Journal of Mathematics and Statistics

Original Research Paper

Sufficient Descent Condition of the Polak-Ribière-Polyak (PRP) Conjugate Gradient Method without Line Search

[1]Abdelrahman A., [3]KH. I. Osman, [1]Osman O.O. Yousif, [2]Mogtaba Mohammed, [3]Arafa Dawood, [3]B. A. Younis

[1]*Department of Mathematics, University of Gezira, Sudan*
[2]*Department of Mathematics, College of Science, Majmaah University, Zulfi 11932, Saudi Arabia, Sudan*
[3]*Faculty of Sciences and Arts in Sarat Abeda, Department of Mathematics, King Khalid Universitty, Abha, Saudi Arabia*

Article history
Received: 01-09-2021
Revised: 10-11-2021
Accepted: 18-11-2021

Corresponding Author:
Awad Abdelrahman Abdalla
Department of Mathematics,
University of Gezira, Sudan
E-mail: awad.abdalla26@yahoo.com

Abstract: Nonlinear conjugate gradient methods for unconstrained optimization problems are used in many aspects of theoretical and applied sciences. They are iterative methods, so at any iteration a step length is computed using a method called line search. In most cases, the sufficient descent condition plays an important role to prove the global convergence of a conjugate gradient method. Due to its outperformance in practical computation, the Polak-Ribière-Polyak (PRP) conjugate gradient method is widely used for solving nonlinear unconstrained optimization problems. However, the sufficient descent condition of PRP has not established without line search yet. In this study, we established the sufficient descent condition without line search based on the conditions $0 < \beta_k^{PRP} \leq \xi \beta_k^{FR}$ and $1 < \beta_k^{PRP} \leq \mu \beta_k^{FR}$, where $0 < \xi < 1$ and $\mu > 1$. As a result, we found that under certain conditions, the sufficient descent condition is satisfied when the PRP implemented without line search.

Keywords: Conjugate Gradient Method, Unconstrained Optimization, Sufficient Descent Condition.

Introduction

The optimization formula that we utilized is as follows:

$$\min_{x \in R^n} f(x), \tag{1.1}$$

where $f: R^n \to R$ is a nonlinear, continuously differentiable function, and its gradient is denoted by $g(x)$, which should be available, when applied to solve the problem (1.1), starting from an initial point $x_1 \in R^n$ and follows the iteration formula:

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, ..., \tag{1.2}$$

where $\alpha_k > 0$ is a step-size. The step-size is determined by some line search and d_k is the search direction defined by:

$$d_k = \begin{cases} -g_k & if\ k = 0, \\ -g_k + \beta_k d_k - 1 & if\ k \geq 1, \end{cases} \tag{1.3}$$

where, $g_k = \nabla f(x_k)$ is the gradient of $f(x_k)$ and β_k is a scalar. The famous classical CG methods formulas for β_k are:

$$\beta_k^{HS} = \frac{g_K^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}, \tag{1.4}$$

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, \tag{1.5}$$

$$\beta_k^{PRP} = \frac{g_K^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}, \tag{1.6}$$

$$\beta_k^{CD} = \frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}, \tag{1.7}$$

$$\beta_k^{LS} = \frac{g_k^T (g_k - gk - 1)}{d_{k-1}^T g_{k-1}}, \tag{1.8}$$

$$\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1}) d_{k-1}^T}. \tag{1.9}$$

The other types of the conjugate gradient methods can be obtained based on the corresponding different choices

DOI: 10.3844/jmsssp.2022.11.15

for important scalars β_k. Therefore, to get an efficiency in practical computation and robust convergence, the scalar must satisfy the global convergence and obtained choicely performance in computation. Historically, the linear Conjugate Gradient (CG) method had been suggested by Hestenes and Stiefel (1952); for solving symmetric positive definite linear systems of equations in Hestenes and Stiefel (1952) and nonlinear conjugate gradient scalar is proposed for solving linear systems of equations, independent, that is:

$$\beta_k^{HS} = \frac{g_k^T\left(g_k - g_{k-1}\right)}{\left(g_k - g_{k-1}\right)T_{d_{k-1}}}. \tag{1.10}$$

