Extended Spectral Nonlinear Conjugate Gradient methods for solving unconstrained problems

Dr. Basim A. Hassan
Asst. Professor, College of Computers Sciences and Math., University of Mosul, Iraq

ABSTRACT

In this paper, we present extension forms of Dai, Yuan (DY), Fletcher, Revers (FR) and Conjugate Descent (CD) CG algorithms. The extended method have the sufficient descent and globally convergence properties under certain conditions. These new algorithms are tested on some standard test functions and compared with the original FR algorithm showing considerable improvements over all these comparisons.

Keywords: Conjugate gradient method, Spectral Conjugate gradient method, sufficient descent property, global convergent methods.

INTRODUCTION

The nonlinear conjugate gradient (CG) method is designed to solve the following unconstrained optimization problem

$$\min \{ f(x) \mid x \in \mathbb{R}^n \}$$ ...........(1)

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuously differentiable nonlinear function whose gradient is denoted by $g$. Due to its simplicity and its very low memory requirement, the CG method has played a special role for solving large scale nonlinear optimization problems. The iterative formula of the CG method is given by

$$x_{k+1} = x_k + \alpha_k d_k$$ ...........(2)

where $\alpha_k > 0$ is a step length which is computed by carrying out a line search and satisfies the standard Wolfe (SW ) conditions:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta_1 \alpha_k d_k^T g_k$$ ...........(3)

$$g(x_k + \alpha_k d_k)^T d_k \geq \delta_2 d_k^T g_k$$ ...........(4)

with $0 < \delta_1 < \delta_2 < 1$, and $d_{k+1}$ is the search direction defined by

$$d_{k+1} = \begin{cases} -g_k & k = 1, \\ -g_{k+1} + \beta_k d_k & k > 1, \end{cases}$$ ...........(5)

where $d_k$ is a descent direction. Different conjugate gradient algorithms correspond to different choices for the scalar parameter $\beta_k$ see [7]. The well-known formula of $\beta_k$ are given by

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k},$$ ...........(6)

$$\beta_k^{CD} = \frac{g_{k+1}^T g_{k+1}}{-g_k^T d_k},$$ ...........(7)

$$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{y_k^T d_k},$$ ...........(8a)

which are called Fletcher and Reeves (FR) [4], Conjugate Descent (CD) [3] and Dai and Yuan (DY) [2], respectively. In fact, utilizing (5), $\beta_k^{DY}$ can be rewritten as:
Extended spectral conjugate gradient method and algorithm

In this paper we suggest a new type of spectral conjugate gradient methods for solution of the \( \min f(x) \). In [5] we consider a condition that a descent search direction is generated, and we extend the DY method. We make such a direction inductively. Suppose that the current search direction \( d_k \) is a descent direction, namely, \( g_k^T d_k < 0 \) at the \( k^{th} \) iteration. Now we need to find a \( \beta_k \) that produces a descent search direction \( d_{k+1} \). This requires that

\[
g_k^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_k g_k^T d_k. \tag{11}
\]

Letting \( \gamma_{k+1} \) be a positive parameter, we define

\[
\beta_k = \frac{\|g_{k+1}\|^2}{\tau_{k+1}}. \tag{12}
\]

Equation (11) is equivalent to

\[
\tau_{k+1} > g_k^T d_k. \tag{13}
\]

Taking the positively of \( \gamma_{k+1} \) in to consideration, we have

\[
\tau_{k+1} > \max \left\{ g_k^T d_k, 0 \right\}. \tag{14}
\]

Therefore if condition (14) is satisfied for all \( k \), the conjugate gradient method with (12) produces a descent search direction at every iteration. From (12) we can get various kinds of conjugate gradient method by choosing various \( \tau_{k+1} \).

Hideaki and Yasushi proposed a new conjugate gradient method which was obtained by modifying the DY method and called MDY method. A nice property of the MDY method is that it generates sufficient descent directions. The parameter \( \beta_k \) in MDY method is given by

\[
\beta_k^{MDY} = \frac{g_k^T g_{k+1}}{\tau_{k+1}}. \tag{15}
\]

where

\[
\tau_{k+1} = \frac{2}{\alpha_k} (f_k - f_{k+1}). \tag{16}
\]

The definition of search direction and Formula (15) ensure that
\[ g_k ^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_k^{MD} g_k ^T d_k = (-\tau_{k+1} + g_k ^T d_k) \beta_k^{MD}, \quad \ldots \ldots \] (17)

and hence,
\[ 0 < \beta_k = -\frac{g_k ^T d_{k+1}}{\tau_{k+1} - g_k ^T d_k} = \frac{\psi_{k+1}}{\psi_k}. \quad \ldots \ldots \] (18)

performs more effective. More details can be found in [5].

