ICIC Express Letters
Volume **13**, Number **8**, August **2019**

ICIC International ©2019 ISSN 1881-803X
pp. 681–687

A NEW HYBRID CONJUGATE GRADIENT METHOD FOR OPTIMIZATION PROBLEMS AND ITS GLOBAL CONVERGENCE

Rui Shan[1], Lijiu Jia[1], Jingyi Zhao[2] and Wen Liu[1]

[1]School of Science
[2]School of Mechanical Engineering
Yanshan University
No. 438, West Hebei Avenue, Qinhuangdao 066004, P. R. China
sr@ysu.edu.cn; 1328849200@qq.com; zjylwsr@126.com; liuwen1961@hotmail.com

Received December 2018; accepted March 2019

Abstract. *In order to achieve the theoretically effective and numerically efficient method for solving optimization problems, we propose a new hybridization of the Wei-Yao-Liu (WYL) and Fletcher-Reeves (FR). In the method, the hybridization parameters are determined by assuming the descent condition. Under proper conditions, property of the method is attractive. Moreover, the global convergence of the method does not require the assumption that the objective function is convex. The numerical results show that the presented method is competitive to the FR and the WYL method.*

Keywords: Conjugate gradient method, Unconstrained optimization, Global convergence

1. **Introduction.** Consider the unconstrained optimization problem:

$$\min_{x \in R^n} f(x), \tag{1}$$

where f: $R^n \to R^1$ is a continuously differentiable function. We denote its gradient by g.

Commonly, the conjugate gradient method (CG) is a very important technique for solving this problem, especially the large dimensional or nonlinear problem [1]. CG method arises from the study of quadratic function minimization, but it can be extended to deal with the minimization of non-quadratic functions. The iterative process of the CG method is given by

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

x_k is the current iterate, α_k is the step length which is determined by some line search and d_k is the search direction defined by:

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

β_k is the regulatory parameter of search direction d_k. CG methods are determined by different scalar forms of β_k, and CG methods have different search directions d_k [2]. Some common forms of β_k are defined as follows:

$\beta_k^{HS} = \dfrac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}$ HS (Hestenes-Stiefel) algorithm [3]

$\beta_k^{FR} = \dfrac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$ FR algorithm [4]

$\beta_k^{DY} = \dfrac{g_k^T g_k}{d_{k-1}^T(g_k - g_{k-1})}$ DY (Dai-Yuan) algorithm [7]

DOI: 10.24507/icicel.13.08.681

$\beta_k^{PRP} = \dfrac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}$ PRP algorithm [5,6]

Generally, FR and DY algorithms are considered to have good convergence property, but the numerical results are poor; the PRP and HS methods have superior numerical result. It is well-known, when the objective function is convex, Polak and Ribeire proved that PRP method is globally convergent with the exact line search [5]. However, Powell demonstrated that the PRP method cannot achieve global convergence when the objective function is non-convex, and suggested that β_k should not be less than zero [8]. Recently, Wu made a further study and inferred a modified three-term PRP CG algorithm [9]. Du et al. proposed four modified CG methods and proved these methods possess sufficient descent property [10]. Gao and He proposed a three-term CG method for solving nonlinear monotone equations with convex constraints and the method applied to solving the large-scale non-smooth nonlinear monotone equations [11]. Based on a singular value analysis on an extension of the PRP method, Babaie-Kafaki and Ghanbari put a nonlinear CG method and proved it is globally convergent with the Weak Wolfe-Powell (WWP) line search technique [12]. And the WWP conditions are represented as follows:

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

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k. \tag{5}$$

This line search technique is often used to study the convergence of CG method [13,14]. To ensure that the β_k is not negative, Wei, Yao, and Liu (WYL) propose a new CG formula

$$\beta_k^{WYL} = \frac{g_k^T \left(g_k - \dfrac{||g_k||}{||g_{k-1}||} g_{k-1} \right)}{||g_{k-1}||^2}. \tag{6}$$

