New scaled algorithm for non-linear conjugate gradients in unconstrained optimization

Ghada M. Al-Naem, Ahmed H. Sheekoo
Department of Energy Mathematics, Faculty of Computer Science and Mathematics, Mosul University, Mosul, Iraq

ABSTRACT
A new scaled conjugate gradient (SCG) method is proposed throughout this paper, the SCG technique may be a special important generalization conjugate gradient (CG) method, and it is an efficient numerical method for solving nonlinear large scale unconstrained optimization. As a result, we proposed the new SCG method with a strong Wolfe condition (SWC) line search is proposed. The proposed technique's descent property, as well as its global convergence property, are satisfied without the use of any line searches under some suitable assumptions. The proposed technique's efficiency and feasibility are backed up by numerical experiments comparing them to traditional CG techniques.

Keywords: CG method, Large-scale nonlinear, SCG method, Sufficient descent property, Unconstrained optimization

This is an open access article under the CC BY-SA license.

1. INTRODUCTION
CG method is universal method for solving nonlinear large-scale unconstrained optimization problems, because it has simple iterations, low memory requirements and very fast convergence properties [1]. Therefore, in this work, we considered this general unconstrained optimization problem: indexing and abstracting services depend on the accuracy of the title, extracting from it keywords useful in cross-referencing and computer searching. An improperly titled paper may never reach the audience for which it was intended, so be specific.

\[ \text{Min}\{f(x): x \in \mathbb{R}^n \} \] (1)

Where \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) is smooth and its gradient vector defined \( g_n = \nabla f(x_n) \), and the initial point \( x_0 \in \mathbb{R}^n \) is usually solved iteratively according to the recursive formula,

\[ x_{n+1} = x_n + \tau_n d_n, n \geq 0 \] (2)

where \( x_n \) is current iteration, \( \tau_n > 0 \) is the step-size calculated by the SWC,

\[ \frac{f(x_n + \tau_n d_n) - f(x_n) + \delta \tau_n \Delta g_n}{g(x_n + \tau_n d_n)^T d_n} \leq -\sigma g_n^T d_n \] (3)

Journal homepage: http://ijeecs.iaescore.com
where $0 < \delta < \sigma < 1$ and $d_n$ is a search direction. The classical search direction $d_{n+1}$ are frequently defined by,

$$d_{n+1} = \begin{cases} -g_{n+1}, & \text{if } n = 0 \\ -g_{n+1} + \beta_n d_n, & \text{if } n \geq 1 \end{cases}$$ (4)

generally, the parameter $\beta_n$ is selected so that if $f(x)$ is a strictly convex quadratic function and if $\tau_n$ is calculated by the exact line search, then (2) and (4) can be simplified to the linear conjugate gradient technique [2]. Several formulas, such as hestenes and stiefel (HS), fletcher and reeves (FR), conjugate descent (CD), Polak-Ribiere (PRP), Liu and Storey (LS) and Dai-Yuan method (DY), have been proposed [3]-[9]. As demonstrated by the formula,

$$\beta_{HS} = \frac{g_n^T(g_n - g_{n-1})}{(g_n - g_{n-1})^T d_{n-1}}; \quad \beta_{PRP} = \frac{\gamma_n g_n}{g_n^T g_{n-1}}; \quad \beta_{DY} = \frac{\gamma_n}{\gamma_n - 1}$$

$$\beta_{CD} = -\frac{g_n^T g_{n-1} d_{n-1}}{\gamma_n^2 - 1} \quad \beta_{LS} = \frac{\gamma_n}{\gamma_n - 1}$$

the primary distinction between SCG and CG is the calculation of the search direction. SCG's typical search direction is as follows,

$$d_{n+1} = \begin{cases} -g_{n+1}, & \text{if } n = 0 \\ -\vartheta_n g_{n+1} + \beta_n d_n, & \text{if } n \geq 1 \end{cases}$$ (5)

where $\vartheta_n$ denotes a spectral parameter. Barzilai and Borwien [10] proposed the SCG method and developed their unconstrained optimization. Instead of global convergence, the idea is to use only teasing trends. Birgin and Martinez [11] proposed an accelerated CG technology that uses the Newton method to improve the CG method's performance. Following on from this thought, Farvaneh and Keyvan [13] proposed a new SCG [14]-[20] contain additional references in this field.

