| g_{k+1} \right\|^{2} = - (1 + \sigma d) \left\| g_{k+1} \right\|^{2} \quad (16) \end{aligned}$$

by setting  $|d_{k}^{T}g_{k}|/||g_{k}||^{2}$  to d, where d is a constant and satisfies the inequality  $0 \le d < 1$ . When  $\varphi_{k} \le 0$  (Line 8 of Algorithm 1), then  $\beta_{k}^{HFP} = \beta_{k}^{PRP}$  and we obtain the following:  $d_{k+1}^{T}g_{k+1} = -||g_{k+1}||^{2} + \beta_{k}^{PRP}d_{k}^{T}g_{k+1}$  $= -||g_{k+1}||^{2} + \frac{g_{k+1}^{T}y_{k}}{||g_{k}||^{2}}d_{k}^{T}g_{k+1}$  $\le -||g_{k+1}||^{2} + \frac{g_{k+1}^{T}(g_{k+1} - g_{k})}{||g_{k}||^{2}}|d_{k}^{T}g_{k+1}|$  $\le -||g_{k+1}||^{2} + \frac{(||g_{k+1}||^{2} - g_{k+1}^{T}g_{k})}{||g_{k}||^{2}} \cdot -\sigma d_{k}^{T}g_{k}$ 

$$\leq -\|g_{k+1}\|^{2} + \frac{\left(\|g_{k+1}\|^{2} + 0.2\|g_{k+1}\|^{2}\right)}{\|g_{k}\|^{2}} \cdot -\sigma d_{k}^{T} g_{k}$$

$$= -\|g_{k+1}\|^{2} + \frac{1.2\|g_{k+1}\|^{2}}{\|g_{k}\|^{2}} \cdot -\sigma d_{k}^{T} g_{k}$$

$$= \left(-1 - 1.2\frac{\sigma d_{k}^{T} g_{k}}{\|g_{k}\|^{2}}\right) \|g_{k+1}\|^{2} \leq \left(-1 - 1.2\frac{\sigma |d_{k}^{T} g_{k}|}{\|g_{k}\|^{2}}\right) \|g_{k+1}\|^{2}$$

$$= \left(-1 - 1.2\sigma d\right) \|g_{k+1}\|^{2} = -\left(1 + 1.2\sigma d\right) \|g_{k+1}\|^{2}. \quad (17)$$

When  $\beta_k$  is computed according to Eq. (4) (that is, Line 6 of Algorithm 1), the following result holds:

$$\begin{aligned} d_{k+1}^{T}g_{k+1} &= -\|g_{k+1}\|^{2} + \beta_{k}d_{k}^{T}g_{k+1} \\ &\leq -\|g_{k+1}\|^{2} + |\beta_{k}^{FR}||d_{k}^{T}g_{k+1}| + |\beta_{k}^{PRP}||d_{k}^{T}g_{k+1}| \\ &\leq -\|g_{k+1}\|^{2} + \sigma|\beta_{k}^{FR}||d_{k}^{T}g_{k}| + \sigma|\beta_{k}^{PRP}||d_{k}^{T}g_{k}| \\ &= -\|g_{k+1}\|^{2} + \sigma\frac{\|g_{k+1}\|^{2}}{\|g_{k}\|^{2}} \cdot |d_{k}^{T}g_{k}| + \sigma\frac{|g_{k+1}^{T}y_{k+1}|}{\|g_{k}\|^{2}} \cdot |d_{k}^{T}g_{k}| \\ &\leq -\|g_{k+1}\|^{2} + \sigma\frac{|d_{k}^{T}g_{k}|}{\|g_{k}\|^{2}} \cdot (2.2\|g_{k+1}\|^{2}) \\ &\leq -\|g_{k+1}\|^{2} + \sigma d(2.2\|g_{k+1}\|^{2}) = -(1-2.2\sigma d)\|g_{k+1}\|^{2}. (18) \end{aligned}$$

Inequalities (16), (17) and (18) show that (12) holds for k+1 and hence the proof is completed.

# 4 The Global Convergence of Algorithm 1

In order to establish the global convergence result for Algorithm 1, we first state the following necessary and general assumptions on the objective function f(x).