Polak and Ribière (1969; Polyak, 1969), proposed nonlinear conjugate gradient scalar for solving unconstrained independent optimization problems, that is given as the following:

$$\beta_k^{PRP} = \frac{g_k^T\left(g_k - g_{k-1}\right)}{\left\|g_{k-1}\right\|^2}. \tag{1.11}$$

The PRP and HS methods have excellent performance in practical computation, due to their possession of an approximate restart feature when jamming occurs. Nevertheless, their convergence properties don't seem to be perfect(Jiang and Jian, 2019). The global convergence of the PRP method with exact line search has been established by (Polak and Ribière, 1969) under a powerful convexity assumption for *f*. The numerical performance of the FR method (Fletcher and Reeves,1964); conjugate gradient method has often been much slower than that of the PRP conjugate gradient method (Awad Abdelrahman, 2020). (Gilbert and Nocedal, 1992), proceeded classificatory analysis and an exhibited that the PRP method is globally convergent if PRP is constrained to be non-negative and α_k is determined by a line search step satisfying the sufficient descent condition $g_K^T d_k \le -c\left\|g_k\right\|^2$. According to, the standard Wolfe conditions (Li and Li, 2011), proposed two derivative-free approaches stood on modified PRP conjugate gradient techniques for solving large-scale nonlinear equations. Moreover, the numerical efficiency of the PRP method is principally associated with an automatic restart feature that avoids jamming (Powell, 1977).

In this study, in order to obtain sufficient descent condition without line search, we do a little analysis on the method of PRP (1.11).

Analysis Attributes of PRP Method

Before narrating the attributes of PRP scalar, we introduce the following assumption, which is a very important to study the functions behavior.

Assumption A

f(x) is bounded below on the level set on R^n and is continuously and differentiable in a neighborhood *N* of the level set $\Gamma = \{x \in R^n | f(x) \le f(x_0)\}$ at the initial point x_0, there exists a constant $B > 0$ such that

$$\left\|X - y\right\| \le B, \forall x, y \in N. \tag{2.1}$$

The gradient $g(x) = \nabla f(x)$ is Lipschitz continuous in N, so a constant $L > 0$ exists, such that;

$$\left\|g(x) - g(y)\right\| \le L\left\|x - y\right\|, \textit{ for any } x, y\varepsilon N \tag{2.2}$$

By using the Assumption of *f*, there exists a constant $\tau \ge 0$ such that:

$$\left\|g(x)\right\| \le \tau\, \forall x \in \Gamma. \tag{2.3}$$

The following is significant attributes of PRP (1.11) method, which have been utilized in establishing the sufficient descent condition in the intervals (0,1) and (1,∞).

$$\beta_k^{PRP} = \frac{g_K^T\left(g_k - g_{k-1}\right)}{\left\|g_{k-1}\right\|^2} = \beta_k^{FR}\left(1 - \frac{g_k^T g_{k-1}}{\left\|gk\right\|^2}\right). \tag{2.4}$$

From (2.4), if $g_K^T g_{k-1} \le 0$, then:

$$\le \frac{-g_K^T g_{k-1}}{\left\|g_{k-1}\right\|^2} \le \frac{\left\|g_k\right\|\left\|g_{k-1}\right\|}{\left\|gk\right\|^2} = \frac{\left\|g_{k-1}\right\|}{\left\|gk\right\|},$$

$$1 - \frac{-g_K^T g_{k-1}}{\left\|gk\right\|^2} \le 1 + \frac{\left\|g_{k-1}\right\|}{gk},$$

From above inequality and (2.4), we have:

$$0 \le \beta_K^{PRP} \le \beta_k^{FR} + \frac{\left\|g_k\right\|^2}{\left\|g_{k-1}\right\|^2} = \frac{\left\|g_{k-1}\right\|}{\left\|g_k\right\|} = \beta_K^{FR}\left(1 + \frac{\left\|g_{k-1}\right\|}{\left\|g_k\right\|}\right),$$
$$0 \le \beta_K^{PRP} \mu \beta_k^{FR} \textit{ , where, } \mu > 1, \mu = 1 + \frac{\left\|g_{k-1}\right\|}{\left\|g_k\right\|}. \tag{2.5}$$