### 2. NEW ALGORITHM AND THE DESCENT PROPERTY

Obviously, for SCG, the method for selecting the spectral parameter $\vartheta_n$ and conjugate parameter $\beta_n$ is critical. In this section, we explain how our proposed SCG is dependent on the parameter $\beta_n$ proposed by Wei et al. [21], which is defined as (6).

$$\beta_n^{WY} = \frac{||g_n||^2 - ||g_n|| ||\vartheta_n d_{n+1}||}{||g_{n+1}||^2}$$ (6)

The new spectral parameter $\vartheta_n$ is proposed by (7),

$$\vartheta_n = 1 + \frac{g_{n+1}^T (g_{n+1} g_n) g_{n+1}^T d_n}{||g_n||^2 ||g_n||^2} \quad \text{if } ||g_n|| \leq \varepsilon$$ (7)

note that, if an exact line search is used then $\vartheta_n = 1$, so (5) reduced to (4).

Algorithm SCG

Step1: Select a starting point $x_0 \in R$, given constant $0 < \delta < \sigma < 1$, stopping criteria $\varepsilon = 10^{-6} > 0$; Set $d_0 = -g_0$.

Step2: Compute $||g_n||$, if $||g_n|| \leq \varepsilon$, stop. Otherwise, continues.

Step3: Calculate $\beta_n^{WY}$, $\vartheta_n$ by (6) and (7) respectively and compute step length $\tau_n$ by (3).

Step4: Update the new point by (2). Compute $g_{n+1} = g(x_{n+1});$ if $||g_{n+1}|| \leq \varepsilon$, stop; Otherwise, continues.

Step5: Compute search direction $d_{n+1}$ by (5).

Step6: If the Powell restart criteria

$$||g_{n+1}^T g_n|| \geq 0.2 ||g_{n+1}||^2$$ (8)
is satisfied, set \( d_{n+1} = -g_{n+1} \) and go back to Step 3; otherwise continues.  
Step 7: Put \( n = n + 1 \) and go to step 3.  

We will discuss the sufficient descent property of the Algorithm SCG above without depending to any line search.  

### 2.1. Theorem

It can be concluded that the SCG method with the line search direction (5), \( \beta_n^{WY} \), \( \vartheta_n \) defined in (6) and (7) respectively, and then,

\[
g_{n+1}^T d_{n+1} \leq -\xi \|g_{n+1}\|^2, \quad \xi \geq 0
\]

holds for \( \forall n \geq 0 \).  

Proof: To stimulate this confirmation, we use induction, if \( n = 0 \), then \( g_0^T d_0 = -\|g_0\|^2 \), as a result; condition (9) is established. Now, condition (9) is also true in order to notify that every \( n \geq 0 \) is true. Multiply both sides of (5) by \( g_{n+1}^T \) to obtain,

\[
g_{n+1}^T d_{n+1} = -\left( 1 + \frac{g_{n+1}^T d_n - \langle g_{n+1}^n, g_{n+1} \rangle}{\|g_n\|^2} \right) \|g_{n+1}\|^2 + \frac{\|g_{n+1}\|^2 \|g_{n+1} \|^2}{\|g_n\|^2} g_{n+1}^T d_n
\]

\[
= -\|g_{n+1}\|^2 - \frac{\|g_{n+1}\|^2 \|g_{n+1} \|^2}{\|g_n\|^2} g_{n+1}^T d_n + \frac{\|g_{n+1}\|^2 \|g_{n+1} \|^2}{\|g_n\|^2} g_{n+1}^T d_n
\]

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

therefore, the Algorithm SCG can satisfy the sufficient descent conditions without using any line searches.

### 3. THE GLOBAL CONVERGENCE ANALYSIS

The general situation of the objective function required for the overall global convergence of general CG in psychological analysis is as follows.