Assumption 1: (Lipschitz Condition) In some neighbourhood N of the level set  $S = \{x \in \mathbb{R}^n \mid f(x) \le f(x_0)\}$ , the gradient, g(x), of f(x) is Lipschitz continuous, that is, there exists a constant  $L < \infty$  such that  $\|g(x) - g(y)\| \le L \|x - y\| \quad \forall x, y \in N.$  (19)

Assumption 2: (Boundedness) The level set S is bounded. That is, there is a positive constant  $m < \infty$ such that  $||x|| \le m \quad \forall x \in S$ . Assumptions 1 and 2 together implies that for any such positive constant m,

$$\left\|g\left(x\right)\right\| \le m \quad \forall \ x \in N.$$
<sup>(20)</sup>

Assumption 3: (Zoutendijk Condition) Let an iterative scheme of the form (2) where  $d_k$  is a descent direction and  $\alpha_k$  satisfies the Wolfe line search conditions

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k, \quad g_{k+1}^T d_k \ge \sigma g_k^T d_k, \quad (21)$$

where  $0 < \delta \le \sigma < 1$ . If Assumptions 1 and 2 hold, then,

$$\sum_{k=0}^{\infty} \frac{\left(g_{k}^{T} d_{k}\right)^{2}}{\left\|d_{k}\right\|^{2}} < +\infty.$$
(22)

Obviously if  $\alpha_k$  satisfies the strong Wolfe conditions (10) and (11) as suggested in our algorithm above, then it must satisfy the weak Wolfe line search conditions. Hence, Assumption 3 also holds under strong Wolfe conditions.

From the definition of  $d_k^T y_k$  and (2), it can be shown that  $\alpha_k \ge \frac{-(1-\sigma)}{L} \frac{d_k^T g_k}{\|d_k\|^2}$ . Thus, from Theorem 1 it becomes obvious, if we assume  $g_k \ne 0$ , that  $\alpha_k$  is not zero. Hence, there must exist

 $g_k \neq 0$ , that  $\alpha_k$  is not zero. Hence, there must exist a constant  $\gamma > 0$  such that  $\alpha_k \ge \gamma$ , for all  $k \ge 0$ . This leads to the following global convergence results for the proposed algorithm.

**Theorem 2:** Suppose that Assumptions 1-3 hold. Consider a conjugate gradient method of the form (2)-(3), where  $\beta_k$  is estimated by (4), and where  $\alpha_k$  satisfies (10) and (11), with  $0 \le \delta < \sigma < 1$ . Then,

$$\lim_{k \to \infty} \inf \left\| g_k \right\| = 0.$$
<sup>(23)</sup>

**<u>Proof:</u>** If  $\liminf_{k \to \infty} ||g_k|| \neq 0$ , then there exists a constant b > 0 such that  $||g_k|| \ge b$ , (24)

for sufficiently large k. Starting with (4) we have

$$\begin{aligned} \left|\beta_{k}\right| &\leq \left|\beta_{k}^{PRP}\right| + \left|\beta_{k}^{FR}\right| = \frac{\left|g_{k+1}^{T}y_{k}\right|}{\left\|g_{k}\right\|^{2}} + \frac{\left\|g_{k+1}\right\|^{2}}{\left\|g_{k}\right\|^{2}} \\ &= \frac{\left|g_{k+1}^{T}g_{k+1} - g_{k+1}^{T}g_{k}\right|}{\left\|g_{k}\right\|^{2}} + \frac{\left\|g_{k+1}\right\|^{2}}{\left\|g_{k}\right\|^{2}} \\ &\leq \frac{\left\|g_{k+1}\right\|^{2}}{\left\|g_{k}\right\|^{2}} - \frac{\left|g_{k+1}^{T}g_{k}\right|}{\left\|g_{k}\right\|^{2}} + \frac{\left\|g_{k+1}\right\|^{2}}{\left\|g_{k}\right\|^{2}} = 2\frac{\left\|g_{k+1}\right\|^{2}}{\left\|g_{k}\right\|^{2}} - \frac{\left|g_{k+1}^{T}g_{k}\right|}{\left\|g_{k}\right\|^{2}} \\ &\leq 2\frac{\left\|g_{k+1}\right\|^{2}}{\left\|g_{k}\right\|^{2}} \leq \frac{2m^{2}}{b^{2}} = J. \end{aligned}$$

$$(25)$$

The first inequality in (25) was established from the fact that  $\varphi_k \in (0,1)$ , the first equality follows from the definitions of  $\beta_k^{PRP}$  and  $\beta_k^{FR}$ , the first part of the second equality uses the definition of  $y_k$ , the second equality applies the Cauchy Schwartz inequality, while the third inequality uses (20) and (24).

