Generalizations of the New Technique for Spectral Conjugate Gradient Methods

Basim A. Hassan
Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, IRAQ
*Correspondent author email: basimbas39@gmail.com

Abstract
In this article, we will present a generalization of the new technique for spectral conjugate gradient methods based on the descent condition that by using a simple method to prove the worldwide convergence of the new method without the Wolfe line searches. Depending on our numerical experiments we can confirm that our proposed methods are preferable and in general better to the classical conjugate gradient methods in terms of good organization.

Keywords: Conjugate gradient, Spectral conjugate gradient, Sufficient descent condition, worldwide convergence, Numerical results.

Introduction
The most well-known minimization technique for unconstrained problems is nonlinear conjugate gradient (CG) method. We take into account solving the unrestrained minimization function:

$$\text{minimize } f(x), \quad x \in \mathbb{R}^n$$ (1)

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is an uninterrupted specifiable function, limited from below. The general structure of nonlinear conjugate gradient method can be summarized as follows, preliminary from an initial point $x_1 \in \mathbb{R}^n$, the conjugate gradient method yields a cluster $x_k \in \mathbb{R}^n$ such that:

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

where $\alpha_k > 0$ is a step length, received from the line search, and the direction $d_k$ are given by:

$$d_1 = -g_1 , \quad d_{k+1} = -g_{k+1} + \beta_k d_k$$ (3)

In the previous relation, $\beta_k$ is the conjugate gradient parameter. Now, we denote $g_k = \nabla f(x_k)$, $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$.

More details can be found in [10]. Different conjugate gradient methods communicate to different selected for the formula $\beta_k$ such as:

$$\beta_k^{PR} = \frac{s_k^T g_{k+1}}{g_k^T g_k} , \quad \beta_k^{CD} = -\frac{s_k^T g_{k+1}}{g_k^T d_k}$$ (4)

$$\beta_k^{HS} = \frac{s_k^T y_k}{y_k^T d_k} , \quad \beta_k^{PR} = \frac{s_k^T y_k}{y_k^T g_k}$$ (5)


The step size $\alpha_k$ is computed by exact or inexact cable search. The Wolfe states is used in familiar:

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k$$ (6)
\[ g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \]  
(7)

where \( d_k \) is a descent direction and \( 0 < \delta \leq \sigma < 1 \).

Different from the classical conjugate gradient method, in a spectral conjugate gradient (SCG) method, the search direction \( d_{k+1} \) is defined as follows:

\[
d_{k+1} = -\vartheta_k g_{k+1} + \beta_k d_k,
\]
(8)

where \( \vartheta_k \) is called a spectral coefficient. It is easy to see that (8) reduces to (3) if \( \vartheta_k = 1 \).

Since there are two types of parameters can be suitably chosen to obtain a search direction in (3), it is possible that (3) combines the advantages of spectral method and conjugate gradient method. More details can be found in [6].

In this paper, we will mention our generalizations of the new technique for spectral conjugate gradient methods and its algorithm in part 2. Whereas in part 3, we demonstrate the adequate descent condition and the worldwide convergence realization of the new technique. Whereby in in part 4, we tackle the mathematical results and elaboration and at last we draw a conclusion in the part 5.

**Generalizations of the new technique for SCG - methods**

One motivation for our new formula for \( \vartheta_k \) is the descent property of the conjugate descent method.

We assume that the search direction is descent direction for all \( k \) steps, and we hope the conclusion also for \( k + 1 \) steps, so we have:

\[ g_{k+1}^T d_{k+1} < 0 \]  
(9)

From the (10) and multiplying by \( g_{k+1}^T \) and from the (18) we have:

\[-\vartheta_k g_{k+1}^T g_{k+1} + \beta_k g_{k+1}^T d_k < 0 \]  
(10)

We observe the formula (10), let the form of \( \vartheta_k \) to be \( \vartheta_k = 1 + \eta_k \), and we assume \( \beta_k > 0 \), we have:

\[
\frac{\vartheta_k \| g_{k+1} \|^2}{\beta_k} > g_{k+1}^T d_k
\]
(11)

\[
\frac{(1 + \eta_k)\| g_{k+1} \|^2}{\beta_k} > g_{k+1}^T d_k
\]
(12)

From (11) we get:

\[
\| g_{k+1} \|^2 + \frac{\eta_k \| g_{k+1} \|^2}{\beta_k} > g_{k+1}^T d_k
\]
(13)

From (12) we let:

\[
\frac{\| g_{k+1} \|^2}{\beta_k} = d_k^T Gv_k
\]
(14)

and

\[
\frac{\eta_k \| g_{k+1} \|^2}{\beta_k} = g_{k+1}^T d_k
\]
(15)

From (15) and (14), we know that:

\[
\beta_k = \frac{\| g_{k+1} \|^2}{d_k^T Gv_k}
\]
(16)

and

\[
\vartheta_k = 1 + \frac{g_{k+1}^T d_k}{d_k^T Gv_k}
\]
(17)

