An International Journal of Optimization and Control: Theories & Applications
ISSN:2146-0957 eISSN:2146-5703
Vol.7, No.2, pp.177-185 (2017)
http://doi.org/10.11121/ijocta.01.2017.00339

RESEARCH ARTICLE

A modified quadratic hybridization of Polak-Ribière-Polyak and Fletcher-Reeves conjugate gradient method for unconstrained optimization problems

Sindhu Narayanan, P. Kaelo* and M.V. Thuto

Department of Mathematics, University of Botswana, Private Bag UB00704, Gaborone, Botswana
sindhu.op@gmail.com, kaelop@mopipi.ub.bw, thutomv@mopipi.ub.bw

ARTICLE INFO

Article History:
Received 15 December 2016
Accepted 23 February 2017
Available 15 July 2017

Keywords:
Hybridization
Conjugate gradient
Wolfe line search conditions
Global convergence

AMS Classification 2010:
90C06, 90C30, 65K05

ABSTRACT

This article presents a modified quadratic hybridization of the Polak–Ribière–Polyak and Fletcher–Reeves conjugate gradient method for solving unconstrained optimization problems. Global convergence, with the strong Wolfe line search conditions, of the proposed quadratic hybrid conjugate gradient method is established. The new method is tested on a number of benchmark problems that have been extensively used in the literature and numerical results show the competitiveness of the new hybrid method.

1. Introduction

Nonlinear conjugate gradient method is a very powerful technique for solving large scale unconstrained optimization problems

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function. It has advantages over Newton and quasi-Newton methods in that it only needs the first order derivative and hence less storage capacity is needed. It is also relatively simple to program.

Given an initial guess $x_0 \in \mathbb{R}^n$, the nonlinear conjugate gradient method generates a sequence $\{x_k\}$ for problem (1) as

$$x_{k+1} = x_k + \alpha_k d_k, \qquad k = 0, 1, 2, \dots, \tag{2}$$

where α_k is a step length which is determined by a line search and d_k is a descent direction of f at x_k generated as

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

where $g_k = \nabla f(x_k)$ is the gradient of f at x_k and β_k is a parameter.

Conjugate gradient methods differ in their way of defining the parameter β_k. Over the years, several choices of β_k, which give rise to different conjugate gradient methods, have been proposed. The most famous formulas for β_k are Fletcher-Reeves (FR) method [20]

$$\beta_k^{FR} = \frac{||g_k||^2}{||g_{k-1}||^2},$$

Polak-Ribiére-Polyak (PRP) method [32, 33]

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

Dai-Yuan (DY) method [12]

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

and the Hestenes-Stiefel (HS) method [18, 23]

*Corresponding Author

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}},$$

where $y_{k-1} = g_k - g_{k-1}$ and $||\cdot||$ denotes the Euclidean norm of vectors. These were the first scalars β_k for nonlinear conjugate gradient methods to be proposed. Since then, other parameters β_k have been proposed in the literature (see for example [1,2,4–6,14,15,17,19,22,28,35,39,40] and references therein).

From the literature, it is well known that FR and DY methods have strong convergence properties. However, they may not perform well in practice. On the other side, PRP and HS methods are known to perform better numerically but may not converge in general. Given this, researchers try to devise some new methods, which have the advantages of these two kinds of methods. This has been done mostly by combining two or more β_k parameters in the same conjugate gradient method to come up with hybrid methods. Thus, hybrids try to combine attractive features of different algorithms. For example, Touati-Ahmed and Storey [36] proposed this hybrid method

$$\beta_k^{TS} = \max\left\{0, \min(\beta_k^{FR}, \beta_k^{PRP})\right\}$$

to take advantage of the attractive convergence properties of β_k^{FR} and numerical performance of β_k^{PRP}.

Many other hybrids have been proposed by parametrically combining different parameters β_k. In Dai and Yuan [11], for instance, a one-parameter family of conjugate gradient methods is proposed as

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

where the parameter λ_k is such that $\lambda_k \in [0,1]$. Liu and Li [28] proposes a convex combination of β_k^{LS} and β_k^{DY} to get

$$\beta_k = (1-\gamma_k)\beta_k^{LS} + \gamma_k \beta_k^{DY},$$

where $\beta_k^{LS} = -\frac{g_k^T y_{k-1}}{d_{k-1}^T g_{k-1}}$ is the Liu-Storey (LS) [26] parameter and $\gamma_k \in [0,1]$. Other hybrid conjugate gradient methods can be found in [2,4–8,13,21,22,24,25,27,29,35,38,41].

