Improve Performance of Fletcher-Reeves (FR) Method

Salah Gazi Shareef¹, Ahmed Anwer Mustafa²
¹Department of Mathematics, Faculty of Science, University of Zakho
²Department of Mathematics, Faculty of Science, University of Zakho
Kurdistan Region-Iraq

Abstract: Conjugate gradient (CG) methods are famous for solving nonlinear unconstrained optimization problems because they required low computational memory. In this paper, we propose a new conjugate gradient (βₖ⁰ₙᵉʷ) which possesses global convergence properties using exact line search and inexact line search. The given method satisfies sufficient descent condition under strong Wolfe line search. Numerical results based on the number of iterations (NOI) and number of function (NOF), have shown that the new βₖ⁰ₙᵉʷ performs better than Flecher-Reeves (FR) CG methods.

Keywords: Unconstrained optimizations, Conjugate gradient method, Sufficient Descent Condition, Global Convergent.

1. Introduction

The conjugate gradient method (CG) plays an important role in solving the unconstrained optimization problem. In general, the method has the following form:

\[ \text{Min } f(x) \quad (1.1) \]
\[ x \in \mathbb{R}^n \]

where, \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is continuously differentiable. The CG method is an iterative method of the form,

\[ x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \ldots \]  
\[ (1.2) \]

where \( x_k \) is the current iterate point, \( \alpha_k > 0 \) is a step size and \( d_k \) is the search direction. Basically \( d_k \) is defined by

\[ d_k = \begin{cases}  -g_k, & k = 0 \\ -g_{k+1} + \beta_k d_k, & k \geq 1 \end{cases} \]  
\[ (1.3) \]

where, \( g_k \) is the gradient of \( f(x) \) at the point \( x_k \). \( \beta_k \in \mathbb{R} \) is known as conjugate gradient and different \( \beta_k \) will yield different CG methods. Some well-known formulas are given as follows:

\[ \beta_{k}^{\text{HS}} = \frac{\gamma}{d_k^T y_k} \]  
\[ (1.4) \]
\[ \beta_{k}^{\text{FR}} = \frac{\gamma}{g_k^T y_k} \]  
\[ (1.5) \]
\[ \beta_{k}^{\text{PR}} = \frac{\gamma}{d_k^T g_k} \]  
\[ (1.6) \]
\[ \beta_{k}^{\text{DX}} = \frac{\gamma}{d_k^T g_k} \]  
\[ (1.7) \]
\[ \beta_{k}^{\text{BA}2} = \frac{\gamma}{d_k^T g_k} \]  
\[ (1.8) \]
\[ \beta_{k}^{\text{LS}} = \frac{\gamma}{d_k^T g_k} \]  
\[ (1.9) \]
\[
\beta_k^{DY} = \frac{g_k^{T}g_{k+1}}{d_k^Ty_k} \quad (1.10)
\]
\[
\beta_k^{RMIL} = \frac{g_k^{T}g_{k+1}}{d_k^T(d_k - g_{k+1})} \quad (1.11)
\]
\[
\beta_k^{AMRI} = \frac{\|g_{k+1}\|^2 - \|g_k + 1\|^2}{\|d_k\|^2} \quad (1.12)
\]

Where, \(g_k\) and \(g_{k+1}\) are the gradients of \(f(x)\) at the point \(x_k\) and \(x_{k+1}\) respectively. The above corresponding methods, HS is known as Hestenes and Steifel [7], FR is Fletcher and Reeves [9], PR is Polak and Ribiere [4], DX is Dixon[3],BA3 is AL-Bayati, A.Y. and AL-Assady[2], LS is Liu and Storey[11], DY is Dai and Yuan [10] , RMIL is Rivaie, Mustafa, Ismail and Leong[8] and lastly AMRI denotes Abdelrhaman Abashar, Mustafa Mamat, Mohd Rivaie and Ismail Mohd[1].

In this paper, we propose our new \(\beta_k^{\text{New1}}\) and compared its performance with standard formulas of (FR) method.

The remaining sections of the paper are arranged as follows. in section 2 , the new conjugate gradient formula and algorithm method presented, in section 3, we showed the sufficient descent condition and the global convergence proof of our new method. In section 4 numerical results, percentages, graphics and discussion. Lastly, In section 5 conclusion.

### 2. New proposed method and algorithm

In this algorithm, we modification the numerator in the proposed by Fletcher and Reeves method in 1964, where he proposed that:

\[
\beta_k^{PR} = \frac{g_k^{T}g_{k+1}}{g_k^Ty_k} \quad (2.1)
\]

Our proposal is

\[
g_{k+1} = g_{k+1} - \gamma \frac{g_{k+1}^Tv_k}{v_k^Ty_k} y_k \quad (2.2)
\]

where, \(\gamma \in (0,1]\)

The new method is as follows:

\[
\beta_k^{\text{New1}} = \frac{g_k^{T}g_{k+1}}{g_k^Ty_k} \quad (2.3)
\]

We programmed the new method and compared with the numerical results of the method Fletcher and Reeves and we noticed superiority of the new method that proposed on the method of Fletcher and Reeves.