From (2.5), β^{PRP} is positive and belong to the interval (1,∞) and also.
$-g_K^T g_{k-1} \ge \delta^2 > 0$, where $\delta \in (0, 1)$ and from (2.3):

$$0 \le \frac{\left\|g_k\right\|}{\left\|g_{k-1}\right\|} \le \frac{\left\|g_k\right\|^2}{-g_k^T g_{k-1}} \le \frac{\tau^2}{\delta^2}. \tag{2.6}$$

From (2.4) and if $g_k^T g_{k-1} > 0$, then:

$$0 < \frac{g_k^T g_{k-1}}{\|g_k\|^2} \le \frac{\|g_k\|\|g_{k-1}\|}{\|g_k\|^2} = \frac{\|g_{k-1}\|}{\|g_k\|},$$
$$0 < \gamma \le \frac{g_k^T g_{k-1}}{\|g_k\|^2} \le \frac{\|g_k\|\|g_{k-1}\|}{\|g_k\|^2} = \frac{\|g_{k-1}\|}{\|g_k\|},$$

where, γ is small real number, $\gamma \in (0,1)$:

$$1-\gamma \ge 1 - \frac{g_k^T g_{k-1}}{\|g_k\|^2} \ge 1 - \frac{\|g_{k-1}\|}{\|g_k\|},$$

Multiply the above inequality by β_k^{FR}, we obtain:

$$\beta_k^{FR}\left(1-\frac{\|g_{k-1}\|}{\|g_k\|}\right) \le \beta_k^{FR}\left(1-\frac{g_k^T g_{k-1}}{\|g_k\|^2}\right) \le \beta_k^{FR}(1-\gamma),$$
$$-\beta_k^{FR} \le \beta_k^{FR} - \frac{\|g_k\|}{\|g_{k-1}\|} \le \beta_k^{PRP} \le \beta_k^{FR}(1-\gamma) \le \beta_k^{FR} \tag{2.7}$$

$$\varsigma\,\beta_k^{FR} \le \beta_k^{PRP} \le \xi\beta_k^{FR}, where\, \varsigma < 0\, 0 < \xi < 1. \tag{2.8}$$

Observe that in case of $g_k^T g_{k-1} > 0,$ the β_k^{PRP} is in the interval (0,1) and the PRP method may not satisfying the sufficient descent condition in this interval (Gilbert and Nocedal, 1992); So we try to overcome that defect by obtaining the sufficient descent condition without line search in this interval. Furthermore, from (2.7), we obtain the following results:

$$-\frac{\|g_k\|^2}{\|g_{k-1}\|^2} \le \frac{\|g_k\|^2}{\|g_{k-1}\|^2} - \frac{\|g_k\|}{\|g_{k-1}\|} \le \frac{\|g_k\|^2 - g_k^T g_{k-1}}{\|g_{k-1}\|^2} \le \frac{\|g_k\|^2}{\|g_{k-1}\|^2}(1-\gamma),$$

Multiply the above inequality by $\|g_{k-1}\|^2$ we obtain:

$$-\|g_k\|^2 \le \|g_k\|^2 - \|g_k\|\|g_{k-1}\| \le \|g_k\|^2 - g_k^T g_{k-1} \le \|g_k\|^2(1-\gamma),$$

Subtractions the above inequality by $\|g_k\|^2$, we obtain:

$$-2\|g_k\|^2 \le -\|g_k\|\|g_{k-1}\| \le -g_k^T g_{k-1} - \gamma\|g_k\|^2,$$

From above inequality, we get:

$$\gamma\|g_k\| \le \|g_{k-1}\| \le 2\|g_k\|. \tag{2.9}$$

and:

$$0 < g_k^T g_{k-1} \le 2\|g_k\|^2. \tag{2.10}$$

Sufficient Descent Condition

The following theorems present the proof of the sufficient descent condition for the PRP method without using the line search.