The numerical experiments show that WYL is superior to PRP method, under the sufficient descent condition, it is global convergence with the WWP line search [14]. On the basis of the modified secant equation, Yin et al. acquired a modified PRP CG method and the search direction generated by the presented method possesses both the sufficient descent and trust region properties without carrying out any line search [15]. Li et al. presented a modified CG algorithm by line search method with acceleration scheme for nonlinear symmetric equations. The proposed method not only possesses descent property but also owns global convergence in mild conditions [16]. It is also an effective way to obtain high efficiency CG algorithm by constructing hybrid conjugate gradient method. Considering the good numerical result of HS method and the strong convergence property of DY method, Dai and Yuan projected the HS method and the DY method obtained the hybrid CG method. Under the WWP, the authors proved the global convergence, and a better numerical result than the PRP method is obtained in [17]. Andrei used PRP and DY method to construct convex combination hybrid CG method, and the parameters of hybridization are determined by conjugate conditions [18]. Recently, Babaie-Kafaki and Ghanbari dealt with the convex combination of the PRP and FR method [19], the parameter β_k is derived by solving the least-squares problem $\min\limits_{\theta_k} \left\| d_k^{HCG} - d_k^{ZZL} \right\|$, d_k^{HCG} is the search direction of this article, and d_k^{ZZL} is the search direction of the three-term CG method by Zhang et al. [20]. This motivates us to propose a hybrid method combining the FR and the WYL methods. The hybrid method will possess some better properties of the FR and the WYL method: (i) The scalar $\beta_k \geq 0$ holds automatically. (ii) The global convergence with the WWP line search of the presented method is established for non-convex objective function.

This article is divided into the following parts. In the next part, the algorithm is stated. The global convergence is proved in Part 3, and the numerical results are reported in Part 4. The last part gives the conclusion.

2. **New Algorithm.** Now we describe the following convex combining the new algorithm of the WYL and FR methods:

$$\beta_k^{HWF} = (1-\theta_k)\beta_k^{WYL} + \theta_k\beta_k^{FR}. \tag{7}$$

β_k^{HWF} is called the hybridization parameter. If $\theta_k = 0$, then $\beta_k^{HWF} = \beta_k^{WYL}$, and if $\theta_k = 1$, then $\beta_k^{HWF} = \beta_k^{FR}$. Then, we aim at finding a suitable choice for the hybridization parameter θ_k. In this case, the search direction of new CG method with the parameter (7) can be written as:

$$d_0^{HWF} = -g_0, \quad d_k^{HWF} = -g_k + \beta_k^{WYL} d_{k-1} + \theta_k\left(\beta_k^{FR} - \beta_k^{WYL}\right) d_{k-1}, \quad \forall k \geq 0. \tag{8}$$

θ_k are determined by assuming descent condition $d_k^T g_k < 0$, from (8), after some algebraic manipulations, θ_k given by:

$$\theta_k^* = \frac{1}{\cos\varphi} - 1. \tag{9}$$

φ is the included angle of g_{k-1} and g_k. θ_k^* is well-defined when $\cos\varphi < 1 - \frac{\|g_{k-1}\|^2}{\|g_k\|^2}$. However, θ_k^* given by (9) may be outside the $[0,1]$. In order to maintain the convexity in our hybridization of the WLY and FR method, if $\theta_k^* < 0$, then we set $\theta_k^* = 0$, and if $\theta_k^* > 1$, then we set $\theta_k^* = 1$. That is, we deal with the following CG parameter:

$$\beta_k^{HWF} = (1-\theta_k)\beta_k^{WYL} + \theta_k\beta_k^{FR}.$$

Based on the above discussion, the hybridization parameter θ_k is computed by:

$$\theta = \begin{cases} 0, & \text{if } \theta_k^* < 0, \text{ or } g_k^T g_{k-1} = 0 \\ \theta_k^*, & \text{if } \cos\varphi < 1 - \dfrac{\|g_{k-1}\|^2}{\|g_k\|^2}, \text{ and } \theta_k^* \in [0,1] \\ 1, & \text{if } \theta_k^* > 1 \end{cases} \tag{10}$$

Now we describe the algorithm as follows.

Algorithm 2.1 (the convex hybrid method).