The step length α_k is often chosen to satisfy certain line search conditions. It is very important in the convergence analysis and implementation of conjugate gradient methods. The line search in the conjugate gradient methods is often based on the weak Wolfe conditions

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

and

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

or the stronger version of the Wolfe line search conditions

$$f(x_k + \alpha_k d_k) \le f(x_k) + \mu \alpha_k g_k^T d_k \tag{6}$$

and

$$|g(x_k + \alpha_k d_k)^T d_k| \le -\sigma g_k^T d_k, \tag{7}$$

where $0 < \mu < \sigma < 1$. More information on these line search methods and other line search methods can be found in the literature [9, 14, 25, 31, 34, 37, 39, 41]. In this paper, we suggest another approach to get a new hybrid nonlinear conjugate gradient method.

The rest of the paper is organised as follows. In section 2, we present the proposed method. In Section 3 we prove that the proposed algorithm (method) globally converges. Section 4 presents some numerical experiments and conclusion is given in Section 5.

2. A new hybrid conjugate gradient method

We now present our proposed hybrid conjugate gradient method. The hybrid method we propose is motivated by the work of Babaie-Kafaki [4, 5] and Mo, Gu and Wei [29]. Babaie-Kafaki [4, 5] suggested a quadratic hybridization of β_k^{FR} and β_k^{PRP} method of the form

$$\beta_k^{HQ\pm} = \begin{cases} \beta_k^{+}(\theta_k^{\pm}), & \theta_k^{\pm} \in [-1,1], \\ \beta_k^{PRP+}, & \theta_k^{\pm} \in \mathbb{C}, \\ -\beta_k^{FR}, & \theta_k^{\pm} < -1, \\ \beta_k^{FR}, & \theta_k^{\pm} > 1, \end{cases} \tag{8}$$

where

$$\beta_k^{+}(\theta_k) = (1-\theta_k^2)\beta_k^{PRP} + \theta_k \beta_k^{FR}, \qquad \theta_k \in [-1,1],$$

and

$$\theta_k^{\pm} = \frac{\beta_k^{FR} \pm \sqrt{(\beta_k^{FR})^2 - 4\beta_k^{PRP}(\beta_k^{HS} - \beta_k^{PRP})}}{2\beta_k^{PRP}}$$

is the solution of the quadratic equation

$$\theta_k^2 \beta_k^{PRP} - \theta_k \beta_k^{FR} + \beta_k^{HS} - \beta_k^{PRP} = 0.$$

Thus, the author suggested two methods β_k^{HQ+} and β_k^{HQ-}. The parameter

$$\beta_k^{PRP+} = \max\{0, \beta_k^{PRP}\}$$

is a hybrid parameter that was suggested by Gilbert and Nocedal [21] to improve on the convergence properties of β_k^{PRP}.

In Mo, Gu and Wei [29], the authors suggest a β_k^* defined by

$$\beta_k^* = \beta_k^{PRP} + \frac{2g_k^T g_{k-1}}{\|g_{k-1}\|^2}, \quad (9)$$

which then modifies the Touati-Ahmed and Storey method [36] to give

$$\beta_k = \max\left\{0, \min(\beta_k^{FR}, \beta_k^{PRP}, \beta_k^*)\right\}.$$

This method by Mo et al. [29] was shown to be very competitive with the other hybrids in the literature and it was shown to perform much better than the original β_k^{PRP}.

Now, motivated by this suggestion (9) from [29] and the work of Babaie-Kafaki [4,5], in this work we modify Babaie-Kafaki's method by introducing β_k^S as

$$\beta_k^S = \begin{cases} \beta_k^+(\theta_k), & \theta_k \in [-1,1], \\ \max\{0, \beta_k^*\}, & \theta_k \in \mathbb{C}, \\ -\beta_k^{FR}, & \theta_k < -1, \\ \beta_k^{FR}, & \theta_k > 1, \end{cases} \quad (10)$$

where

$$\theta_k = \frac{\beta_k^{FR} - \sqrt{(\beta_k^{FR})^2 - 4\beta_k^*(\beta_k^{HS} - \beta_k^*)}}{2\beta_k^*}$$

and

$$\beta_k^+(\theta_k) = (1-\theta_k^2)(\max\{0, \beta_k^*\}) + \theta_k \beta_k^{FR}, \quad \theta_k \in [-1,1], \quad (11)$$

where β_k^* is as defined in (9), and then define

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_{k-1}^S d_{k-1}, & k \geq 1. \end{cases} \quad (12)$$

This leads to our hybrid conjugate gradient method presented below.

Algorithm 1. *New Hybrid β_k^S Conjugate Gradient Method*

Step 1 *Give initial guess $x_0 \in \mathbb{R}^n$, and the parameters $0 < \mu < \sigma < 1$ and $\epsilon > 0$.*

Step 2 *Set $d_0 = -g_0$ and $k = 0$. If $||g_0|| < \epsilon$, stop.*

Step 3 *Compute α_k using the strong Wolfe conditions (6) and (7).*

Step 4 *Set $x_{k+1} = x_k + \alpha_k d_k$, $k = k+1$.*

Step 5 *If $||g_k|| < \epsilon$, stop.*

Step 6 *Compute β_k using (10–11).*

Step 7 *Compute $d_k = -g_k + \beta_k^S d_{k-1}$, go to Step 3.*

3. Global convergence of the proposed method

The global convergence analysis in this section follows that of Babaie-Kafaki [4, 5]. To analyze the global convergence property of our hybrid method, the following assumptions are required. These assumptions have been used extensively in the literature for the global convergence analysis of conjugate gradient methods.