Theorem 3.1. Let $g_k^T g_{k-1} \le 0$, where $\beta_k = \beta_k^{PRP} \le \mu\beta_k^{FR}$ if μ >1 given in (2.5). For any $X_1 \in R^n$ consider the sequence $\{x_k\}$, which is generated by (1.2) and (1.3), then the following inequality holds:

$$g_k^T d_k \le -c\|g_k\|^2\ \forall k \ge 0, \tag{3.1}$$

where:

$$c = 2 - \frac{\delta^2}{\delta^2 - \mu\tau^2}.$$

Proof. We prove this Theorem by induction. We begin proving the descent condition $g_K^T d_k \le 0$ as follows. For $k = 0$, is true $g_0^T d_0 = -\|g_0\|^2 \le 0,$ supposing that $g_i^T d_i < 0$ holds for $i \le k-1$ we deduce that the sufficient descent condition holds by proving that $g_i^T d_i < 0$ holds for $i = k$ as follow. Consider $i = k$, we have the following properties without line search:

$$d_k = -g_k + \beta_k d_{k-1}. \tag{3.2}$$

We multiply (3.2) by g^{k-1}, to obtain:

$$g_{k-1}^T d_k = -g_k^T g_{k-1} + \beta_k g_{k-1}^T d_{k-1},$$
$$g_{k-1}^T d_k \left(\frac{\|g_k\|^2}{\|g_k\|^2}\right) = -g_k^T g_{k-1} + \beta_k g_{k-1}^T d_{k-1},$$
$$\frac{g_k^T d_k}{\|g_k\|^2}\left(g_k^T g_{k-1}\right) = -g_k^T g_{k-1} + \beta_k g_{k-1}^T d_{k-1}, \tag{3.3}$$
$$\frac{g_k^T d_k}{\|g_k\|^2} = -1 + \frac{\beta_k g_{k-1}^T d_{k-1}}{g_k^T g_{k-1}}.$$

Also we multiply (3.2) by g_k ' we get:

$$g_k^T d_k = -\|g_k\|^2 + \beta_k g_k^T d_{k-1},$$
$$\frac{g_k^T d_k}{\|g_k\|^2} = -1 + \frac{\beta_k g_k^T d_{k-1}}{\|g_k\|^2} \tag{3.4}$$

From (3.3) and (3.4), we have:

$$-1+\frac{\beta_k g_{k-1}^T d_{k-1}}{g_k^T g_{k-1}} = -1+\frac{\beta_k g_{k-1}^T d_{k-1}}{\|g_k\|^2},$$
$$\frac{g_{k-1}^T d_{k-1}}{g_k^T g_{k-1}} = \frac{g_{k-1}^T d_{k-1}}{\|g_k\|^2} \tag{3.5}$$

We substitute (3.5) in (3.4) and from (2.5), we get:

$$\frac{g_k^T d_k}{\|g_k\|^2} -1+\frac{\beta_k g_{k-1}^T d_{k-1}}{g_k^T g_{k-1}} = \leq -1+\frac{\mu\|g_k\|^2}{\|g_{k-1}\|^2}\frac{g_{k-1}^T d_{k-1}}{g_k^T g_{k-1}}$$
$$\leq -1+\mu\frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2}\left(\frac{\|g_k\|^2}{g_k^T g_{k-1}}\right), \tag{3.6}$$

From (3.6) and (2.6), we obtain:

$$\frac{g_k^T d_k}{\|g_k\|^2} = -1+\mu\frac{\tau^2}{\delta^2}\frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2},$$

Repeating this process and using the fact that $g_k^T d_1 = -\|g_1\|^2$, which implies that:

$$\frac{g_k^T d_k}{\|g_k\|^2} = -2+\frac{\delta^2}{\delta^2-\mu\tau^2}.$$

Therefore, we can deduce that (3.1) holds for $k \geq 1$.

Theorem 3.2. Let $g_k^T g_{k-1} > 0$ where, $\beta_k = \beta_k^{PRP} \leq \xi\, \beta_K^{FR}$, such that $0 < \xi < 1$ is given as (2.8). For any $x_1 \in R^n$, consider the sequence $\{x_k\}$ generated by (1.2) and (1.3), then the following inequality holds:

$$g_k^T d_k \leq -c\|g_k\|^2 \; \forall k \geq 0, \tag{3.7}$$

where:

$$c = \leq 2-\frac{r^2}{r^2-2\xi}.$$

Proof. We prove this Theorem by induction. We begin to prove the descent condition $g_k^T d_k < 0$ as follows. For $k = 0$, is true $g_0^T d_0 = -\|g_0\|^2 \leq 0$, supposing that $g_i^T d_i < 0$ holds for $\leq k$ -1,we deduce that the sufficient descent condition holds by proving that $g_i^T d_i < 0$ holds for $i = k$ as follow. Consider $i = k$, we have the following properties without line search.