Step 1. Choose an initial point $x_0 \in \Re^n$, $\varepsilon \in (0,1)$. *Set* $d_0 = -g_0 = -\nabla f(x_0)$, $k := 0$.

Step 2. If $\|g_k\| \leq \varepsilon$, *then stop; otherwise go to the next step.*

Step 3. Compute steplength α_k *by the Wolfe line search criteria.*

Step 4. Let $x_{k+1} = x_k + \alpha_k d_k$. *If* $\|g_{k+1}\| \leq \varepsilon$, *then stop.*

Step 5. Calculate the search direction by (8).

Step 6. Set $k := k+1$, *and go to Step 3.*

3. **The Global Convergence Properties.** In this section, we prove the global convergence of new algorithm and the following assumptions are often needed (see [7,14,21]).

Assumption 3.1. *The level set* $\Omega = \{x \in R^n | f(x) \leq f(x_0)\}$ *is bounded when* x_0 *is the initial point.*

Assumption 3.2. *The function* f *is continuously differentiable in a neighborhood* Φ *of* Ω, *and the function gradient satisfies the Lipschitz continuity condition, that is, there exists a positive constant* L *such that the following holds*

$$\|g(x_1) - g(x_2)\| \leq L\|x_1 - x_2\|, \quad \forall x_1, x_2 \in \Phi. \tag{11}$$

We will follow the convergence analysis of [7], and the following important condition is essential [13].

Lemma 3.1. *(Zoutendijk condition) Suppose that Assumptions 3.1 and 3.2 hold. Consider the iterative method (2) in which, for all $k \geq 0$, the search direction d_k is a descent direction and the steplength α_k satisfies the Wolfe conditions (4) and (5). Then*

$$\sum_{k=0}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < \infty. \tag{12}$$

Proof: By (5) and Assumption 3.2, we have

$$-(1-\sigma) g_k^T d_k \leq (g_{k+1} - g_k)^T d_k \leq \alpha_k L \|d_k\|^2,$$

this means that $\alpha_k \geq \frac{-(1-\sigma) g_k^T d_k}{L\|d_k\|^2}$, together with the descent condition $g_k^T d_k < 0$, and (4) implies that

$$\frac{\delta(1-\sigma)(g_k^T d_k)^2}{L \|d_k\|^2} \leq f_k - f_{k+1},$$

summing up this inequality from $k = 0$ to ∞, and using Assumption 3.1, we can obtain this lemma.

Then, we will show the Property(*), proposed by Gilbert and Nocedal which ensures that β_k is small when the step α_k is small [22].

Property(*). *Consider a method of the form (2)-(3), and suppose that*

$$0 < \overline{\gamma} \leq \|g_k\| \leq \gamma, \tag{13}$$

for all $k \geq 0$. Under this assumption, we say the method has Property(*) if there exist constants $b > 1$ and $\lambda > 0$ such that for all k,

$$|\beta_k| \leq b, \tag{14}$$

and

$$\|s_k\| \leq \lambda \Rightarrow |\beta_k| \leq \frac{1}{2b}. \tag{15}$$

Lemma 3.2. *Suppose that Assumptions 3.1 and 3.2 hold. Consider the method (2)-(3) with the following three properties: (i) $\beta_k \geq 0$, for all $k \geq 0$; (ii) the line search satisfies $\{x_k\}_{k\geq 0} \subset \Omega$, the sufficient descent condition and the Zoutendjk condition; (iii) Property(*) holds [22]. Then, the method converges in the sense that*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{16}$$

Now, we can prove the following global convergence lemma for our hybrid CG method.

Lemma 3.3. *Suppose that Assumptions 3.1 and 3.2 hold. Consider the CG method (2)-(3) with the CG parameter β^{HWF} defined by (7), in which the steplength α_k is determined by Wolfe condition (4) and (5). If the sufficient descent condition holds, $\theta_k \in [0,1]$, and there exists a positive constant M such that*

$$\theta_k \leq M \|s_k\|, \tag{17}$$

then the method converges in the sense that (16) holds.

Proof: Firstly, because of the descent condition and (4), the sequence $\{x_k\}_{k\geq 0}$ is a subset of the level set Ω. Also, from Lemma 3.1, the Zoutendijk condition holds. Consider Lemma 3.2, it is enough to complete the proof if the method has Property(*).