Assumption 1. *Let the level set*

$$\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\},$$

where x_0 is the initial guess, be bounded. That is, there exists a positive constant B such that

$$\|x\| \leq B, \quad \forall x \in \Omega. \quad (13)$$

Assumption 2. *In some neighbourhood N of Ω, the function f is continuously differentiable and its gradient, $g(x) = \nabla f(x)$, is Lipschitz continuous, that is, there exists a constant $L > 0$ such that*

$$\|g(x) - g(y)\| \leq L\|x - y\|$$

for all $x, y \in N$.

These assumptions imply that there exists a positive constant $\hat{\gamma}$ such that

$$\|g(x)\| \leq \hat{\gamma}. \quad (14)$$

Also, under **Assumptions** 1 and 2, the following lemma can be established.

Lemma 1 (Zoutendijk lemma)**.** *Consider any iteration of the form $x_{k+1} = x_k + \alpha_k d_k$, where d_k is a descent direction and α_k satisfies the weak Wolfe conditions (4) and (5). Suppose* ***Assumptions*** *1 and 2 hold, then*

$$\sum_{k=0}^{\infty} \cos^2 \theta_k \|g_k\|^2 < \infty.$$

It follows from **Lemma 1** and the sufficient descent condition with the Wolfe line search that

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{15}$$

Lemma 2. *Suppose that* ***Assumptions*** *1 and 2 hold. Consider any conjugate gradient method in the form of*

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \cdots,$$

and (12) in which, for all $k \geq 0$, the search direction d_k is a descent direction and the step length α_k is determined to satisfy the Wolfe conditions. If

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} = \infty, \tag{16}$$

then the method converges in the sense that

$$\liminf_{k \longrightarrow \infty} \|g_k\| = 0. \tag{17}$$

Lemma 3. *Suppose that* ***Assumptions*** *1 and 2 hold. Consider any conjugate gradient method in the form of*

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \cdots,$$

and (12), with the conjugate gradient parameter $\beta_k^+(\theta_k)$ defined by (11), in which the step length α_k is determined to satisfy the strong Wolfe conditions (6) and (7).
Also assume that the descent condition

$$d_k^T g_k < 0, \quad \forall k \geq 0 \tag{18}$$

holds and there exists a positive constant ξ such that

$$|\theta_k| \leq \xi \alpha_k, \quad \forall k \geq 0. \tag{19}$$

If, for a positive constant γ, we have

$$\|g_k\| \geq \gamma, \quad \forall k \geq 0, \tag{20}$$

then $d_k \neq 0$ and

$$\sum_{k=0}^{\infty} \|u_{k+1} - u_k\|^2 < \infty, \tag{21}$$

where $u_k = \frac{d_k}{\|d_k\|}$.

Proof. Firstly, note that the descent condition (18) guarantees that $d_k \neq 0$. So, u_k is well-defined. Moreover, from (20) and **Lemma 2**, we have

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} < \infty, \tag{22}$$

since otherwise (17) holds contradicting (20). Now, we divide $\beta_k^+(\theta_k)$ into two parts as

$$\beta_k^{(1)} = (1 - \theta_k^2) \max(0, \beta_k^*), \quad \beta_k^{(2)} = \theta_k \beta_k^{FR},$$

and, for all $k \geq 0$, we define

$$r_{k+1} = \frac{v_{k+1}}{\|d_{k+1}\|}, \quad \delta_{k+1} = \beta_k^{(1)} \frac{\|d_k\|}{\|d_{k+1}\|},$$

where

$$v_{k+1} = -g_{k+1} + \beta_k^{(2)} d_k.$$

Therefore, from (12) we obtain that

$$u_{k+1} = r_{k+1} + \delta_{k+1} u_k. \tag{23}$$

Since $\|u_k\| = \|u_{k+1}\| = 1$, from (23) we can write

$$\|r_{k+1}\| = \|u_{k+1} - \delta_{k+1} u_k\| = \|\delta_{k+1} u_{k+1} - u_k\|. \tag{24}$$

Because $\theta_k \in [-1, 1]$, we have $\delta_{k+1} \geq 0$. Using the condition $\delta_{k+1} \geq 0$, the triangle inequality and (24), we get

$$\begin{aligned} \|u_{k+1} - u_k\| &\leq \|(1 + \delta_{k+1}) u_{k+1} - (1 + \delta_{k+1}) u_k\| \\ &\leq \|u_{k+1} - \delta_{k+1} u_k\| + \|\delta_{k+1} u_{k+1} - u_k\| \\ &= 2\|r_{k+1}\|. \end{aligned} \tag{25}$$