#### 2.1 Algorithm of the New1 Method

**Step (1):** Given \(x_0 \in \mathbb{R}^n, \epsilon > 0, 0 < \gamma \leq 1\)

Set \(k = 0\), Compute \(f(x_0), g_0, d_k = -g_k\)

**Step (2):** If \(\|g_{k+1}\| < \epsilon\) stop.

**Step (3):** Compute \(a_k > 0\) satisfying the strong Wolfe condition

\[x_{k+1} = x_k + a_k d_k\]

**Step (4):** Compute \(d_{k+1} = -g_{k+1} + \beta_k^{\text{New1}} d_k\).

\[g_{k+1} = g_{k+1} - \gamma \frac{g_{k+1}^Tv_k}{v_k^Ty_k} y_k\]

\[\beta_k^{\text{New1}} = \frac{g_{k+1}^Tg_{k+1}}{g_k^Ty_k}\]

**Step (5):** If \(\|g_{k+1}\|^2 \geq \|g_{k+1}\|^2\) go to step (1) else continue.

**Step (6):** Set \(k = k + 1\), go to step (2)
3. The Global Convergent Analysis of the New Method

The convergence properties of \( \beta_k^{\text{New}} \) will be studied. For an algorithm to converge, it is necessary to show that the sufficient descent condition and the global convergence properties.

3.1 Sufficient Descent Condition

For the sufficient condition to hold, then
\[
\beta_k^{\text{New}} d_k \leq -C\|g_k\|^2 \quad \text{for } k \geq 0 \quad \text{and } C > 0 \quad (3.1)
\]

**Theorem 3.1**

Consider a CG method with search direction (1.3) and \( \beta_k^{\text{New}} \) defined as (2.3), assume that \( \alpha_k \) is satisfies strong Wolfe condition then, condition (3.1) will holds for all \( k \geq 0 \) in both cases exact line search and inexact line search.

**Proof**

By using induction mathematical

If \( k = 0 \), then we will have \( g_0^T d_0 \leq -C\|g_0\|^2 \). Hence condition (3.1) hold.

We need to show that for \( k \geq 1 \), condition (3.1), we also holds.

Now we prove the current search direction satisfies (3.1) at the iteration \((k + 1)\). From (1.3), multiply by \( g_{k+1} \) then
\[
g_{k+1}^T d_{k+1} = g_{k+1}^T (-g_{k+1} + \beta_k^{\text{New}} d_{k+1})
\]
\[
= -\|g_{k+1}\|^2 + \beta_k^{\text{New}} g_{k+1}^T d_k
\]

The proof is complete if the line search is exact, then \( g_{k+1}^T d_k = 0 \), and thus,
\[
g_{k+1}^T d_k = -\|g_{k+1}\|^2
\]

Which implies that \( d_{k+1} \) is a sufficient descent condition.

Now, if the line search is an inexact line search which requires \( g_{k+1}^T d_k \neq 0 \).

Therefore, we get
\[
g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} d_k^T g_{k+1} (1 - \gamma)
\]

By strong Wolfe condition, we have
\[
g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} (1 - \gamma)
\]

Since \( d_k^T y_k \) and \( \alpha_k \) are scalars, then
\[
g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2 + \|g_{k+1}\|^2 (1 - \gamma)
\]

**Lemma 3.1**

The norm of consecutive search direction are given by below expression
\[
\|d_{k+1}\| \leq \|\beta_k^{\text{New}}\|\|d_k\|, \text{ for all } k
\]

**Proof**

From (1.3), we have
\[
d_{k+1} + g_{k+1} = \beta_k^{\text{New}} d_k.
\]

By take norm both sides, we have
\[
\|d_{k+1} + g_{k+1}\| = \|\beta_k^{\text{New}} d_k\|, \text{ By using triangular inequality, we get}
\]
\[ \|d_{k+1}\| \leq \|d_{k+1} + g_{k+1}\| = \|\beta_k^{new} d_k\|. \] Hence, we get
\[ \|d_{k+1}\| \leq \|\beta_k^{new}\| \|d_k\|. \] for all \( k \)

**Lemma 3.2**

The norm of search direction and the norm of gradient are the same that is
\[ \|d_k\|^2 = \|g_k\|^2 \quad (3.5) \]

**Proof**

Multiply this equation \( d_k = -g_k \) by \( g_k^T \), we get
\[ g_k^T d_k = -\|g_k\|^2 \quad (3.6) \]

