Indonesian Journal of Electrical Engineering and Computer Science
Vol. 11, No. 3, September 2018, pp. 1188~1193
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v11.i3.pp1188-1193

Global Convergence of a New Coefficient Nonlinear Conjugate Gradient Method

Nur Syarafina Mohamed[1], Mustafa Mamat[2], Mohd Rivaie[3], Shazlyn Milleana Shaharuddin[4]
[1]Universiti Kuala Lumpur, Malaysian Institute of Industrial Technology, Malaysia
[2]Fakulti Informatik dan Komputeran, Universiti Sultan Zainal Abidin (UNISZA), Malaysia
[3]Jabatan Sains Komputer dan Matematik, Universiti Teknologi Mara (UiTM), Malaysia
[4]Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, Malaysia

Article Info

Article history:

Received May 13, 2018
Revised Jun 14, 2018
Accepted Jun 28, 2018

Keywords:

Conjugate gradient method
Global convergence
Strong wolfe-powell line search

ABSTRACT

Nonlinear conjugate gradient (CG) methods are widely used in optimization field due to its efficiency for solving a large scale unconstrained optimization problems. Many studies and modifications have been developed in order to improve the method. The method is known to possess sufficient descend condition and its global convergence properties under strong Wolfe-Powell search direction. In this paper, the new coefficient of CG method is presented. The global convergence and sufficient descend properties of the new coefficient are established by using strong Wolfe-Powell line search direction. Results show that the new coefficient is able to globally converge under certain assumptions and theories.

Corresponding Author:

Nur Syarafina Mohamed,
Universiti Kuala Lumpur,
Malaysian Institute of Industrial Technology, Malaysia.
Email: nursyarafina@unikl.edu.my

1. INTRODUCTION

Conjugate gradient (CG) method is used in finding the minimum value for unconstrained optimization problem. This method can be expressed in general form such;

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

where $f : R^n \to R$ is a continuously differentiable nonlinear function. $g(x)$ is denoted as a gradient of the function. Equation (1) can be solved by using several methods such Steepest Descent and Newton method but CG method is the most preferred due to its simplicity [1]. The nonlinear CG method generates a sequence of $\{x_k\}$ by using the recurrence;

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

where x_k is the current iterative point and $x_0 \in R^n$ is set to be a starting point of the sequence. From (2), $\alpha_k > 0$ is known as a step size and d_k is the search direction defined by the rule:

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

The most common technique in inexact line search used widely is strong Wolfe-Powell line search, where $\alpha_k > 0$ satisfies

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \,, \; \left| g(x_k + \alpha_k d_k)^T \leq \sigma \left| g_k^T d_k \right| \right| \tag{4}$$

with $0 < \delta < \sigma < 1$ are both constants. Distinct choice of the parameter β_k yields different numerical performance. Past study has shown at least six formulae for β_k, which are given as follows;

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} \tag{5}$$

$$\beta_k^{FR} = \frac{g_k^T g_k}{\|g_{k-1}\|^2} \tag{6}$$

$$\beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} \tag{7}$$

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}} \tag{8}$$

$$\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}} \tag{9}$$

$$\beta_k^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{\|d_{k-1}\|^2} \tag{10}$$

From (5)-(10), g_k and g_{k-1} are the abbreviations of $g(x_k)$ and $g(x_{k-1})$ which are the gradients of $f(x)$ at points x_k and x_{k-1} respectively. $\|\bullet\|$ denotes the Euclidean norm of the vectors. From (5)-(10), the above corresponding methods are respectively called as HS [2], FR [3], PRP [4], CD [5], DY [6], and lastly RMIL denotes for Rivaie, Mustafa, Ismail and Leong [7]. Zoutendijk [8] has proved the FR method in (6) to be globally converged under exact line search on a general function [9]. After that, [10] extended the result under strong Wolfe-Powell line search. A new modification of a conjugate gradient method is presented but it did not prove the global convergence under inexact line search though it has possess global convergence properties under exact line search [7&11]. The strong Wolfe-Powell line search is considered due to the higher cost of exact line search [12].

In this paper, a new coefficient with more simple β_k is proposed. Section two will elaborate the motivation of the coefficient together with the algorithm. Section three will discuss and prove the sufficient descend condition and global convergence properties. Finally, conclusion and recommendation for future study are wrapped up at section four.