Let us try to derive a new type method we need the next direction \( d_{k+1} \) to be descent. Assume that \( \beta_k > 0 \). By this, we have for any \( \xi \in (0, 1) \), and from (13) the following inequality holds:
\[ \beta_k \tau_{k+1} > \beta_k g_k ^T d_k, \quad \ldots \ldots \] (19)
i.e.,
\[ -\xi\|g_{k+1}\|^2 < \beta_k \tau_{k+1} + \beta_k g_k ^T d_k < 0. \quad \ldots \ldots \] (20)

Now we can rewrite the above inequality as
\[ g_k ^T \left[ -\xi - \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2} \right] g_{k+1} + \beta_k d_k < 0. \quad \ldots \ldots \] (21)

Hence, we obtain our new directions as follows:
\[ d_{k+1} = -\left( \xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2} \right) g_{k+1} + \beta_k d_k. \quad \ldots \ldots \] (22)

Then we can rewrite (22) as
\[ d_{k+1} = -\varphi_k g_{k+1} + \beta_k d_k, \quad \ldots \ldots \] (23)
where
\[ \varphi_k = \xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2}. \quad \ldots \ldots \] (24)

This method includes the Zhang and Wang (ZW) method as a special case. By setting \( \tau_{k+1} = y_k ^T d_k \), direction (22) reduces to the Zhang and Wang (ZW) method which defined in (10).

Now we can obtain the a new conjugate gradient algorithms, as follows:
The New Algorithm (2.1)

**Step 1.** Initialization : Select \( x_1 \in R^n \) and the parameters \( 0 < \delta_1 < \delta_2 < 1 \). Compute \( f(x_1) \) and \( g_1 \). Consider \( d_1 = -g_1 \) and set the initial guess \( \alpha_1 = 1/\|g_1\| \).

**Step 2.** Test for continuation of iterations. If \( \|g_{k+1}\| \leq 10^{-6} \), then stop. else step3.

**Step 3.** Line search : Compute \( \alpha_{k+1} > 0 \) satisfying the Wolfe line search condition (6) and update the variables \( x_{k+1} = x_k + \alpha_k d_k \).

**Step 4.** Conjugate gradient parameter which defined in (6) or (7) or (8).

**Step 5.** Direction computation \( d_{k+1} \) which defined in (22). If the restart criterion of Powell \[ \left| g_k ^T g_k \right| \geq 0.2 \|g_{k+1}\|^2 \], is satisfied, then set \( d_{k+1} = -g_{k+1} \) otherwise define \( d_{k+1} \).

**GLOBAL CONVERGENCE**

In this section, we establish convergence of the proposed method, the following assumptions for the objective function are needed.

Assumption (3.1)
The level set \(L = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}\) is bounded.

ii- In some neighborhood \(U\) of \(L\), \(f(x)\) is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant \(\mu > 0\) such that
\[
\|g(x_{k+1}) - g(x_k)\| \leq \mu\|x_{k+1} - x_k\|, \quad \forall x_{k+1}, x_k \in U. \quad \text{...........(25)}
\]
Assumption 3.1 imply that there exists a positive constant \(\gamma\) such that
\[
\|g_{k+1}\| \leq \gamma, \quad \forall x \in U. \quad \text{...........(26)}
\]

Here we have to present sufficient descent property.

**Theorem 1**
Let \(\{x_{k+1}\}\) and \(\{d_{k+1}\}\) be generated by (2) and (22), where \(\alpha_k\) satisfies Wolfe line search conditions, then holds of the sufficient descent property
\[
g_{k+1}^T d_{k+1} < -c\|g_{k+1}\|^2. \quad \text{...........(27)}
\]

**Proof:**
Then conclusion can be proved by induction. When \(k = 0\), we have \(g_0^T d_0 < -\|g_0\|^2 < 0\).
Suppose that \(g_k^T d_k < -c\|g_k\|^2\). From (13) and (22) we have
\[
g_{k+1}^T d_{k+1} = g_{k+1}^T \left[ -\left( \xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|} \right)g_{k+1} + \beta_k d_k \right] \quad \text{...........(28)}
\]
\[
\leq -\xi\|g_{k+1}\|^2 - \beta_k \xi + \beta_k g_{k+1}^T d_k
\]
\[
\leq -\xi\|g_{k+1}\|^2 - \beta_k g_{k+1}^T d_k + \beta_k g_{k+1}^T d_k
\]
\[
g_{k+1}^T d_{k+1} \leq -\xi\|g_{k+1}\|^2 \leq -c\|g_{k+1}\|^2. \quad \text{...........(29)}
\]
where \(c = \xi\). Thus the theorem is proved.