In particular, Hideaki and Yasushi (HY) [5] made a modification on the CG method and developed a modified method. Their method expression of the denominator \( d_k^T Gv_k \), we know that:

\[
d_k^T Gv_k = 2(f_k - f_{k+1})/\alpha_k
\]
(18)

Accordingly, we give the parameters \( \vartheta_k^GB \) in the above algorithm:

\[
\vartheta_k = 1 + \frac{g_{k+1}^T d_k}{2(f_k - f_{k+1})/\alpha_k}
\]
(19)

Different choices for \( d_k^T Gv_k \) lead to different spectral conjugate gradient methods.
Then, we can propose the following generalizations of the new technique for spectral nonlinear conjugate gradient methods (GB - methods):

**New Algorithm:**

**Step 1.** Give \( x_1 \in R^n \), \( \varepsilon \geq 0 \). Set \( d_1 = -g_1 \) and set the initial \( \alpha_i = 1 / \| g_i \| \).

**Step 2.** If \( \| g_{k+1} \| \leq 10^{-\varepsilon} \), then stop, or else be present at Step 3.

**Step 3.** Find \( \alpha_k \) satisfying Wolfe conditions

\[
(6-7) \text{ and new iterative } x_{k+1} = x_k + \alpha_k d_k.
\]

**Step 4.** Compute \( \beta_k \) by (15) with \( g^G_k \) by (16) respectively.

**Step 5.** Compute direction \( d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k \).

Set \( k = k+1 \) and continue with step 2.

**Convergent analysis**

Within this part, the convergent features of GB will be considered. For a method to converge, it should meet the descent state and the worldwide convergence properties.

**Sufficient descent condition**

For the sufficient state to hold,

\[
g^T_{k+1} d_{k+1} \leq -c \| g_{k+1} \|^2, \quad c > 0 \quad (20)
\]

**Theorem 1.**

Consider a CG method with the search direction (3) with \( \beta_k \) and \( g^G_k \) given as (16) and (17), then state (20) holds for all \( k + 1 \).

**Proof.**

If \( k = 1 \), then \( g^T_1 d_1 \leq -c \| g_1 \|^2 \). Thus, state (20) holds true. We too require to prove that for \( k > 1 \), state (20), will also hold true. From (8), multiply by \( g^T_{k+1} \) then:

\[
g^T_{k+1} d_{k+1} = -\theta_k g^T_{k+1} g_{k+1} + \beta_k g^T_{k+1} d_k
\]

\[
= -\left( 1 + \frac{g^T_{k+1} d_k}{d^T_k G_k} \right) g^T_{k+1} g_{k+1} + \frac{\| g_{k+1} \|^2}{d^T_k G_k} g^T_{k+1} d_k
\]

\[
= -\| g_{k+1} \|^2
\]

Furthermore, from above analysis, we also get \( \beta_k > 0 \). The proof is completed.

**Lemma 1.**

Consider any method (2), (8), where (15), (16) and the step-size \( \alpha_k \) be determined by the Wolfe line search, then:

\[
\sum_{k=1}^{\infty} \frac{(g^T_k d_k)^2}{\| d_k \|^2} < \infty \quad (22)
\]

The above lemma (1) regularly called the Wolfe condition. It was originally known by Zoutendijk. [11]

**Global convergence properties**

In the direction of study, the global convergence properties, first if we chose any value \( d^T_k G_k \) with sufficient descent condition that \( \beta_k \) are always satisfies the relations:

\[
0 < \beta_k \leq \frac{-g^T_{k+1} d_{k+1}}{-g^T_k d_k} \quad (23)
\]

This formula is very important in our convergence analysis.

The following assumption that may be needed in many theorem proofs:

**Assumption (A)**

\( A_1 : f : R^n \rightarrow R^1 \) is bounded below.

\( A_2 : \text{The level set } S = \{ x \in R^n | f(x) \leq f(x_0) \} \) is bounded, i.e., there exists a positive constant \( \zeta > 0 \) such that:

\[
\| x \| \leq \zeta, \quad \forall x \in S \quad (24)
\]

\( A_3 : \nabla f \text{ satisfied the Lipschitz state namely, } \| g(x_{k+1}) - g(x_k) \| \leq L \| x_{k+1} - x_k \|, \quad \forall x_{k+1}, x_k \in U \quad (25)\)

where \( L > 0 \) Lipschitz constant.

Under these assumptions of \( f(x) \), there exists a constant \( \Gamma \geq 0 \) such that:

\[
\| g(x) \| \leq \Gamma \quad (26)
\]

More details can be found in [9].
Theorem 2.
Suppose that Assumption (A) holds. Consider any method (2), (8), where (15), (16) and the step-length $\alpha_k$ be verify by the Wolfe line search, then:
\[
\liminf_{k \to \infty} \|g_k\| = 0 
\] (27)

Proof:
Suppose by disagreement that near exists a positive constant $\varepsilon > 0$ such that:
\[
\|g_{k+1}\| > \varepsilon, 
\] (28)

From (8), we have:
\[
\|d_{k+1}\|^2 = (\beta_k)^2 \|d_k\|^2 - 2g_k^T g_{k+1} d_{k+1} - (\gamma_k^2) \|g_{k+1}\|^2 
\] (29)