Let us consider β_k^{WYL}. Denoting $Y_k = g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}$, by (11), we get

$$\begin{aligned} \|Y_k\| &= \left\| g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right\| \\ &= \left\| g_k - g_{k-1} + g_{k-1} - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right\| \\ &\le \|g_k - g_{k-1}\| + \left| \|g_{k-1}\| - \|g_k\| \right| \\ &\le 2 \|g_k - g_{k-1}\| . \end{aligned} \tag{18}$$

By (18), (6) and (7), since $\theta_k \in [0, 1]$, from (13) we have

$$\begin{aligned} \left|\beta_k^{NWF}\right| &\le \beta_k^{WYF} + \beta_k^{FR} \\ &\le \frac{g_k^T Y_k}{\|g_{k-1}\|^2} + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \\ &\le \frac{\|g_k\| \cdot 2 \|g_k - g_{k-1}\|}{\|g_k\|^2} + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \\ &\le \frac{\gamma \cdot 4\gamma}{\overline{\gamma}^2} + \frac{\gamma^2}{\overline{\gamma}^2} \\ &\le 5 \frac{\gamma^2}{\overline{\gamma}^2}, \end{aligned} \tag{19}$$

and from (11) and (17) we get

$$\begin{aligned} \left|\beta_k^{NWF}\right| &\le \beta_k^{WYL} + \theta_k \beta_k^{FR} \\ &\le \frac{g_k^T Y_k}{\|g_{k-1}\|^2} + M \|s_k\| \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \\ &\le \frac{2L\gamma \|s_k\|}{\overline{\gamma}^2} + M \|s_k\| \frac{\gamma^2}{\overline{\gamma}^2} \\ &\le \frac{2L\gamma + M\gamma^2}{\overline{\gamma}^2} \|s_k\| . \end{aligned} \tag{20}$$

Therefore, from (19) and (20), if we let

$$b = 5\frac{\gamma^2}{\overline{\gamma}^2}, \text{ and } \lambda = \frac{\overline{\gamma}^2}{2b(2L\gamma + M\gamma^2)}, \tag{21}$$

then (14) and (15) hold. Thus, the method has Property(*). NWF algorithm is global convergence. The proof is completed.

4. **Numerical Experiments.** In this section, we present some numerical results obtained by MATLAB implementations of the NWF method with the hybridization parameter θ_k given by (10), WYL, and FR. We chose 10 test functions, and during the test, the parameters are set as follows:

$$\delta = 0.2, \quad \sigma = 0.4, \quad \varepsilon = 10^{-6}, \quad NI \le 9999.$$

We stop the iteration if the $NI \le 9999$, three CG methods are compared in numerical performance and the numerical results are given in Table 1.

In Table 1, a pair of numbers means the number of iterations and the number of functional evaluations. Obviously, for the same function, the NWF method has fewer iterations and the calculation seems to be more accurate. This means the new algorithm has good performance for solving unconstrained optimization problems. In particular, the hybrid CG method of FR and WYL seems to be the best one among the three

TABLE 1. Number of iterations and number of functional evaluations

Function	FR	WYL	NWF
Booth	21\5.9384e-10	19\4.3694e-10	18\3.2853e-09
Matayas	156\1.2151e-07	173\1.1748e-07	88\4.0745e-07
Testf1	14\−4.0000	8\−4.0000	8\−4.0000
Dixon-price	21\7.0838e-10	39\6.0448e-10	19\3.7934e-10
Rosenbrock	26\1.0759e-12	23\4.5433e-06	17\2.3105e-10
Srn	5\2.0000	7\2.0000	7\2.0000
Simionescu	76\4.4353e-08	76\4.4353e-08	43\4.4353e-08
Testf2	16\−0.0937	17\−0.0937	40\−0.0937
Step	9\1.0280e-11	11\1.0890e-10	7\1.7885e-10
Trid	11\−2.0000	10\−2.0000	8\−2.0000

algorithms because it uses the least number of iterations and functional evaluations when the algorithms reach the same precision. All numerical results show that the efficiency of NWF method is encouraging.