2. NEW CG COEFFICIENT

The new CG coefficient introduced is known as β_k^{SMR}. β_k^{SMR} is motivated mainly from [7] where the denominator is retained as same as in (10). Whilst, the nominator in (10) is given as $g_k^T (g_k - g_{k-1})$ which is as same as used in (5) and (7). During expansion, the nominator becomes $g_k^T g_k - g_k^T g_{k-1}$ which implies

$\|g_k\|^2 - (g_k^T g_{k-1})$. Choosing the right nominator is important due to its role as a restart property in avoiding problems associated with jamming, [13-14]. Preventing any negative value, modifications has been made [15]; hence the new CG coefficient and the simplified version are as follows;

$$\beta_k^{SMR} = \max\left\{0, \frac{\|g_k\|^2 - |g_k^T g_{k-1}|}{\|d_{k-1}\|^2}\right\} \tag{11}$$

Before proceeds with more details steps, β_k^{SMR} needs to be simplified;

$$\beta_{k+1}^{SMR} = \begin{cases} 0 & \text{for} \quad \|g_{k+1}\|^2 < |g_{k+1}^T g_{k-1}| \\ \dfrac{\|g_{k+1}\|^2 - |g_{k+1}^T g_k|}{\|d_k\|^2} \leq \dfrac{\|g_{k+1}\|^2}{\|d_k\|^2} \end{cases} \quad 0 \leq \beta_{k+1}^{SMR} \leq \frac{\|g_{k+1}\|^2}{\|d_k\|^2} \tag{12}$$

Algorithm 2.1: Conjugate Gradient Method

A complete algorithm of CG method could be generated as follows:

Step 1: Initialization. Set $k = 0$ and select $x_0 \in \Re^n$, $d_0 = -g_0$, if $g_0 = 0$, stop.

Step 2: Based on (11), compute β_k^{SMR} .

Step 3: Compute search directions d_k based on (3). If $\|g_k\| \leq \varepsilon$, then stop. Otherwise, go the next step.

Step 4: Based on (4), solve for α_k .

Step 5: Updating new initial point using (2). If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| \leq \varepsilon$ then, stop. Otherwise go to Step 3 with $k = k+1$.

3. THEORETICAL ANALYSIS

This section discussed and analysed the sufficient descent property for the new coefficient under strong Wolfe-Powell line search direction. Before proceed, let assume that $g_k \neq 0$ for all k or else, the stationary point has been found. For any iterative method to be globally convergent, it is important to suffice its descent property, that is;

$$g_k^T d_k < 0.$$

3.1. Sufficient Descent Property

Before proceed, let assume that $g_k \neq 0$ for all k or else, the stationary point has been found. For any iterative method to be globally convergent, it is important to suffice its descent property, that is;

$$g_k^T d_k \leq -c\|g_k\|^2 \tag{13}$$

where c is a positive constant, is crucial to ensure the global convergences of the nonlinear conjugate gradient method under strong Wolfe-Powell line search direction [16]. Sufficient descent property is important to show that the function $f(x)$ can be reduces along the search direction. The proving steps below are modified from [11-12].

Theorem 3.1

If g_k and d_k are generated by algorithm 2.1 with $\sigma < \frac{6}{25}$, then, for all $k \geq 0$, it becomes;

$$\frac{\|g_k\|}{\|d_k\|} \le \frac{5}{3} \quad (14)$$

Proof

The proving steps are performed by inductions. For $k = 0$ and $\frac{\|g_0\|}{\|d0\|} = 1 < \frac{5}{3}$, hence (14) holds for $k = 0$.

Suppose for some $k \ge 0$, (14) holds true. Rearrange (3) and multiplying it with g_{k+1}^T, then

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

From strong Wolfe-Powell condition and absolute values properties in (4), expression (15) becomes;

$$\|g_{k+1}\|^2 \le \left| -g_{k+1}^T d_{k+1} \right| + \left| \beta_{k+1} g_{k+1}^T d_k \right| \; \|g_{k+1}\|^2 \le \left| g_{k+1}^T d_{k+1} \right| + \sigma \left| \beta_{k+1} \right| \left| g_k^T d_k \right| \quad (16)$$

Since $\beta_{k+1}^{SMR} \ge 0$, then

$$\|g_{k+1}\|^2 \le \left| g_{k+1}^T d_{k+1} \right| + \sigma \beta_{k+1} \left| g_k^T d_k \right| \quad (17)$$