Also, from (13), (14), (19) and (20) we have

$$\begin{aligned}
\|v_{k+1}\| &= \|-g_{k+1} + \beta_k^{(2)} d_k\| \\
&= \|-g_{k+1} + \theta_k \tfrac{\|g_{k+1}\|^2}{\|g_k\|^2} d_k\| \\
&\leq \|g_{k+1}\| + |\theta_k| \tfrac{\|g_{k+1}\|^2}{\|g_k\|^2} \|d_k\| \\
&\leq \hat{\gamma} + \tfrac{\xi \alpha_k \hat{\gamma}^2}{\gamma^2} \|d_k\| \qquad (26) \\
&= \hat{\gamma} + \tfrac{\xi \hat{\gamma}^2 \|x_{k+1} - x_k\|}{\gamma^2} \\
&\leq \hat{\gamma} + \tfrac{\xi \hat{\gamma}^2 (\|x_{k+1}\| + \|x_k\|)}{\gamma^2} \\
&\leq \hat{\gamma} + \tfrac{2B\xi\hat{\gamma}^2}{\gamma^2}.
\end{aligned}$$

Now, from (22), (25), and (26) we have

$$\begin{aligned}
\sum_{k=0}^{\infty} \|u_{k+1} - u_k\|^2 &\leq 4\sum_{k=0}^{\infty} \|r_{k+1}\|^2 \\
&= 4\sum_{k=0}^{\infty} \frac{\|v_{k+1}\|^2}{\|d_{k+1}\|^2} \\
&\leq 4\left(\hat{\gamma} + \tfrac{2B\hat{\gamma}^2\xi}{\gamma^2}\right)^2 \sum_{k=0}^{\infty} \frac{1}{\|d_{k+1}\|^2} \\
&\leq 4\left(\hat{\gamma} + \tfrac{2B\hat{\gamma}^2\xi}{\gamma^2}\right)^2 \tfrac{1}{\gamma^4} \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_{k+1}\|^2} \\
&< \infty
\end{aligned} \qquad (27)$$

□

We now define the following property, called property (*).

Definition 1. *[10] Consider any conjugate gradient method in the form of*

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \cdots,$$

and (12). Suppose that for a positive constant γ the inequality (20) holds. Under this assumption, we say that the method has property (*) if and only if there exist constants $b > 1$ and $\lambda > 0$ such that for all $k \geq 0$,

$$|\beta_k| \leq b, \qquad (28)$$

and

$$\|\alpha_k d_k\| \leq \lambda \Rightarrow |\beta_k| \leq \frac{1}{b}. \qquad (29)$$

Theorem 1. *Suppose that* ***Assumptions*** *1 and 2 hold. Consider any conjugate gradient method in the form of*

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \cdots,$$

and (12), with the conjugate gradient parameter $\beta_k^+(\theta_k)$ defined by (11), in which the step length α_k is determined to satisfy the strong Wolfe conditions (6) and (7). If the search directions satisfy the descent condition (18) and there exists a positive constant η such that

$$|\theta_k| \leq \eta \|\alpha_k d_k\|, \quad \forall k \geq 0, \qquad (30)$$

then the method converges in the sense that

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

Proof. Because of the descent condition and strong Wolfe conditions, we have proven that the sequence $\{x_k\}_{k\geq 0}$ is a subset of the level set Ω. Also, since all the assumptions of **Lemma 2** hold, the inequality (21) holds. Now, to prove the convergence, it is enough to show that the method has property (*).
Since $\theta_k \in [-1, 1]$, from (11), (14), and (20) we have

$$\begin{aligned}
|\beta_k^+(\theta_k)| &= |(1-\theta_k)^2(\max\{0, \beta_k^*\}) + \theta_k \beta_k^{FR}| \\
&\leq |(1-\theta_k)^2| \left| \beta_k^{PRP} + \tfrac{2g_{k+1}^T g_k}{\|g_k\|^2} \right| + |\theta_k||\beta_k^{FR}| \\
&\leq |\beta_k^{PRP}| + \tfrac{|2g_{k+1}^T g_k|}{\|g_k\|^2} + \beta_k^{FR} \\
&\leq \tfrac{\|g_{k+1}\|(\|g_{k+1}\| + \|g_k\|)}{\|g_k\|^2} + \tfrac{2\|g_{k+1}\|\|g_k\|}{\|g_k\|^2} + \tfrac{\|g_{k+1}\|^2}{\|g_k\|^2} \\
&\leq \tfrac{2\hat{\gamma}^2}{\gamma^2} + \tfrac{2\hat{\gamma}^2}{\gamma^2} + \tfrac{\hat{\gamma}^2}{\gamma^2} \\
&= \tfrac{5\hat{\gamma}^2}{\gamma^2}.
\end{aligned} \qquad (31)$$