In a similar version, the result in (25) may also be established for  $\varphi_k \notin (0,1)$ , i.e., if either  $\varphi_k \ge 1$  or  $\varphi_k \leq 1$  for all  $k \geq 1$ , in accordance to Step 5 of the algorithm in section 2.

Now from (3), we have

$$\|d_{k+1}\| \le \|g_{k+1}\| + |\beta_k| \|d_k\| = \|g_{k+1}\| + |\beta_k| \frac{\|x_{k+1} - x_k\|}{\alpha}$$
$$= \|g_{k+1}\| + |\beta_k| \frac{\|s_k\|}{\alpha} \le m + \frac{JM}{\gamma} = D,$$
(26)

п

where  $M = \max \{ ||x - y|| | x, y \in N \}$  is the diameter of N. This result implies

$$\sum_{k\ge 0} \frac{1}{\|d_k\|^2} = +\infty.$$
 (27)

If we multiply the left-hand side of (27) by the constant  $c^2b^2$ , we obtain

$$\sum_{k\geq 0} \frac{1}{\|d_k\|^2} \cdot c^2 b^2 \leq \underbrace{\sum_{k\geq 0} \frac{c^2 \|g_k\|^4}{\|d_k\|^2}}_{\text{From (24)}} \leq \underbrace{\sum_{k\geq 0} \frac{\left(d_k^T g_k\right)^2}{\|d_k\|^2}}_{\text{From (12)}} < +\infty \quad (28)$$

Inequality (28) contradicts (27) which invariably means  $||g_k|| < b$ , and thus (23) is established.

## 5 The Numerical Implementation of Algorithm 1

By the above analysis, the main questions we want to investigate is whether the proposed method outperforms some classical non-hybrid methods especially those of FR and PRP since the proposed method in this paper can be reduced to any of these methods.

The codes were written and run on Matlab R2016a. All the tests were performed on a PC with an installed memory of 2GB and 64-bits operating system. The experiments were performed on a set of 56 nonlinear unconstrained problems retrieved from the collections of Andrei [3], some of which may also be found in the CUTE library [9]. The dimension of the problem is taken as 500, 1000, 5000 and 10000. However, due to the peculiarity of some problems, we used much-reduced dimensions so as to ensure the problems were solvable by all the methods.

The new hybrid method was compared with other existing methods based on four attributes: number of iterations, CPU time, function and gradient evaluations. The Dolan and Moré [14] performance profiling technique was used to assess the approximate performance of the different methods. They introduced the notion of performance profile as a means to evaluate and compare the performance of the set of solvers S on a set of problem P. Suppose there exist  $n_s$  solvers and  $n_p$  problems, for each problem p and solver s, they defined:  $r_{p.s}$  = computing cost (number of iterations, CPU time, function or gradient evaluations) required to solve problem p by solver s. If  $I_{p,s}$  is an attribute (say, number of iterations or function evaluation, etc) for problem  $p \in P$  by method  $s \in S$ , then the comparison between the different methods is based on the performance ratio given by

$$r_{p,s} = \frac{\mathbf{I}_{p,s}}{\min\left\{\mathbf{I}_{p,s} : s \in S\right\}}.$$
(29)

Equation (29) gives the required number of iterations for solving problem  $p \in P$  with method  $s \in S$ . Readers may refer to Dolan and Moré [14]

for further guides on the application of the technique.

All algorithms terminate when  $||g_k|| \le 10^{-6}$ . No restriction was placed on the number of iterations. The strong Wolfe line searches (10) and (11) were imposed, with parameters  $\delta = 0.0001$ ,  $\sigma = 0.9$ , t = 1.

The following Figures 1-4 were generated in accordance with the performance profile theory of [14]. As earlier mentioned, the performance of the proposed hybrid method was checked against five other traditional methods based on four attributes. Particularly, the proposed HFP method found solutions to all the selected test functions.

Observe from Figure 1 that HFP competes very well with the five traditional methods. At  $\tau = 1$ , HFP is only behind DY and FR. This shows that the rate of convergence of the proposed method is good.



**Fig.1:** The performance profile of HFP method against FR, PRP, LS, DY and CG\_DESCENT methods according to the CPU times.



**Fig.2:** The performance profile of HFP method against FR, PRP, LS, DY and CG\_DESCENT methods according to the number of iterations.



**Fig.3:** The performance profile of HFP method against FR, PRP, LS, DY and CG\_DESCENT methods according to the values of the objective functions.



**Fig.4:** The performance profile of HFP method against FR, PRP, LS, DY and CG\_DESCENT methods according to the values of the norm of the gradient of the objective functions.