By using Cauchy inequalities and substituting (12) in (17),

$$\|g_{k+1}\|^2 \le \|g_{k+1}\| \|d_{k+1}\| + \sigma \frac{\|g_{k+1}\|^2}{\|d_k\|^2} \|g_k\| \|d_k\| \quad (18)$$

Implies;

$$\|g_{k+1}\|^2 \le \|g_{k+1}\| \|d_{k+1}\| + \sigma \|g_{k+1}\|^2 \frac{\|g_k\|}{\|d_k\|} \quad (19)$$

Applying the induction hypothesis in (14),

$$\|g_{k+1}\|^2 \le \|g_{k+1}\| \|d_{k+1}\| + \frac{5}{3}\sigma \|g_{k+1}\|^2 \quad \|g_{k+1}\|^2 (1 - \frac{5}{3}\sigma) \le \|g_{k+1}\| \|d_{k+1}\| \quad \frac{\|g_{k+1}\|}{\|d_{k+1}\|} \le \frac{3}{(3 - 5\sigma)} \quad (20)$$

Therefore, if $\sigma < \frac{6}{25}$, then $\frac{\|g_{k+1}\|}{\|d_{k+1}\|} \le \frac{5}{3}$. Hence, (14) is true for $k+1$. The proof is completed.

3.2. Global Convergence Properties

The following assumption is needed in order to proceed with the proof of global convergence properties. The proof modifications are from [11-12,17-19]

Assumption 4.1

1) f is bounded below on the level set R^n and is continuous and differentiable in a neighborhood N of the level set $\ell = \{x \in R^n \mid f(x) \le f(x_0)\}$ at the initial point x_0.

2) The gradient $g(x)$ is Lipschitz continuous in N, so there exists a constant $L > 0$ such that;

$$\|g(x) - g(y)\| \le L\|x - y\| \quad \textit{for any} \;\; x, y \in N \quad (21)$$

From (11) and (13),

$$\left|\beta_k^{SMR}\right| \le \begin{cases} \dfrac{\|g_k\|^2}{\|d_{k-1}\|^2} & \text{if } g_k^T d_k \le 0 \\ \dfrac{\|g_k\|}{\|d_{k-1}\|} & \text{if } g_k^T d_k > 0 \end{cases} \tag{22}$$

Theorem 4.2

Suppose that Assumption 4.1 holds. Consider any CG method in the form of (2) and (3) where α_k is obtained from (4). If the descend condition holds, then;

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \tag{23}$$

Proof

To prove Theorem 4.2, contradiction method is used. That is, if Theorem 4.2 is not true, then there exists a constant $\varepsilon > 0$, such that;

$$\|g_k\| \ge \varepsilon \tag{24}$$

Rewriting (3) as $d_{k+1} + g_{k+1} = \beta_{k+1} d_k$ and squaring both sides the equation;

$$\|d_{k+1}\|^2 = -\|g_{k+1}\|^2 - 2g_{k+1}^T d_{k+1} + (\beta_{k+1})^2 \|d_k\|^2 \tag{25}$$

From (22), if $g_k^T d_k \le 0$, then;

$$\|d_{k+1}\|^2 = -\|g_{k+1}\|^2 - 2g_{k+1}^T d_{k+1} + \frac{\|g_{k+1}\|^4}{\|d_k\|^2} \tag{26}$$

Divide (26) by $\|g_{k+1}\|^4$ and from (13),

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \le \frac{1}{\|g_{k+1}\|^2} + \frac{2c}{\|g_{k+1}\|^2} \tag{27}$$

Suppose that (23) does not hold, then there exists $\varepsilon > 0$ such that (24) holds for all $k \ge 0$

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \le \frac{1}{\varepsilon^2} + \frac{2c}{\varepsilon^2} = \frac{2c+1}{\varepsilon^2}, \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \ge \frac{\varepsilon^2}{2c+1}, \sum_{k+0}^{\infty} \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \ge \infty \tag{28}$$

Also, from (22), if $g_k^T d_k > 0$, then,

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \le \frac{1}{\|g_{k+1}\|^2} + \frac{c}{\|g_{k+1}\|^2}, \frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \le \frac{1}{\varepsilon^2} + \frac{c}{\varepsilon^2} = \frac{c+1}{\varepsilon^2},$$
$$\frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \ge \frac{\varepsilon^2}{c+1}, \sum_{k+0}^{\infty} \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \ge \infty \tag{29}$$

From (28) and (29), this shows that (23) holds. The proof is completed.