Moreover, from **Assumption** 2 and equations (11), (14), (20), and (30) we get

$$\begin{aligned}
|\beta_k^+(\theta_k)| &= |(1-\theta_k)^2(\max\{0, \beta_k^*\}) + \theta_k \beta_k^{FR}| \\
&\leq |(1-\theta_k)^2| \left| \beta_k^{PRP} + \tfrac{2g_{k+1}^T g_k}{\|g_k\|^2} \right| + |\theta_k||\beta_k^{FR}| \\
&\leq |\beta_k^{PRP}| + \tfrac{|2g_{k+1}^T g_k|}{\|g_k\|^2} + |\theta_k|\beta_k^{FR} \\
&\leq \tfrac{\|g_{k+1}\|\|g_{k+1} - g_k\|}{\|g_k\|^2} + \tfrac{2\|g_{k+1}\|\|g_k\|}{\|g_k\|^2} + |\theta_k| \tfrac{\|g_{k+1}\|^2}{\|g_k\|^2}
\end{aligned}$$

$$
\begin{aligned}
&\leq \tfrac{L\hat{\gamma}\|x_{k+1}-x_k\|}{\gamma^2} + \tfrac{2\hat{\gamma}^2}{\gamma^2} + \tfrac{\eta\hat{\gamma}^2}{\gamma^2}\|\alpha_k d_k\| \\
&\leq \tfrac{L\hat{\gamma}}{\gamma^2}\|\alpha_k d_k\| + \tfrac{2\eta\hat{\gamma}^2}{\gamma^2}\|\alpha_k d_k\| + \tfrac{\eta\hat{\gamma}^2}{\gamma^2}\|\alpha_k d_k\| \qquad (32) \\
&= \tfrac{L\hat{\gamma}+3\eta\hat{\gamma}^2}{\gamma^2}\|\alpha_k d_k\|.
\end{aligned}
$$

So, from (31) and (32), if we let

$$b = \frac{5\hat{\gamma}^2}{\gamma^2} \quad \text{and} \quad \lambda = \frac{\gamma^2}{b(L\hat{\gamma} + 3\eta\hat{\gamma}^2)},$$

then (28) and (29) hold and consequently, the method has property (*). □

4. Numerical Experiments

We now present numerical experiments obtained by our method on some test problems chosen from Morè, et al. [30] and Andrei [3] to analyse its efficiency and effectiveness. A number of these test problems are widely used in the literature for testing unconstrained optimization methods. We present these test problems in Table 1, where the columns 'Prob' and 'Dim', respectively, represent the name and dimension of the test problem, and the dimensions of the problems range from 2 to 20000.

We compare our proposed new hybrid conjugate gradient method (β_k^S) with the quadratic hybridization β_k^{HQ-} of Babaie-Kafaki [4, 5] and the method β_k^* by Mo, Gu and Wei [29]. In [4], β_k^{HQ-} was shown to be the better hybridization as compared to β_k^{HQ+}, hence our comparison will only focus on β_k^{HQ-}. For all the methods, we considered the stopping condition to be $\epsilon = 10^{-5}$, that is, the algorithms (methods) were stopped once the condition $||g_k|| < 10^{-5}$ was satisfied, or the maximum number of iterations of 5000 was reached. For the line search, the strong Wolfe conditions (6) and (7) were used to find the step length α_k, with $\mu = 0.0001$ and $\sigma = 0.16$. All the methods were coded in MATLAB R2015b and numerical results are compared based on number of gradient evaluations, function evaluations and CPU time.

In Table 1, we present the number of functions evaluations (NFE) and gradient evaluations (NGE) obtained for the methods β_k^{HQ-}, β_k^S and β_k^*, where the best results for each problem are indicated in bold. We observe from the table that, overall, the incorporation of β_k^* in the quadratic hybridization has a positive effect on β_k^{HQ-}, even though for some problems it is worse off.

We also compare the methods using the performance profiles tool suggested by Dolan and Moré [16] which, over the years, has been used extensively to judge the performance of different methods on a given set of test problems. The tool evaluates and then compares the performance of the set of methods S on a set P of test problems. That is, using the ratio

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

where $t_{p,s}$ is (function, gradient, CPU time) evaluations required to solve p by method s, the overall performance profile function is

$$\rho_s(\tau) = \frac{1}{n_p}\text{size}\{p : 1 \leq p \leq n_p, \log(r_{p,s}) \leq \tau\},$$

where n_p is the total number of problems in P and $\tau \geq 0$.

In case the method s fails to solve problem p, the ratio $r_{p,s}$ is set to some sufficiently large number. The function $\rho_s(\tau)$ is then plotted against τ to give the performance profile. Notice that the function $\rho_s(\tau)$ takes the values $\rho_s(\tau) \in [0, 1]$ and so the inequality $\rho_s(\tau_1) < \rho_t(\tau_1)$ shows that the method t outperforms the method s at τ_1.