#### 3.1. Assumption

- The function \( f(x) \) is constrained from below to the level set \( \Phi = \{ x : x \in \mathbb{R}^n / f(x) \leq f(x_0) \} \), where the point of departure is \( x_0 \), i.e., there is a constant \( \alpha > 0 \), which means \( \| x_n \| \leq \alpha \forall x \in \Phi \).
- In certain neighborhood \( N \) of the level set \( \Phi \), the function \( f(x) \) is continuously differentiable and its gradient \( g(x) \) is Lipschitz continuous, i.e. \( \exists \) a constant \( l_\xi > 0 \) s. t.

\[
d_{n+1} = \\begin{cases} 
- g_{n+1}, & \text{if } n = 0 \\
- g_{n+1} + \beta_n d_n, & \text{if } n \geq 1
\end{cases}
\]

Assumption (I) clearly implies the existence of a constant \( \omega > 0 \), s. t.,

\[
0 < \| g_{n+1} \| \leq \omega, \forall x \in \Phi \quad [22]
\]

the following Lemma, known as the Zountendijk condition Zountendijk [23], proposed it and is frequently used to demonstrate global convergence of CG techniques.

#### 3.2. Lemma

Suppose Assumption (I) holds. Suppose a general iterative method (2) and the direction (4) is descent direction. So, we have got,

\[
\sum_{n=0}^{\infty} \frac{(\beta_n d_n)^2}{\|d_n\|^2} < \infty
\]

according to Assumptions (3.1), Theorem (2.1) and Lemma (3.1), the following results can be proved.
3.3. Theorem
Suppose that Assumption (I) holds. Any CG method of the form (2) and (5) with $d_n$ is a descending search direction and $\tau_n$ satisfies SWC. Then,

$$\liminf_{n \to \infty} \|g_n\| = 0$$  \hspace{1cm} (14)

or,

$$\sum_{n=1}^{\infty} \frac{\|g_n\|^4}{\|d_n\|^2} < +\infty$$  \hspace{1cm} (15)

Proof: Assume, for the sake of argument that the conclusion is not true. Then there exists a positive constant $\bar{\omega} > 0$ s.t. $\|g_{n+1}\| \geq \bar{\omega}, \forall n$. We can deduce from (5) that $d_{n+1} + \varrho_n g_{n+1} = \beta_n^{\text{WLY}} d_n$. When we square both sides of this equation, we get,

$$(d_{n+1} + \varrho_n g_{n+1})(d_{n+1} + \varrho_n g_{n+1}) = (\beta_n^{\text{WLY}})^2 \|d_n\|^2$$

$$\|d_{n+1}\|^2 = -(\varrho_n)^2 \|g_{n+1}\|^2 - 2\varrho_n g_n^T d_{n+1} + (\beta_n^{\text{WLY}})^2 \|d_n\|^2$$

dividing both sides of the above equation by $\|g_{n+1}\|^4$, and use (10) we get,

$$\frac{\|d_{n+1}\|^2}{(\varrho_n)^2 \|g_{n+1}\|^2} = \frac{\|g_{n+1}\|^2}{\|g_{n+1}\|^2} - \frac{2\varrho_n}{\|g_{n+1}\|^2} + (\beta_n^{\text{WLY}})^2 \frac{\|d_n\|^2}{\|g_{n+1}\|^2}$$

$$\frac{\|d_{n+1}\|^2}{(\varrho_n)^2 \|g_{n+1}\|^2} = \frac{\|g_{n+1}\|^2}{\|g_{n+1}\|^2} - \frac{2\varrho_n}{\|g_{n+1}\|^2} + (\beta_n^{\text{WLY}})^2 \frac{\|d_n\|^2}{\|g_{n+1}\|^2}$$

$$\frac{\|d_{n+1}\|^2}{\|g_{n+1}\|^2} = \frac{1}{\|g_{n+1}\|^2} + (\beta_n^{\text{WLY}})^2 \frac{\|d_n\|^2}{\|g_{n+1}\|^2} - \left(\frac{1}{\|g_{n+1}\|^2} + \frac{(\varrho_{n+1})^2}{\|g_{n+1}\|^2}\right)$$

$$\leq \frac{1}{\|g_{n+1}\|^2} + (\beta_n^{\text{WLY}})^2 \frac{\|d_n\|^2}{\|g_{n+1}\|^2}$$

in [16] they proved $0 \leq \beta_n^{\text{WLY}} \leq \frac{2\|g_{n+1}\|^2}{\|g_n\|^2}, \forall n \geq 0$