We multiply (3.2) by g_k we obtain:

$$g_k^T d_k = -\|g_k\|^2 + \beta_k g_k^T d_{k-1},$$
$$g_k^T d_k \leq -\|g_k\|^2 + \xi\beta_k^{FR} g_k^T d_{k-1} = -\|g_k\|^2 + \xi\beta_k^{FR} g_k^T d_{k-1}\left(\frac{\|g_{k-1}\|^2}{\|g_{k-1}\|^2}\right),$$
$$g_k^T d_k \leq -\|g_k\|^2 + \xi\frac{\|g_k\|^2}{\|g_{k-1}\|^2} g_k^T g_{k-1}\frac{g_{k-1}^T}{\|g_{k-1}\|^2}.$$

From (2.9), (2.10) and the above inequality, we get:

$$g_k^T d_k \leq -\|g_k\|^2 + \frac{2\xi}{r^2}\|g_k\|^2\frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2},$$
$$\frac{g_k^T d_k}{\|g_k\|^2} \leq -1+\frac{2\xi}{r^2}\frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2}.$$

Repeating this process and using the fact that $g_1^T d_1 = \|g_1\|^2$, which implies that:

$$\frac{g_k^T d_k}{\|g_k\|^2} \leq -2+\frac{\gamma^2}{\gamma^2-2\xi}.$$

Therefore, we can deduce that (3.7) holds for $k \geq 1$.

Conclusion

In this study, based on (Polyak, 1969), we have established only sufficient descent condition without line search. The proofs of the sufficient descent condition under the following conditions $0 < \beta_k^{PRP} \leq \xi\, \beta_k^{FR}$ and $1 < \beta_k^{PRP} \leq \xi\, \beta_k^{FR}$, where $0 < \xi < 1$ and $\mu > 1$, for the method of PRP. Moreover, this study opened many keys for established the global convergence of the method of PRP in two cases.

Acknowledgement

The authors are grateful to the editor and the anonymous reviewers for their valuable comments and suggestions, which have substantially improved this study.

Authors' contributions

All authors contributed equally to the writing of this study. All authors read and approved the final version of this study.

Ethics

The authors declare that this article is original, contains unpublished material and no ethical issues involved.

References

Abdelrahman, A. (2020). A Robust Modification of Hestenes Stiefel Conjugate Gradient Method with Strong Wolfe Line Search, Journal of Mathematics and Statistics, 16(2020), 54-61.

Fletcher, R., & Reeves, C. M. (1964). Function minimization by conjugate gradients. The computer journal, 7(2), 149-154. doi.org/10.1093/comjnl/7.2.149

Gilbert, J. C., & Nocedal, J. (1992). Global convergence properties of conjugate gradient methods for optimization. SIAM Journal on optimization, 2(1), 21-42. doi.org/10.1137/0802003

Hestenes, M. R., & Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems (Vol. 49, No. 1). Washington, DC: NBS. https://www.math.unipd.it/~alvise/AN_2018/LETTURE/hestenes-stiefel.pdf

Jiang, X., & Jian, J. (2019). Improved Fletcher-Reeves and Dai–Yuan conjugate gradient methods with the strong Wolfe line search. Journal of Computational and Applied Mathematics, 348, 525-534. doi.org/10.1016/j.cam.2018.09.012

Li, Q., & Li, D. H. (2011). A class of derivative-free methods for large-scale nonlinear monotone equations. IMA Journal of Numerical Analysis, 31(4), 1625-1635. doi.org/10.1093/imanum/drq015

Polak, E., & Ribiere, G. (1969). Note sur la convergence de méthodes de directions conjuguées. ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique, 3(R1), 35-43. http://www.numdam.org/article/M2AN_1969__3_1_35_0.pdf

Polyak, B. T. (1969). The conjugate gradient method in extremal problems. USSR Computational Mathematics and Mathematical Physics, 9(4), 94-112. doi.org/10.1016/0041-5553(69)90035-44

Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. Mathematical programming, 12(1), 241-254. doi.org/10.1007/BF01593790