With respect to the numbers of iterations reported for the test functions (Figure 2), HFP produces a better performance than the other methods except DY at  $\tau = 1$ . In order to get a clearer information on the performance of the methods with respect to the numerical values of the objective functions and their norms, the authors decided to adjust the scales of the performance profile graphs (see Figures 3 and 4). The figures reveal that the values generated for HFP are in consonance with those of the compared traditional methods.

#### 6 Conclusion

This paper introduces a new and efficient hybrid conjugate gradient parameter through the convex combination of the conjugate gradient parameters of the traditional FR and PRP methods. The proposed method guarantees descent direction and is globally convergent under the strong Wolfe line search technique. Numerical results for mixed large and small-scale problems show that the method is very efficient in terms of convergence and the number of problems it solved when compared to some classical methods. For other attributes such as the function and gradient values, the method is competitive and promising. As part of future work, this method would be tested on more problems and compared with other known efficient hybrid methods. It is also possible to test the method for different specific values of the parameter  $\varphi_k$ .

#### References:

- Alhawarat, A., Mamat, M., Rivaie, M. and Z. Salleh, An Efficient Hybrid Conjugate Gradient Method with the Strong Wolfe-Powell Line Search, *Mathematical Problems in Engineering*, 2015.
- [2] Andrei, N., 2008. A hybrid conjugate gradient algorithm for unconstrained optimization as a convex combination of Hestenes-Stiefel and Dai-Yuan, *Studies in Informatics and Control*, Vol.17, 2008, pp. 55-70.
- [3] Andrei, N., An unconstrained optimization test functions collection, *Adv. Mod. and Optim.*, Vol.10, No.1, 2008, pp. 147-161.
- [4] Babaie-Kafaki, S., A hybrid conjugate gradient method based on a quadratic relaxation of the Dai-Yuan hybrid conjugate gradient parameter, *Optimization*, Vol.62, 2013, pp. 929-941.
- [5] Babaie-Kafaki, S. and N. Mahdavi-Amiri, Two Modified Hybrid Conjugate Gradient Methods Based on a Hybrid Secant Equation, *Mathematical Modelling and Analysis*, Vol.18, 2013, pp. 32-52.
- [6] Babaie-Kafaki, S., Addendum to: A hybrid conjugate gradient method based on a quadratic relaxation of Dai-Yuan hybrid conjugate gradient parameter. *Optimization*, Vol.63, 2014, pp. 657-659.
- [7] Babaie-Kafaki, S. and R. Ghanbari, Two hybrid nonlinear conjugate gradient methods based on a modified secant equation. *Optimization*, Vol.63, 2014, pp. 1027-1042.
- [8] Bamigbola, O.M., Ali, M. and E. Nwaeze, An efficient and convergent method for unconstrained nonlinear optimization. In *Proceedings of International Congress of Mathematicians*. Hyderabad, India, 2010.
- [9] Bongartz, I., Conn, A. R., Gould, N. I. M. and P. L. Toint, CUTE: constrained and unconstrained testing environments, *ACM*

*Trans. Math. Software*, Vol.21, 1995, pp. 123-160.

- [10] Dai, Y.H. and L. Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, *Appl. Math. Optim.*, Vol.43, 2001, pp. 87-101.
- [11] Dai, Y.H. and Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J. Optim.*, Vol.10, 1999, pp. 177-182.
- [12] Dai, Y.H. and Y. Yuan, An Efficient Hybrid Conjugate Gradient Method for Unconstrained Optimization, *Annals of Operations Research*, Vol.103, 2001, pp. 33-47.
- [13] Daniel, J. W., The conjugate gradient method for linear and nonlinear operator equations, *SIAM J. Numer. Anal.*, Vol.4, 1967, pp. 10-26.
- [14] Dolan, E. D. and J. J. Moré, Benchmarking optimization software with performance profiles, *Mathematical Programming*, Vol.91, 2002, pp. 201-213.
- [15] Fletcher, R. and C. Reeves, Function minimization by conjugate gradients, *Comput.* J., Vol.7, 1964, pp. 149-154.
- [16] Fletcher, R., Practical Methods of Optimization, John Wiley & Sons, New York. 1987.
- [17] Gilbert, J.C. and J. Nocedal, Global convergence properties for conjugate gradient methods for optimization, *SIAM J. Optim.*, Vol.2, 1992, pp. 21-42.
- [18] Hager, W.W. and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, *SIAM J. Optim.*, Vol.16, No.1, 2005, pp. 170-192.
- [19] Hager, W. W. and H. Zhang, A survey of nonlinear conjugate gradient methods, *Pacific Journal of Optimization*, Vol.2, 2006, pp. 35-58.
- [20] Hestenes, M.R. and E.L. Stiefel, 1952. Method of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Standards*, Vol.49, 1952, pp. 409-436.
- [21] Li, W. Y. and Y. T. Yang, A nonmonotone hybrid conjugate gradient method for unconstrained optimization. *Journal of Inequalities and Applications*, 2015.
- [22] Liu, J.K. and S.J. Li, New hybrid conjugate gradient method for unconstrained optimization, *Appl. Math. Comput.*, Vol.245, 2014, pp. 36-43.
- [23] Liu, Y. and C. Storey, Efficient generalized conjugate gradient algorithms, Part 1: Theory, *J. Optim. Theory Appl.*, Vol.69, 1991, pp. 129-137.