$$\frac{\|d_{n+1}\|^2}{\|g_{n+1}\|^2} \leq \frac{1}{\|g_{n+1}\|^2} + \left(\frac{2\|g_{n+1}\|^2}{\|g_n\|^2}\right)^2 \frac{\|d_n\|^2}{\|g_{n+1}\|^2}$$

$$= \frac{1}{\|g_{n+1}\|^2} + \frac{4\|d_n\|^2}{\|g_n\|^2}$$

$$\|d_{n+1}\|^2 \leq 4\|d_n\|^2 + \frac{1}{\|g_{n+1}\|^2}$$

In terms of $\frac{\|d_t\|^2}{(\varrho_t d_t)^2} = \frac{1}{\|g_t\|^2}$, together with the above relations and $\|g_n\|^2 \geq \omega$, we have,

$$\frac{\|d_{n+1}\|^2}{\|g_{n+1}\|^2} \leq 4\|d_n\|^2 + \frac{1}{\|g_{n+1}\|^2} + \frac{1}{\|g_n\|^2}$$

$$\leq \cdots \leq \sum_{i=1}^{n} \frac{1}{\|g_i\|^2} \leq \frac{n}{\omega^2}$$
that is, $\frac{\|g_{n+1}\|^2}{\|d_{n+1}\|^2} \geq \frac{\sigma^2}{n}$. Hence $\sum_{n \geq 1} \frac{\|d_{n+1}\|^2}{\|g_{n+1}\|^2} \geq +\infty$, this is contradicts lemma (3.1). Therefore, the proof is complete.

4. **THE NUMERICAL RESULTS**

In this section, we will present the outcomes of various test functions. To evaluate the new method, some test functions were chosen. These functions are taken into account by CUTE test function [24], [25]. Using SWC line search, the new SCG method, the classic [21] (WYL) method, the FR method, and the LS method are compared in terms of the number of iterations (NI) and the number of function evaluations (NF). All symbols are written in FORTRAN 77 double precision and collected as Visual FORTRAN (F6.6). The new SCG method is implemented using the SWC line search (3), and with $\delta = 0.001, \sigma = 0.9$, we tested 15 well-known test functions, the dimensions of which are given below (1000, 5000, 10000, 50000, and 100000). This algorithm’s stopping criterion is $\|g_{n+1}\| \leq 10^{-6}$ and we enter 600 if the (NI) equal to or more than 600. The results obtained by the newly proposed method outperform those obtained by the other methods mentioned in the Table 1.

