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

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

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

**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-Ribiére-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 $0 < \beta_{\text{PRP}} \leq \xi \beta_{\text{FR}}$ and $1 < \beta_{\text{PRP}} \leq \mu \beta_{\text{FR}}$, 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:

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

(1.1)

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $x_1 \in \mathbb{R}^n$ and follows the iteration formula:

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

(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:

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

(1.3)

where, $g_k = \nabla f(x_k)$ is the gradient of $f(x_k)$ and $\beta_k$ is a scalar.

The famous classical CG methods formulas for $\beta_k$ are:

$$
\beta_k^{\text{FR}} = \frac{g_k^T (g_k - g_{k+1})}{(g_k - g_{k-1}) d_{k-1}},
$$

(1.4)

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

(1.5)

$$
\beta_k^{\text{PRP}} = \frac{g_k^T (g_k - g_{k+1})}{g_{k-1}^T g_{k-1}},
$$

(1.6)

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

(1.7)

$$
\beta_k^{\text{FR}} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T g_{k-1}},
$$

(1.8)

$$
\beta_k^{\text{FR}} = \frac{g_k^T g_k}{(g_k - g_{k-1}) d_{k-1}^T},
$$

(1.9)

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

© 2022 Abdelrahman A., KH. I. Osman, Osman O.O. Yousif, Mogtaba Mohammed, Arafa Dawood, B. A. Younis. This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.
for important scalars $\beta_k$. 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:

$$
\beta^t_{HS} = \frac{g_k^T (g_k - g_{k+1})}{g_{k+1}^T g_{k+1}}.
$$

(1.10)

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

$$
\beta^t_{PRP} = \frac{g_k^T (g_k - g_{k+1})}{g_{k+1}^T g_{k+1}}.
$$

(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 Ribi`ere, 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 $c_k$ is determined by a line search step satisfying the sufficient descent condition $g_k^T d_k \leq -\|g_k\|^2$. 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 $\Gamma= \{x \in R^n \mid f(x) \leq f(x_0)\}$ at the initial point $x_0$, there exists a constant $B > 0$ such that

$$
\|X - y\| \leq B, \forall x, y \in N.
$$

(2.1)

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

$$
\|g(x) - g(y)\| \leq L\|x - y\|, \text{for any } x, y \in N
$$

(2.2)

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

$$
\|g(x)\| \leq \tau \forall x \in \Gamma.
$$

(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 $(1,\infty)$.

$$
\beta^t_{PRP} = \frac{g_k^T (g_k - g_{k+1})}{g_{k+1}^T g_{k+1}} = \beta^t_{c}\left(1 - \frac{g_k^T g_{k+1}}{\|g_k\|^2}\right).
$$

(2.4)

From (2.4), if $g_k^T g_{k+1} \leq 0$, then:

$$
-\frac{g_k^T g_{k+1}}{\|g_k\|^2} \leq \left[\frac{\|g_{k+1}\|}{\|g_k\|} - \frac{g_k^T g_{k+1}}{\|g_k\|^2}\right],
$$

and $1 - \frac{g_k^T g_{k+1}}{\|g_k\|^2} \leq 1 + \frac{\|g_{k+1}\|}{\|g_k\|}$.

From above inequality and (2.4), we have:

$$
0 \leq \beta^t_{PRP} \leq 0 + \|g_{k+1}\| = \|g_k\| = \beta^t_{c}\left(1 + \|g_{k+1}\| \|g_k\|^{-1}\right).
$$

(2.5)

From (2.5), $\beta^t_{PRP}$ is positive and belong to the interval $(1,\infty)$ and also. $-g_k^T g_{k+1} \geq \delta^2 > 0$, where $\delta \in (0, 1)$ and from (2.3):

$$
0 \leq \frac{g_k}{\|g_k\|} - \frac{g_{k+1}}{\|g_{k+1}\|} \leq \frac{\tau^2}{\delta^2}.
$$

(2.6)
From (2.4) and if $g^T_i g_{i-1} > 0$, then:

$$0 < g^T_i g_{i-1} \leq \frac{\|g_i\|}{\|g_{i-1}\|} \leq \frac{\|g_i\|}{\|g_{i-1}\|},$$

$$0 < \gamma \leq \frac{g^T_i g_{i-1} \leq \frac{\|g_i\|}{\|g_{i-1}\|}}{\|g_i\|} \leq \frac{\|g_i\|}{\|g_{i-1}\|}.$$

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

$$1 - \gamma \geq 1 - \frac{g^T_i g_{i-1} \leq \frac{\|g_i\|}{\|g_{i-1}\|}}{\|g_i\|},$$

Multiply the above inequality by $\beta_k^{FR}$, we obtain:

$$\beta_k^{FR} \left(1 - \frac{\|g_i\|}{\|g_{i-1}\|} \leq \frac{\|g_i\|}{\|g_{i-1}\|} \right) \leq \beta_k^{FR} (1 - \gamma),$$

$$-\beta_k^{FR} \leq \beta_k^{FR} - \frac{\|g_i\|}{\|g_{i-1}\|} \leq \beta_k^{FR} (1 - \gamma) \leq \beta_k^{FR}.$$