From the above equation and (23), we have:
\[
\|d_{k+1}\|^2 \leq \left( \frac{\alpha_k^2}{\|d_k\|^2} \right) \|d_k\|^2 - 2g_k^T g_{k+1} d_{k+1} - (\gamma_k^2) \|g_{k+1}\|^2 
\] (30)

Dividing both inequality by $(g_k^T d_k + 1)^2$, we have:
\[
\frac{\|d_k\|^2}{(g_k^T d_k + 1)^2} \leq \frac{\|d_k\|^2}{(g_k^T d_k + 1)^2} - (\gamma_k^2) \|g_{k+1}\|^2 - \frac{1}{g_k^T d_k} \frac{1}{\|g_{k+1}\|^2} 
\] (31)

Using (31) recursively and noting that $\|d_k\|^2 = -g_k^T d_k = \|g_k\|^2$, we obtain:
\[
\|d_k\|^2 \leq \sum_{i=1}^{k} \frac{1}{\|g_i\|^2} 
\] (32)

From (31) and (28) we obtain:
\[
\frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \varepsilon_k^2 
\] (33)

which indicates
\[
\frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{\varepsilon_k^2}{k} 
\] (34)

This does not coincide with his (22). So, the result (27) holds.

Numerical results
We implement 15 classical unrestricted optimization questions from [1] successively to check algorithm. These experiments are done by employing Fortran. In the First place we set $\delta_1 = 0.001$ and $\delta_2 = 0.9$ and if $\|g_{k+1}\| \leq 10^{-6}$ then stop. The mathematical results of the tests are placed in Table 1. The initial pillar “Problem” stands for the label of the examined problem. Dim refers to the dimension of the test problems. NI, NF refer to the figure of iterations and function valuations, successively. Out of the mathematical results, it is denoted that the suggested proposed spectral conjugate gradient method is showing excellence.

In Table 1, it can be seen from the comparison given above that the new algorithm in this paper is more efficient than FR method and HY method for solving unconstrained optimization.

Conclusions
In this paper, we contain derived a generalization of the new technique for spectral conjugate gradient method based on our descent condition. We have preliminary numerical outcome to show its efficiency. As demonstrated in Section 4, the information mathematical results illustrate that the modified GB performs better than the FR in [2] and HY in [5].

Table 1: Comparison of different CG methods with different test problems and different dimensions.

<table>
<thead>
<tr>
<th>P. No.</th>
<th>n</th>
<th>NI</th>
<th>NF</th>
<th>NI</th>
<th>NF</th>
<th>NI</th>
<th>NF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>19</td>
<td>35</td>
<td>19</td>
<td>36</td>
<td>18</td>
<td>34</td>
</tr>
<tr>
<td>1000</td>
<td>38</td>
<td>65</td>
<td>34</td>
<td>62</td>
<td>36</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>47</td>
<td>93</td>
<td>42</td>
<td>84</td>
<td>31</td>
<td>64</td>
</tr>
<tr>
<td>1000</td>
<td>78</td>
<td>131</td>
<td>35</td>
<td>73</td>
<td>34</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>32</td>
<td>52</td>
<td>14</td>
<td>29</td>
<td>13</td>
<td>26</td>
</tr>
<tr>
<td>1000</td>
<td>22</td>
<td>42</td>
<td>16</td>
<td>32</td>
<td>13</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>10</td>
<td>27</td>
<td>9</td>
<td>25</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>1000</td>
<td>24</td>
<td>191</td>
<td>11</td>
<td>33</td>
<td>11</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>32</td>
<td>64</td>
<td>13</td>
<td>25</td>
<td>11</td>
<td>23</td>
</tr>
<tr>
<td>1000</td>
<td>77</td>
<td>129</td>
<td>19</td>
<td>34</td>
<td>13</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>15</td>
<td>25</td>
<td>14</td>
<td>21</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>1000</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>71</td>
<td>110</td>
<td>26</td>
<td>50</td>
<td>38</td>
<td>71</td>
</tr>
<tr>
<td>1000</td>
<td>47</td>
<td>84</td>
<td>25</td>
<td>50</td>
<td>29</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>101</td>
<td>217</td>
<td>85</td>
<td>206</td>
<td>79</td>
<td>195</td>
</tr>
<tr>
<td>1000</td>
<td>101</td>
<td>214</td>
<td>85</td>
<td>203</td>
<td>81</td>
<td>196</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>23</td>
<td>45</td>
<td>18</td>
<td>33</td>
<td>19</td>
<td>34</td>
</tr>
<tr>
<td>1000</td>
<td>27</td>
<td>55</td>
<td>19</td>
<td>44</td>
<td>22</td>
<td>47</td>
<td></td>
</tr>
</tbody>
</table>
References


[9] Snežana S. D., (2017), " New Hybrid Conjugate Gradient Method as a Convex Combination of LS and CD methods, Published by Faculty of Sciences and Mathematics, University of Niš, Serbia, pp. 1813-1825.