By square (3.6), we have
\[ (g_k^T d_k)^2 = -\|g_k\|^4 \Rightarrow \|g_k\|^2 \|d_k\|^2 = \|g_k\|^4 \]

Since \( g_k \neq 0 \), we get (3.5)

**Lemma 3.3**

The following relation holds for \( k \geq 0 \) in exact line search.
\[ \|g_{k+1} - d_k\|^2 = \|g_{k+1}\|^2 + \|d_k\|^2 \quad (3.7) \]

**Proof**

\[ \|g_{k+1} - d_k\|^2 = (g_{k+1} - d_k)^T (g_{k+1} - d_k) = \|g_{k+1}\|^2 - g_{k+1}^T d_k - d_k^T g_{k+1} + \|d_k\|^2 \]

Since \( g_{k+1}^T d_k = 0 \), we get (3.7)

### 3.2 Global Convergent

The following assumption are often needed to prove the convergence of the nonlinear conjugate gradient method, see [6]

**Assumption1:**

(i) \( f \) is bounded below on the level set \( R^n \) continuous and differentiable in a neighborhood \( N \) of the level set \( L = \{ x \in R^n : f(x) \leq f(x_0) \} \) at the initial point \( x_0 \).

(ii) The gradient \( g(x) \) is Lipschitz continuous in \( N \), so there exists a constant \( L > 0 \) such that \( \|g(x) - g(y)\| \leq L\|x - y\| \) for any \( x, y \in N \).

Based on this assumption, we have the below theorem that was proved by Zoutendijk [5]

**Theorem 3.1**

Suppose that assumption 1 holds. Consider any conjugate gradient of the from (1.3) where \( d_k \) is a descent search direction and we take \( \alpha_k \) in both cases exact line search and inexact line search. Then the following condition known as Zoutendijk condition holds

\[ \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \]

From the previous information, we can obtain the following convergence theorem of the conjugate gradient methods.

**Theorem 3.2**

Suppose that assumption 1 is true. Consider any conjugate gradient method of the form (1.3), where, \( \alpha_k \) is obtained by both cases exact line search and inexact line search and \( d_k \) is a descent search direction then either

\[ \lim_{k \to \infty} \|g_k\| = 0 \Rightarrow \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \]

**Proof**

To prove Theorem 3.2, we use contradiction. If Theorem 3.2 is not true, then there exists a constant \( \mu > 0 \), such that
\[ \|g_k\| \geq \mu, \forall i \geq 0. \quad (3.8) \]

Rewrite (1.3), we get
\[ d_{k+1} + g_{k+1} = \beta_k^{new} d_k \quad (3.9) \]

Squaring the above equation, we get
\( \|d_{k+1}\|^2 = (\beta_k^{\text{New}1})^2 \|d_k\|^2 - 2 g_k^T d_{k+1} - \|g_{k+1}\|^2 \) (3.10)

Dividing both sides of equation (3.10) by \((g_k^T d_{k+1})^2\), therefore we end up with

\[
\frac{\|d_{k+1}\|^2}{(g_k^T d_{k+1})^2} = \frac{(\beta_k^{\text{New}1})^2 \|d_k\|^2}{(g_k^T d_{k+1})^2} - \frac{2}{(g_k^T d_{k+1})^2} \frac{g_k^T d_{k+1}}{\|g_{k+1}\|^2} \frac{1}{1 - \frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2}} + \frac{1}{1 - \frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2}}
\]

Substitute \(\beta_k^{\text{New}1}\), we have

\[
\frac{\|d_{k+1}\|^2}{(g_k^T d_{k+1})^2} \leq \left(\frac{\|g_{k+1}\|^2}{\|g_k\|^2}\right)^2 \frac{\|d_k\|^2}{\|g_{k+1}\|^2} + \frac{1}{\|g_{k+1}\|^2}
\]

From Lemma 3.2, it gives us

\[
\frac{\|d_{k+1}\|^2}{(g_k^T d_{k+1})^2} \leq \frac{1}{\|g_k\|^2} + \frac{1}{\|g_{k+1}\|^2}
\]

Hence for \(k = 0\) the above inequality yield

\[
\frac{\|d_1\|^2}{(g_1^T d_1)^2} \leq \frac{1}{\|g_0\|^2} + \frac{1}{\|g_1\|^2}
\]

Hence for all \(k\), we conclude that

\[
\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{1}{\|g_0\|^2} + \frac{1}{\|g_k\|^2}
\]

Therefore

\[
\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \sum_{i=0}^{k} \frac{1}{\|g_i\|^2}
\]

So, by (3.8)

\[
\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{1}{\mu^2} \sum_{i=0}^{k} \frac{1}{\|g_i\|^2} \Rightarrow \frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{1}{\mu^2} \sum_{i=0}^{k} \frac{1}{\|g_i\|^2} \Rightarrow \frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{k}{\mu^2}
\]

We take summation both sides, we get

\[
\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \mu^2 \sum_{k=0}^{\infty} \frac{1}{k} = \infty
\]

Which contradicts Zoutendijk condition in Theorem 3.1. The proof is then complete.