4. CONCLUSION

By taking a little modification at β_k^{RMIL} in (10), a new coefficient β_k^{SMR} of the conjugate gradient method has been proposed. Results showed that the new coefficient satisfy the sufficient descent conditions and converge globally under strong Wolfe-Powell line search. It is proved that the algorithm is practical and effective to be used. Following with the proving provided, the coefficient will be tested on certain test function for future study.

REFERENCES

[1] Hamizah, N., Rivaie, M, Mamat, M. (2016). *A Modified Form of Conjugate Gradient Method for Unconstrained Optimization Problems*. *AIP* Conference Proceedings, 1739, 020076.

[2] Hestenes, M. R, Stiefel, E. (1962). Method of Conjugate Gradient for Solving Linear Equations. *J. Res. Nat. Bur. Stand*, 49:409-436.

[3] Fletcher, R, Reeves, C. (1964). Function minimization by conjugate gradients. *Computational Journal*, 7:149-154.

[4] Polak, E, Ribiere, G. (1969). Note on The Convergence of Conjugate Directions. *Rev. Francaise Informat Recherche Operationelle, 3E Annee*, 16: 35-43.

[5] Fletcher, R. (1987). *Practical Methods of Unconstrained Optimization*. New York: J. Wiley and Sons.

[6] Dai, Y. H, Yuan, Y. (2000). A Nonlinear Conjugate Gradient with Strong Global Convergence Properties. *SIAM Journal Optimization*, 10: 177-182.

[7] Rivaie, M., Mamat, M., June, L. W, Mohd, I. (2012). A New Class of Nonlinear Conjugate Gradient Coefficients With Global Convergence Properties. *Applied Mathematics and Computational*, 218: 11323-11332.

[8] Zoutendijk, G. (1970). Nonlinear programming computational methods. *J. Abadie (Ed.), Integer and Nonlinear Programming*, North-Holland, Amsterdam, 37–86.

[9] Powell, M. J. D. (1984). *Non-convex minimization calculations and the conjugate gradient method*. *Lecture notes in mathematics*, 1066:122-141. Berlin: Springer.

[10] Al-Baali, M. (1985). Descent property and global convergence of Fletcher–Reeves method with inexact line search IMA. *Journal Numerical Analysis*, 5:121-124.

[11] Abderahman, A. A., Mamat, M., RIvaie, M, Omer. O. (2014). Global Convergence Analysis of a New Nonlinear Conjugate Gradient Coefficient with Strong Wolfe Line Search. *Journal of Quality Measurement and Analysis,* 10(1): 75-85.

[12] Abderahman, A. A., Mamat, M., RIvaie, M, Omer. O. (2014). The Proof of Sufficient Descent Condition for a New Type of Conjugate Gradient Methods. *AIP Conference Proceedings*, 1602, 296.

[13] Hajar, N., Mamat, M., Rivaie, M, Salleh, Z. (2015). A Combination of Polak-Ribiere and Hestenes-Stiefel Coefficient in Conjugate Gradient Method for Unconstrained Optimization. *Applied Mathematical Sciences,* 9(63):3131-3142.

[14] Rivaie, M., Abderahman, A., Mamat, M., I. Mohd. (2014). The Convergence Properties of a New Type of Conjugate Gradient Methods. *Applied Mathematical Sciences*, 8(1):33-34.

[15] Syarafina, N. M., Mamat, M, Rivaie, M. (2016). A New Coefficient of Conjugate Gradient Method for Unconstrained Optimization. *Jurnal Teknologi*. 78:6-4,131-136.

[16] Hamoda, M., Mamat, M., Rivaie, M, Salleh, Z. (2016). A Conjugate Gradient Method with Strong Wolfe-Powell Line Search for Unconstrained Optimization. *Applied Mathematical Science,* 10(15): 721-734.

[17] Dong, X.L., Liu, H., Xu, Y. L., Yang, X. M. (2015). Some Linear Conjugate Gradient Methods with Sufficient Descent Condition and Global Convergence. *Optimization Letters,* 9: 1421-1432.

[18] Can, L., Ling F, Xianglian, C. (2013). Global Convergence of a kind of Conjugate Gradient Method. *Telkomnika Indonesian Journal of Electrical Engineering,* 11(1): 544-549.

[19] Can, L. (2013). A Modified Conjugate Gradient Method for Unconstrained Optimization. *Telkomnika Indonesian Journal of Electrical Engineering,* 11(11): 6373-6380.