We now present the plots of these performance profiles on function evaluations, gradient evaluations and CPU time as figures. The function evaluations performance profile is presented in Figure 1, gradient evaluations in Figure 2 and CPU time in Figure 3. It is clear from the figures that replacing β_k^{PRP} by β_k^* in the quadratic hybridization β_k^{HQ-} has a positive effect. From the figures, we observe that β_k^* is the best method overall. As for β_k^S and β_k^{HQ-}, we see that in Figures 1 and 2, for $\tau \leq 0.5$, β_k^{HQ-} is slightly better than β_k^S. However, in Figure 3, the plot shows that in terms of CPU time, there is not much difference between β_k^S and β_k^{HQ-} with the methods being much competitive. Overall the figures show the influence of β_k^* on the quadratic hybridization over the use of β_k^{PRP}.

Table 1. Results of test problems.

Prob	Dim	β_k^{HQ-}		β_k^S		β_k^*	
		NFE	NGE	NFE	NGE	NFE	NGE
Rosenbrock	2	133	63	**95**	43	**95**	**38**
Freud and Roth	2	40	**14**	45	16	**39**	**14**
Beale	2	**40**	**24**	74	38	48	28
Helical valley	3	**130**	**50**	264	103	152	61
Bard	3	91	53	**66**	**44**	75	50
Gaussian	3	**7**	**6**	9	7	9	7
Box	3	**41**	**30**	44	32	47	37
Powell Singular	4	508	236	351	168	**208**	**110**
Wood	4	592	160	414	134	**211**	**70**
Biggs EXP6	6	**201**	**139**	502	374	338	284
Osborne 2	11	**660**	**344**	684	352	1557	762
Broyden tridiagonal	30	**90**	**33**	**90**	**33**	94	35
Ext. TET	100	**17**	**10**	28	15	20	11
Gen. White & Holst	100	11088	2614	**9477**	**2329**	14012	3451
Ext. Penalty	500	63	20	**51**	**15**	52	**15**
Ext. Maratos	500	**234**	109	630	259	257	**105**
Gen. Rosenbrock	1000	**21985**	**4574**	22010	4581	25069	6558
Fletcher	1000	**15881**	**4664**	**15881**	**4664**	27857	7616
Ext. Rosenbrock	5000	133	63	**100**	45	103	**42**
	10000	133	63	**100**	45	103	**42**
Ext. Powell singular	10000	**257**	**136**	623	294	273	154
	20000	475	236	1370	656	**217**	**121**
Raydan 2	5000	52	52	7	7	**6**	**6**
	10000	101	101	8	8	**6**	**6**
Ext. Beale	10000	77	45	70	42	**69**	**41**
	20000	77	45	70	42	**69**	**41**
Ext. Himmelblau	10000	**32**	**13**	38	15	35	14
	20000	**32**	**13**	38	15	35	14
Ext. DENSCHNB	10000	**15**	**9**	19	13	17	10
Ext. DENSCHNF	10000	75	33	90	36	**64**	**29**
Ext. Freud & Roth	10000	40	14	45	16	**39**	**14**
Ext. White & Holst	10000	300	124	247	92	**137**	**62**
Ext. Wood	10000	757	200	450	145	**224**	**74**
NONSCOMP	10000	**147**	**59**	151	61	173	72
Quartic	10000	81	80	**28**	**27**	44	43

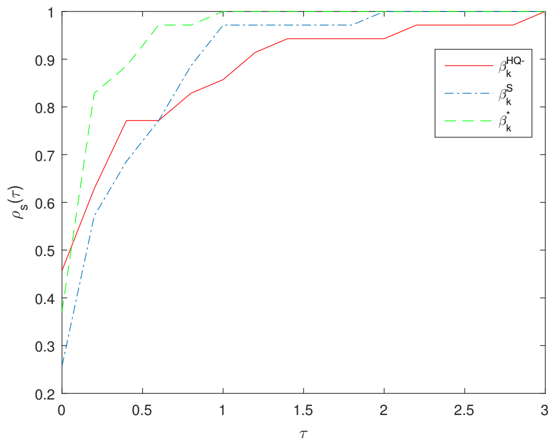

Figure 1. Function evaluations profile.

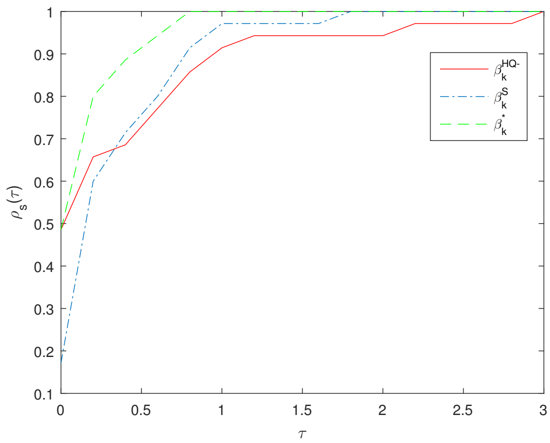

Figure 2. Gradient evaluations profile.