### 4. Numerical Results and Discussions

This section is devoted to test the implement of the new method. We compare the new conjugate gradient algorithm (New1) and standard (F/R). The comparative tests involve well known nonlinear problems (classical test function) with different function \(4 \leq N \leq 5000\). all programs are written in FORTRAN 95 language and for all cases the stopping condition \(\|g_{k+1}\|_{\infty} \leq 1 \times 10^{-5}\) and restarting using Powell condition \(|g_k^T g_{k+1}| \geq 0.2 |g_{k+1}|^2\). The line search routine was a cubic interpolation which uses function and gradient values. The results given in tables (4.1) and (4.2) specifically quote the number of iteration NOI and the number of function NOF. Experimental results in tables (4.1) and (4.2) confirm that the new conjugate gradient algorithm (New1) is superior to standard algorithm (F/R) with respect to the number of iterations NOI and the number of functions NOF.

---

Page | 78
### Table (4.1)

<table>
<thead>
<tr>
<th>No. of test</th>
<th>Test function</th>
<th>N</th>
<th>Standard Formula (FR)</th>
<th>New Formula (New1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOI</td>
<td>NOF</td>
</tr>
<tr>
<td>1</td>
<td>Rosen</td>
<td>4</td>
<td>30</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>30</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>30</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>30</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5000</td>
<td>30</td>
<td>85</td>
</tr>
<tr>
<td>2</td>
<td>Cubic</td>
<td>4</td>
<td>13</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>14</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>15</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>15</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5000</td>
<td>15</td>
<td>44</td>
</tr>
<tr>
<td>3</td>
<td>Powell</td>
<td>4</td>
<td>40</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>42</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>43</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>43</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5000</td>
<td>43</td>
<td>125</td>
</tr>
<tr>
<td>4</td>
<td>Wolfe</td>
<td>4</td>
<td>11</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>14</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>46</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>52</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5000</td>
<td>141</td>
<td>293</td>
</tr>
<tr>
<td>5</td>
<td>Wood</td>
<td>4</td>
<td>27</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>27</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>27</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>27</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5000</td>
<td>27</td>
<td>61</td>
</tr>
<tr>
<td>6</td>
<td>Non-diagonal</td>
<td>4</td>
<td>23</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>27</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>27</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>27</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5000</td>
<td>27</td>
<td>73</td>
</tr>
</tbody>
</table>

### Table (4.2)

<table>
<thead>
<tr>
<th>No. of test</th>
<th>Test function</th>
<th>N</th>
<th>Standard Formula (FR)</th>
<th>New Formula (New1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOI</td>
<td>NOF</td>
</tr>
<tr>
<td>9</td>
<td>G-centeral</td>
<td>4</td>
<td>18</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>24</td>
<td>194</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>28</td>
<td>251</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>28</td>
<td>251</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5000</td>
<td>28</td>
<td>251</td>
</tr>
</tbody>
</table>
Comparing the rate of improvement between the new algorithm (New1) and the standard algorithm (F/R)

Table (4.3)

<table>
<thead>
<tr>
<th>Tools</th>
<th>Standard algorithm (F/R)</th>
<th>New algorithm (New1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOI</td>
<td>100%</td>
<td>91.6501%</td>
</tr>
<tr>
<td>NOF</td>
<td>100%</td>
<td>89.3648%</td>
</tr>
</tbody>
</table>

Table (4.3) shows the rate of improvement in the new algorithm (New1) with the standard algorithm (F/R). The numerical results of the new algorithm is better than the standard algorithm. As we notice that (NOI), (NOF) of the standard algorithm are about 100%. That means the new algorithm has improvement on standard algorithm prorate (8.3499%) in (NOI) and prorate (10.6352%) in (NOF). In general the new algorithm (New1) has been improved prorate (9.49256%) compared with standard algorithm (F/R).
Figure (4.1): shows the comparison between new algorithm (New1) and the standard algorithm (R/F) according to the total number of iterations (NOI) and the total number of functions (NOF).

Conclusion

In this paper, we proposed a new and simple $\beta_k^{\text{New1}}$ that has global convergence properties. Numerical results have shown that this new $\beta_k^{\text{New1}}$ performs better than FR. In the future we can improve the method to HS, PR, DX, DY, LS and other method.

References