Observe that in case of $g^T_i g_{i-1} > 0$, the $\beta_k^{FR}$ 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:

$$\frac{\|g_i\|}{\|g_{i-1}\|} \leq \frac{\|g_i\|}{\|g_{i-1}\|} \leq \frac{\|g_i\|}{\|g_{i-1}\|} \leq \frac{\|g_i\|}{\|g_{i-1}\|} (1 - \gamma),$$

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

$$-\|g_i\| \leq \|g_i\| - \|g_{i-1}\| \leq \|g_i\| - \|g_{i-1}\| \leq \|g_{i-1}\| (1 - \gamma).$$

Subtractions the above inequality by $\|g_{i-1}\|$, we obtain:

$$-2\|g_i\| \leq -\|g_{i-1}\| \leq -\|g_{i-1}\| - \|g_{i-1}\| (1 - \gamma).$$

From above inequality, we get:

$$\gamma \|g_i\| \leq \|g_{i-1}\| \leq 2 \|g_i\|.$$

and:

$$0 < g^T_i g_{i-1} \leq 2\|g_i\|^2. \quad (2.10)$$

**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 $g^T_i g_{i-1} \leq 0$, where $\beta_k = \beta_k^{FR} \leq \mu \beta_k^{FR}$ if $\mu > 1$ given in (2.5). For any $X_i \in R^e$ consider the sequence $\{x_i\}$, which is generated by (1.2) and (1.3), then the following inequality holds:

$$g^T_i d_i \leq -\epsilon \|g_i\|^2 \forall k \geq 0, \quad (3.1)$$

where:

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

Proof. We prove this Theorem by induction. We begin proving the descent condition $g^T_i d_i \leq 0$ as follows. For $k = 0$, is true $g^T_0 d_0 = -\|g_0\|^2 \leq 0$, supposing that $g^T_i d_i < 0$ holds for $i \leq k - 1$ we deduce that the sufficient descent condition holds by proving that $g^T_i d_i < 0$ holds for $i = k$ as follow. Consider $i = k$, we have the following properties without line search:

$$d_k = -g_k + \beta_k d_{k-1}. \quad (3.2)$$

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

$$g^T_k d_k = -g_k^T g_{k-1} + \beta_k g^T_{k-1} d_{k-1},$$

$$g^T_k d_k \left(\begin{array}{c} g_k \\ g_{k-1} \end{array}\right) = -g^T_k g_{k-1} + \beta_k g^T_{k-1} d_{k-1}, \quad (3.3)$$

$$g^T_k d_k = -g^T_k g_{k-1} + \beta_k g^T_{k-1} d_{k-1}.$$

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

$$g^T_k d_k = -g^T_k g_{k-1} + \beta_k g^T_{k-1} d_{k-1},$$

$$g^T_k d_k = -g^T_k g_{k-1} + \beta_k g^T_{k-1} d_{k-1}. \quad (3.4)$$

$$g^T_k d_k = -g^T_k g_{k-1} + \beta_k g^T_{k-1} d_{k-1}.$$

$$g^T_k d_k = -g^T_k g_{k-1} + \beta_k g^T_{k-1} d_{k-1}.$$

$$g^T_k d_k = -g^T_k g_{k-1} + \beta_k g^T_{k-1} d_{k-1}.$$
From (3.3) and (3.4), we have:

\[ -1 + \beta_k \|g_k\| d_{k+1} = -1 + \beta_k \|g_k\| d_{k+1} , \]

(3.5)

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

\[ g_k^T d_k - 1 + \beta_k \|g_k\| d_{k+1} \leq -1 + \mu \|g_k\| \|g_k\| d_{k+1} \]

\[ \leq -1 + \mu \|g_k\| d_{k+1} \left( \|g_k\| \|g_k\| \right) . \]

(3.6)

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

\[ g_k^T d_k = -1 + \mu \frac{\|g_k\| d_{k+1}}{\|g_k\|} . \]

Repeating this process and using the fact that

\[ g_k^T d_k = -1 + \mu \frac{\|g_k\| d_{k+1}}{\|g_k\|} , \]

which implies that:

\[ g_k^T d_k = -2 + \frac{\|g_k\|^2}{\|g_k\|^2} . \]

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 \( x_k \in \mathbb{R}^n \), consider the sequence \( \{x_k\} \) generated by (1.2) and (1.3), then the following inequality holds:

\[ g_k^T d_k \leq -c \|g_k\| \forall k \geq 0 , \]

(3.7)

where:

\[ c = 2 - \frac{r^2}{r^2 - 2\xi} . \]

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 \( g_0^T d_0 = -\|g_0\| \leq 0 \). supposing that \( g_i^T d_i < 0 \) holds for \( i = 1 \) - \( k \), 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:

\[ g_k^T d_k = -1 + \beta_k \|g_k\| d_{k+1} , \]

\[ g_k^T d_k \leq -\|g_k\| + \xi \beta_k^{PRP} \|g_k\| d_{k+1} = -\|g_k\| + \xi \beta_k^{PRP} \|g_k\| d_{k+1} \left( \|g_k\| \|g_k\| \right) . \]

(3.8)

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

\[ g_k^T d_k \leq -\|g_k\| + \frac{2\xi}{\gamma^2} \|g_k\| \|g_k\| d_{k+1} . \]

Repeating this process and using the fact that \( g_k^T d_k = -\|g_k\|^2 \), which implies that:

\[ g_k^T d_k \leq -2 + \frac{\|g_k\|^2}{\|g_k\|^2} . \]

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


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