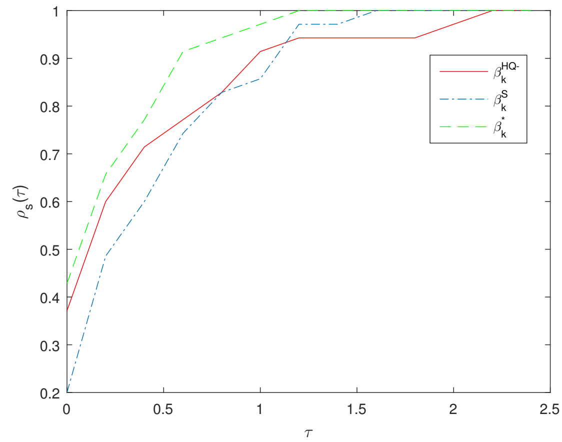

Figure 3. CPU time profile.

5. Conclusion

In this article, a modified quadratic hybridization of Polak–Ribière–Polyak and Fletcher–Reeves conjugate gradient method (β_k^S) was presented. Its global convergence under the strong Wolfe line search conditions was also established. The β_k^S method presented was tested on a number of unconstrained problems that have been extensively used in the literature and compared to the original quadratic hybridization of Polak–Ribière–Polyak and Fletcher–Reeves conjugate gradient method β_k^{HQ-}. The numerical results show that this proposed modification has a positive effect on the performance of β_k^{HQ-}. However, the numerical results from this study show that further research to improve the efficiency and effectiveness of β_k^{HQ-} and other conjugate gradient hybrids is still needed. A number of hybrid conjugate gradient methods have been proposed in the literature but there are many problems that are currently not properly handled by these methods, hence the need for more research in this field.

Acknowledgments

The authors thank the anonymous Referees and the Editors whose comments and suggestions largely improved this work.

References

[1] Al-Baali, M., Narushima, Y. and Yabe, H., A family of three term conjugate gradient methods with sufficient descent property for unconstrained optimization, Comput. Optim. Appl., 60, 89–110 (2015).

[2] Andrei, N., A new three-term conjugate gradient algorithm for unconstrained optimization, Numer. Algorithms, 68, 305–321 (2015).

[3] Andrei, N., An unconstrained optimization test functions collection, Adv. Model. Optim., 10(1), 147–161 (2008).

[4] Babaie-Kafaki, S., A quadratic hybridization of Polak-Ribiere-Polyak and Fletcher-Reeves conjugate gradient methods, J Optim. Theory Appl., 154(3), 916-932 (2012).

[5] Babaie-Kafaki, S., A Note on the global convergence of the quadratic hybridization of Polak-Ribiere-Polyak and Fletcher-Reeves conjugate gradient methods, J Optim. Theory Appl., 157(1), 297-298 (2013).

[6] Babaie-Kafaki, S., Two modified scaled nonlinear conjugate gradient methods, J Comput. Appl. Math., 261, 172–182 (2014).

[7] Babaie-Kafaki, S., Fatemi, M., and Mahdavi-Amiri, N., Two effective hybrid conjugate gradient algorithms based on modified BFGS updates, Numer. Algorithms, 58, 315–331 (2011).

[8] Babaie-Kafaki, S. and Ghanbari, R., A descent extension of the Polak-Ribière-Polyak conjugate gradient method, Comp. Math. Appl., 68, 2005–2011 (2014).

[9] Dai, Y.-H., Conjugate gradient methods with Armijo-type line searches, Acta Math. Appl. Sin. Engl. Ser., 18(1), 123–130 (2002).

[10] Dai, Y.-H. and Liao, L.Z., New conjugacy conditions and related nonlinear conjugate gradient methods, Appl. Math. Optim., 43, 87–101 (2001).

[11] Dai, Y.-H. and Yuan, Y., A class of globally convergent conjugate gradient methods, Sci. China Ser. A., 46(2), 251–261 (2003).

[12] Dai, Y.-H. and Yuan, Y., A nonlinear conjugate gradient method with strong global convergence property, SIAM J. Optim., 10(1), 177-182 (1999).

[13] Dai, Y.-H., and Yuan, Y., An efficient hybrid conjugate gradient method for unconstrained optimization, Ann. Oper. Res., 103, 33–47 (2001).

[14] Dai, Z. and Wen, F., Another improved Wei-Yao-Liu nonlinear conjugate gradient method with sufficient descent property, Appl. Math. Comput., 218:7421–7430 (2012).

[15] Dai, Z.-F. and Tian, B.-S., Global convergence of some modified PRP nonlinear conjugate gradient methods, Optim. Lett., 5, 615–630 (2011).

[16] Dolan, E.D. and Moré, J.J., Benchmarking optimization software with profile performance profiles, Math. Program., 91(2), 201–213 (2002).

[17] Dong, X.L., Liu, H. and He, Y., A self-adjusting conjugate gradient method with sufficient descent condition and conjugacy condition, J Optim. Theory Appl., 165, 225–241 (2015).