The following Lemma [9] is the result for general iterative methods :

**Lemma 1**
Suppose that Assumption 2.2 is satisfied and consider any method with Eq. (2), where \(\alpha_k\) satisfies Eqs. (11) and (12). Then,
\[
\sum_{j=1}^{\infty} \frac{(g_{j+1}^T d_{j+1})^2}{\|d_{j+1}\|^2} < \infty. \quad \text{...........(30)}
\]
From the previous analysis, we can get the following global convergence result for new Algorithm.

**Theorem 2**
Suppose that Assumption (3.1) holds, and these methods have the satisfies sufficient descent condition with \(c = \xi\). Then these method are globally convergent, one has
\[
\lim inf_{k \to \infty} \|g_{k+1}\| = 0 \quad \text{or} \quad \sum_{j=1}^{\infty} \frac{(g_{j+1}^T d_{j+1})^2}{\|d_{j+1}\|^2} < +\infty. \quad \text{...........(31)}
\]

**Proof:**
Now we will prove global convergence. We suppose that the theorem is not true. Suppose by contradiction that there exists \(\varepsilon_1 > 0\) such that
\[ d_0 = -g_0 \quad \text{and} \quad \|g_{k+1}\| > \varepsilon_1 \] ...........(32)

By squaring the sides of (22) and transferring and trimming, we get:

\[ \|d_{k+1}\|^2 = \beta^2 \|d_k\|^2 - \left( \xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|} \right)^2 \|g_{k+1}\|^2 \] ...........(33)

Dividing the previous in equation by \((d^T_{k+1}g_{k+1})^2\), we get:

\[ \frac{\|d_{k+1}\|^2}{(d^T_{k+1}g_{k+1})^2} = \frac{\beta^2 \|d_k\|^2}{(d^T_{k+1}g_{k+1})^2} - \left( \xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|} \right)^2 \frac{\|g_{k+1}\|^2}{(d^T_{k+1}g_{k+1})^2} \] ...........(34)

\[ \frac{\|d_{k+1}\|^2}{(d^T_{k+1}g_{k+1})^2} = \frac{\beta^2 \|d_k\|^2}{(d^T_{k+1}g_{k+1})^2} - \left( \xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|} \right)^2 \frac{\|g_{k+1}\|^2}{(d^T_{k+1}g_{k+1})^2} \] ...........(35)

\[ \frac{\|d_{k+1}\|^2}{(d^T_{k+1}g_{k+1})^2} = \frac{\beta^2 \|d_k\|^2}{(d^T_{k+1}g_{k+1})^2} - \left( \xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|} \right)^2 \frac{\|g_{k+1}\|^2}{(d^T_{k+1}g_{k+1})^2} \] ...........(36)

\[ \frac{\|d_{k+1}\|^2}{(d^T_{k+1}g_{k+1})^2} \leq \frac{\beta^2 \|d_k\|^2}{(d^T_{k+1}g_{k+1})^2} + \frac{1}{(d^T_{k+1}g_{k+1})^2} \] ...........(37)

a. When \( \beta_k = \beta_k^{Dv} \). Then by (8b)

\[ \beta_k^{Dv} = \frac{g^T_{k+1}d_{k+1}}{g_k d_k} \] ...........(38)

and applying (37), we have

\[ \frac{\|d_{k+1}\|^2}{(d^T_{k+1}g_{k+1})^2} \leq \frac{\|d_k\|^2}{(d^T_k g_k)^2} + \frac{1}{\|g_{k+1}\|^2} \] ...........(39)

Noting that

\[ \frac{\|d_k\|^2}{(d^T_k g_k)^2} = \frac{1}{\|g_k\|^2} \] ...........(40)

With this, from (32) we have

\[ \frac{\|d_{k+1}\|^2}{(d^T_{k+1}g_{k+1})^2} \leq \sum_{i=1}^{k+1} \frac{1}{\|g_{i+1}\|^2} \leq \frac{k}{\varepsilon_1^2} \] ...........(41)

Then we get

\[ \frac{(d^T_{k+1}g_{k+1})^2}{\|d_{k+1}\|^2} \geq \frac{\varepsilon_1^2}{k} \] ...........(42)
Which indicates
\[
\sum_{k=1}^{\infty} \frac{(d_{k+1}^T g_{k+1})^2}{\|d_{k+1}\|^2} \geq \sum_{k=1}^{\infty} \frac{\xi_k^2}{k} = \infty .
\] ...........(43)

This is a contradiction to the \((30)\).