<table>
<thead>
<tr>
<th>No</th>
<th>Test Function</th>
<th>Dimension</th>
<th>SCG method ni</th>
<th>SCG method nf</th>
<th>WYL method ni</th>
<th>WYL method nf</th>
<th>FR method ni</th>
<th>FR method nf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ROSEN</td>
<td>1000</td>
<td>26</td>
<td>67</td>
<td>30</td>
<td>8</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>26</td>
<td>67</td>
<td>30</td>
<td>7</td>
<td>20</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>27</td>
<td>70</td>
<td>32</td>
<td>83</td>
<td>21</td>
<td>83</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>WOLFE</td>
<td>1000</td>
<td>110</td>
<td>225</td>
<td>116</td>
<td>230</td>
<td>111</td>
<td>229</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>120</td>
<td>244</td>
<td>123</td>
<td>248</td>
<td>135</td>
<td>280</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EX-BLOCK DIAGONAL</td>
<td>1000</td>
<td>121</td>
<td>247</td>
<td>130</td>
<td>255</td>
<td>134</td>
<td>276</td>
</tr>
<tr>
<td></td>
<td>BDI</td>
<td>100000</td>
<td>1000</td>
<td>26</td>
<td>53</td>
<td>25</td>
<td>53</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>SHALLOW</td>
<td>1000</td>
<td>10</td>
<td>25</td>
<td>10</td>
<td>25</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>11</td>
<td>27</td>
<td>11</td>
<td>27</td>
<td>11</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>WOOD</td>
<td>1000</td>
<td>10</td>
<td>28</td>
<td>64</td>
<td>29</td>
<td>66</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>30</td>
<td>68</td>
<td>29</td>
<td>66</td>
<td>30</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>BEAL</td>
<td>1000</td>
<td>11</td>
<td>27</td>
<td>11</td>
<td>27</td>
<td>11</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>11</td>
<td>27</td>
<td>11</td>
<td>27</td>
<td>11</td>
<td>27</td>
<td>600</td>
</tr>
<tr>
<td>7</td>
<td>POWELL</td>
<td>1000</td>
<td>38</td>
<td>122</td>
<td>56</td>
<td>162</td>
<td>36</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>41</td>
<td>138</td>
<td>56</td>
<td>162</td>
<td>39</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>CUBIC</td>
<td>1000</td>
<td>16</td>
<td>45</td>
<td>16</td>
<td>45</td>
<td>16</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>16</td>
<td>45</td>
<td>16</td>
<td>45</td>
<td>16</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>HIMMELBAU</td>
<td>1000</td>
<td>24</td>
<td>251</td>
<td>26</td>
<td>266</td>
<td>26</td>
<td>276</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>8</td>
<td>391</td>
<td>10</td>
<td>405</td>
<td>10</td>
<td>401</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>DQDRTIC</td>
<td>1000</td>
<td>5</td>
<td>11</td>
<td>6</td>
<td>13</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>5</td>
<td>11</td>
<td>6</td>
<td>13</td>
<td>5</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>DIXMAANB</td>
<td>1000</td>
<td>5</td>
<td>13</td>
<td>6</td>
<td>15</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>6</td>
<td>16</td>
<td>7</td>
<td>16</td>
<td>7</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>STRAIT</td>
<td>1000</td>
<td>6</td>
<td>14</td>
<td>8</td>
<td>21</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>6</td>
<td>14</td>
<td>8</td>
<td>21</td>
<td>6</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>BEAL U63</td>
<td>1000</td>
<td>10</td>
<td>27</td>
<td>13</td>
<td>32</td>
<td>12</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>10</td>
<td>27</td>
<td>13</td>
<td>32</td>
<td>12</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>HILICAL</td>
<td>1000</td>
<td>30</td>
<td>78</td>
<td>29</td>
<td>75</td>
<td>29</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>33</td>
<td>82</td>
<td>30</td>
<td>78</td>
<td>31</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>DENSCHNB</td>
<td>1000</td>
<td>9</td>
<td>21</td>
<td>9</td>
<td>21</td>
<td>9</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>9</td>
<td>21</td>
<td>9</td>
<td>21</td>
<td>9</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. The comparison between the proposed method and the other classical methods

New scaled algorithm for non-linear conjugate gradients in unconstrained... (Ghada M. Al-Naemi)
Table 2 compares the performance percentages of the FR, WYL, and proposed SCG technologies. When compared to the FR-method, the WYL technique saves (NI 23.38%), (NF 6.26%) and the SCG technique saves (NI 53.46%), (NF 22.47%). Under the strong Wolfe line search, the proposed method outperformed the existing methods in terms of number of iterations and number of function evaluations.

<table>
<thead>
<tr>
<th>Measures</th>
<th>FR method</th>
<th>WYL method</th>
<th>SCG method</th>
</tr>
</thead>
<tbody>
<tr>
<td>NI</td>
<td>100%</td>
<td>76.62%</td>
<td>46.54%</td>
</tr>
<tr>
<td>NF</td>
<td>100%</td>
<td>93.74%</td>
<td>77.53%</td>
</tr>
</tbody>
</table>

Table 2. The percentage performance of the proposed methods

5. CONCLUSION
In this paper, a new scaled conjugate gradient algorithm for unconstrained optimization problems is proposed. This method, independent of the line search, satisfies the sufficient descent condition. The proposed method has the advantage of being applicable to large-scale problems. The strong Wolfe line search is used to perform numerical computations on some standard benchmark problems. Preliminary findings indicate that the proposed method is both efficient and promising. As a result, it can be used as a different approach for large-scale unconstrained optimization problems. Furthermore, future research can focus on demonstrating the convergence of this method under different line search methods.

ACKNOWLEDGMENTS
The authors are grateful to the University of Mosul's College of Computer Sciences and Mathematics for their encouragement and support.

REFERENCES
New scaled algorithm for non-linear conjugate gradients in unconstrained... (Ghada M. Al-Naemi)