[18] Dong, X.L., Liu, H.W., He, Y.B. and Yang, X.M., A modified Hestenes-Stiefel conjugate gradient method with sufficient descent condition and conjugacy condition, J Comput. Appl. Math., 281, 239–249 (2015).

[19] Dong, Y., A practical PR+ conjugate gradient method only using gradient, Appl. Math. Comput., 219(4), 2041-2052 (2012).

[20] Fletcher, R. and Reeves, C., Function minimization by conjugate gradients, Comp. J., 7, 149–154 (1964).

[21] Gilbert, J.C., and Nocedal, J., Global convergence properties of conjugate gradient methods for optimization, SIAM J. Optim., 2(1), 21–42 (1992).

[22] Hager, W.W., and Zhang, H., A survey of nonlinear conjugate gradient methods, Pac. J. Optim., 2(1), 35–58 (2006).

[23] Hestenes, M.R., and Stiefel, E., Methods for conjugate gradients for solving linear systems, J Res. Natl. Bur. Stand., 49, 409–436 (1952).

[24] Jia, W., Zong, J. and Wang, X., An improved mixed conjugate gradient method, Systems Engineering Procedia 2, 219–225 (2012).

[25] Liu, H., A mixture conjugate gradient method for unconstrained optimization, Third International Symposium on Intelligent Information Technology and Security Informatics, IEEE, 26–29 (2010).

[26] Liu, Y. and Storey, C., Efficient generalized conjugate gradient algorithms, Part 1: theory, J Optim. Theory Appl., 69, 129-137 (1992).

[27] Liu, J., A hybrid nonlinear conjugate gradient method, Lobachevskii J. Math., 33(3), 195–199 (2012).

[28] Liu, J.K. and Li, S.J., New hybrid conjugate gradient method for unconstrained optimization, Appl. Math. Comput., 245, 36–43 (2014).

[29] Mo, J., Gu, N. and Wei, Z., Hybrid conjugate gradient methods for unconstrained optimization, Optim. Methods Softw., 22(2), 297–307 (2007).

[30] Moré, J.J., Garbow, B.S. and Hillstrom, K.E., Testing unconstrained optimization software, ACM Trans. Math. Software, 7, 17–41 (1981).

[31] Nocedal, J. and Wright, S.J., Numerical Optimization, 2nd Edition, Springer Science+ Business Media, LLC. Printed in the United States of Ameraica (2006).

[32] Polak, E., and Ribière, G., Note sur la convergence de mèthodes de directions conjuguèes, Rev. Française d'Infom. Recherche Opèrationnelle, 3(16), 35–43 (1969).

[33] Polyak, B.T., The conjugate gradient method in Extreme problems, USSR Comput. Math. Math. Phys., 9(4), 94–112 (1969).

[34] Shi, Z.-J., Convergence of line search methods for unconstrained optimization, Appl. Math. Comput., 157, 393–405 (2004).

[35] Sun, M. and Liu, J., Three modified Polak-Ribière-Polyak conjugate gradient methods with sufficient descent property, Journal of Inequalities and Applications, 2015–2125 (2015).

[36] Touati-Ahmed, D. and Storey, C., Efficient hybrid conjugate gradient techniques, J Optim. Theory Appl., 64(2), 379–397 (1990).

[37] Yabe, H. and Sakaiwa, N., A new nonlinear conjugate gradient method for unconstrained optimization, J Oper. Res., 48(4), 284–296 (2005).

[38] Yan, H., Chen, L. and Jiao, B., HS-LS-CD hybrid conjugate gradient algorithm for unconstrained optimization, Second International Workshop on Computer Science and Engineering, IEEE, 264–268 (2009).

[39] Yuan, G. and Lu, X., A modified PRP conjugate gradient method, Ann. Oper. Res., 166, 73–90 (2009).

[40] Zhang, L., Zhou, W. and Li, D.-H., A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence, IMA J. Numer. Anal., 26, 629–640 (2006).

[41] Zhou, A., Zhu, Z., Fan, H. and Qing, Q., Three new hybrid conjugate gradient methods for optimization, Appl. Math., 2(3), 303–308 (2011).

S. Narayanan *graduated with an MSc in Mathematics from the Department of Mathematics, University of Botswana, in 2015. She is currently pursuing her PhD in Mathematics from the Department of Mathematics, University of Botswana. Her research interests are in Optimization and Analysis.*

P. Kaelo *is a lecturer in the Department of Mathematics, University of Botswana. He obtained his PhD in Mathematics from the University of the Witwatersrand, Johannesburg, South Africa, in 2005. His research interests include Optimization, Optimal Control Theory, Numerical Analysis and Functional Analysis.*

M.V. Thuto *is a lecturer in the Department of Mathematics, University of Botswana. He obtained his PhD in Mathematics from the University of Exeter, England, United Kingdom, in 2000. His research interests are in Control Theory, Optimization and Functional Analysis.*