5. **Conclusions.** In this paper, based on the FR and WYL methods, we present a new improved NWF method. The global convergence with the WWP line search is established. Numerical experiments show that the new algorithm is superior to the other two algorithms.

Acknowledgment. Many people have made invaluable contributions, both directly and indirectly to our research. We would like to express our warmest gratitude to our teachers and the authors grateful for the supports of National Natural Science Foundation of China (Nos. 51675461, 11673040).

REFERENCES

[1] E. Polak, *Algorithms and Consistent Approximations*, Springer, Berlin, 1997.
[2] W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, *Pacific Journal of Optimization*, vol.2, no.1, pp.35-58, 2006.
[3] R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Stand*, vol.49, no.6, pp.409-436, 1952.
[4] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients, *Computer Journal*, vol.7, no.2, pp.149-154, 1964.
[5] E. Polak and G. Ribiere, Note sur la convergence de methods de directions conjures, *Revue Francaise Information Recherche Operationnelle*, vol.16, no.16, pp.35-43, 1969.
[6] B. T. Polyak, The conjugate gradient method in extremal problems, *USSR Computational Mathematics & Mathematical Physics*, vol.9, no.4, pp.94-112, 1969.
[7] Y. H. Dai and Y. Yuan, *A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property*, Society for Industrial and Applied Mathematics, 1999.
[8] M. J. D. Powell, Convergence properties of algorithms for nonlinear optimization, *SIAM Review*, vol.28, no.4, pp.487-500, 1986.
[9] Y. Wu, A modified three-term PRP conjugate gradient algorithm for optimization models, *Journal of Inequalities & Applications*, vol.97, 2017.
[10] X. Du, P. Zhang and W. Ma, Some modified conjugate gradient methods for unconstrained optimization, *Journal of Computational & Applied Mathematics*, vol.305, pp.92-114, 2016.
[11] P. Gao and C. He, An efficient three-term conjugate gradient method for nonlinear monotone equations with convex constraints, *Calcolo*, vol.55, no.4, 2018.
[12] S. Babaie-Kafaki and R. Ghanbari, An optimal extension of the Polak-Ribière-Polyak conjugate gradient method, *Numerical Functional Analysis & Optimization*, vol.38, no.4, pp.826-850, 2017.
[13] G. Zoutendijk, Nonlinear programming, computational methods, *Integer & Nonlinear Programming*, vol.143, pp.37-86, 1970.
[14] Z. Wei, S. Yao and L. Liu, The convergence properties of some new conjugate gradient methods, *Applied Mathematics & Computation*, vol.183, no.2, pp.1341-1350, 2006.

[15] J. Yin, L. Wang and X. Jiang, A modified PRP conjugate gradient method with Armijo line search for large-scale unconstrained optimization, *IEEE the 36th Chinese Control Conference*, Dalian, China, 2017.
[16] X. Li et al., A modified conjugate gradient algorithm with backtracking line search technique for large-scale nonlinear equations, *International Journal of Computer Mathematics*, vol.95, no.1, pp.1-14, 2017.
[17] Y. H. Dai and Y. Yuan, An efficient hybrid conjugate gradient method for unconstrained optimization, *Annals of Operations Research*, vol.103, nos.1-4, pp.33-47, 2001.
[18] N. Andrei, Hybrid conjugate gradient algorithm for unconstrained optimization, *Journal of Optimization Theory and Applications*, vol.141, pp.249-264, 2009.
[19] S. Babaie-Kafaki and R. Ghanbari, *A Hybridization of the Polak-Ribière-Polyak and Fletcher-Reeves Conjugate Gradient Methods*, Springer-Verlag New York, Inc., 2015.
[20] L. Zhang, W. Zhou and D. H. Li, A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence, *IMA Journal of Numerical Analysis*, vol.26, no.4, pp.629-640, 2006.
[21] Z. Wei, G. Li and L. Qi, New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems, *Applied Mathematics & Computation*, vol.179, no.2, pp.407-430, 2006.
[22] J. C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM Journal on Optimization*, vol.2, no.1, pp.21-42, 2016.