- [24] Lu, Y. L., Li, W. Y., Zhang, C. M. and Y. T. Yang, A class new conjugate hybrid gradient method for unconstrained optimization, *J. Inf. Comput. Sci.*, Vol.12, No.5, 2015, pp. 1941-1949.
- [25] Mo, J. T., Gu, N. Z. and Z. X. Wei, Hybrid conjugate gradient methods for unconstrained optimization, *Optimization Methods & Software*, Vol.22, 2007, pp. 297-307.
- [26] Polak, E. and G. Ribieré, Note sur la convergence de directions conjugeés, *Rev. Francaise Informat Recherche Operationelle*, *3e Année*, Vol.16, 1969, pp. 35-43.
- [27] Polyak, B.T., The conjugate gradient method in extreme problems, USSR Comp. Math. Math. Phys., Vol.9, 1969, pp. 94-112.
- [28] Powell, M. J. D., Restart procedures of the conjugate gradient method, *Mathematical Programming*, Vol.2, 1977, pp. 241-254.
- [29] Powell, M. J. D., Nonconvex minimization calculations and the conjugate gradient method, *Numerical Analysis (Dundee, 1983), Lecture Notes in Mathematics*, Vol.1066, 1984, pp. 122-141.
- [30] Touati-Ahmed, D. and C. Storey, Efficient hybrid conjugate gradient techniques. *J. Optim. Theory Appl.*, Vol.64, 1990, pp. 379-397.
- [31] Wei, Z., Yao, S. and L. Liu, The convergence properties of some new conjugate gradient methods, *Appl. Math. Comput.*, Vol.183, 2006, pp. 1341-1350.
- [32] Xu, X. and F. Kong, New hybrid conjugate gradient methods with the generalized Wolfe line search, *Springerplus*, Vol.5, 2016, pp. 881.
- [33] Yao, S. W. and B. Qin, A Hybrid of DL and WYL Nonlinear Conjugate Gradient Methods. *Abstract and Applied Analysis*, 2014.
- [34] Zhang, L., An improved Wei-Yao-Liu nonlinear conjugate gradient method for optimization computation, *Applied Mathematics and Computation*, Vol.215, No.6, 2009, pp. 2269–2274.
- [35] Zoutendijk, G., Nonlinear programming, Computational methods, in Integer and Nonlinear Programming, J. Abadie, ed., North-Holland, Amsterdam, 1970, pp. 37-86.
- [36] Adeleke, O. J. and I. A. Osinuga, A Five-term Hybrid Conjugate Gradient Method with Global Convergence and Descent Properties for Unconstrained Optimization Problems, *Asian Journal of Scientific Research*, Vol.11, 2018, pp. 185-194.
- [37] Oladepo, D.A., Adeleke, O.J. and C.T. Ako, A Mixed Hybrid Conjugate Gradient Method for Unconstrained Engineering Optimization

Problems. In: Silhavy R. (eds) Cybernetics and Algorithms in Intelligent Systems. CSOC2018. Advances in Intelligent Systems and Computing, Vol.765. 2019, Springer, Cham.

[38] Adeleke, O. J., Ezugwu, A. E.-S., and I. A. Osinuga, A New Family of Hybrid Conjugate Gradient Methods for Unconstrained Optimization. *Statistics, Optimization & Information Computing*, 2020 https://doi.org/10.19139/soic-2310-5070-480

### **Creative Commons Attribution License 4.0** (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0 <u>https://creativecommons.org/licenses/by/4.0/deed.en\_US</u>