b. When \(\beta_k = \beta_k^{FR}\). Then from \((37)\) and sufficient descent condition with \(c = \xi\).
\[
\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} \leq \frac{\|g_{k+1}\|^2}{\|d_k\|^2} \left(\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|}\right)^2 + \frac{1}{\|g_{k+1}\|^2} \leq \frac{\|d_k\|^2}{c^2\|g_k\|^2} + \frac{1}{\|g_{k+1}\|^2}
\] ............(45)

we also have
\[
\sum_{k=1}^{\infty} \frac{(d_{k+1}^T g_{k+1})^2}{\|d_{k+1}\|^2} \geq \sum_{k=1}^{\infty} \frac{\xi_k^2}{k} = \infty .
\]

c. When \(\beta_k = \beta_k^{CD}\). Then from \((37)\) and sufficient descent condition with \(c = \xi\).
\[
\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} \leq \frac{\|g_{k+1}\|^2}{\|d_k\|^2} \left(\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|}\right)^2 + \frac{1}{\|g_{k+1}\|^2} \leq \frac{\|d_k\|^2}{c^2\|g_k\|^2} + \frac{1}{\|g_{k+1}\|^2}
\] ............(46)

we also have
\[
\sum_{k=1}^{\infty} \frac{(d_{k+1}^T g_{k+1})^2}{\|d_{k+1}\|^2} \geq \sum_{k=1}^{\infty} \frac{\xi_k^2}{k} = \infty .
\]

which contradicts Lemma 1. Therefore, we get this theorem.

**NUMERICAL RESULTS**

In this section, we will report the numerical performance of Algorithm (2.1). We test Algorithm (2.1) by solving the 15 benchmark problems from [1] and compare its numerical performance with that of the other similar methods, which include the standard FR conjugate gradient method in [3]. All codes of the computer procedures are written in Fortran.

The parameters are chosen as follows:
\[
\epsilon = 10^{-6} , \quad \xi = 0.5 , \quad \delta_1 = 0.001 , \quad \delta_2 = 0.9
\]

In Tables 1 and 2, we use the following denotations:

- \(n\) : the dimension of the objective function.
- NOI : the number of iterations.
- NOF : the number of function evaluations.
- FR : the standard FR conjugate gradient method in [3].
- SFR : the new spectral FR method presented in this paper.
- SDY : the new spectral DY method presented in this paper.
- SCD : the new spectral CD method presented in this paper.
In this paper, a new spectral conjugate gradient algorithm has been developed for solving unconstrained minimization problems. Under some mild conditions, the global convergence has been proved. Compared with the other similar algorithm, the numerical performance of the developed algorithm is promising.

Table (4.1) Comparison of the algorithms for \( n = 100 \)

<table>
<thead>
<tr>
<th>Test problems</th>
<th>FR</th>
<th>SFR</th>
<th>SDY</th>
<th>SCD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOI</td>
<td>NOF</td>
<td>NOI</td>
<td>NOF</td>
</tr>
<tr>
<td>Trigonometric</td>
<td>22</td>
<td>40</td>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>Extended Rosenbrock</td>
<td>49</td>
<td>98</td>
<td>53</td>
<td>104</td>
</tr>
<tr>
<td>Extended White &amp; Holst</td>
<td>69</td>
<td>125</td>
<td>45</td>
<td>90</td>
</tr>
<tr>
<td>Penalty</td>
<td>47</td>
<td>1001</td>
<td>13</td>
<td>33</td>
</tr>
<tr>
<td>Generalized Tridiagonal 1</td>
<td>61</td>
<td>904</td>
<td>25</td>
<td>44</td>
</tr>
<tr>
<td>Extended ThreeExpo Terms</td>
<td>34</td>
<td>313</td>
<td>14</td>
<td>21</td>
</tr>
<tr>
<td>Generalized Tridiagonal 2</td>
<td>105</td>
<td>194</td>
<td>44</td>
<td>69</td>
</tr>
<tr>
<td>Diagonal 4</td>
<td>33</td>
<td>51</td>
<td>17</td>
<td>32</td>
</tr>
<tr>
<td>Extended Himmelblau</td>
<td>16</td>
<td>31</td>
<td>15</td>
<td>28</td>
</tr>
<tr>
<td>Broyden Tridiagonal</td>
<td>45</td>
<td>72</td>
<td>31</td>
<td>51</td>
</tr>
<tr>
<td>DENSCHNA (CUTE)</td>
<td>12</td>
<td>23</td>
<td>25</td>
<td>41</td>
</tr>
<tr>
<td>DENSCHNF (CUTE)</td>
<td>22</td>
<td>39</td>
<td>23</td>
<td>40</td>
</tr>
<tr>
<td>Extended Block-Diagonal</td>
<td>28</td>
<td>44</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>Generalized quartic GQ1</td>
<td>10</td>
<td>22</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>Generalized quartic GQ2</td>
<td>50</td>
<td>81</td>
<td>39</td>
<td>69</td>
</tr>
<tr>
<td>Total</td>
<td>603</td>
<td>3038</td>
<td>386</td>
<td>702</td>
</tr>
</tbody>
</table>

Table (4.2) Comparison of the algorithms for \( n = 1000 \)

<table>
<thead>
<tr>
<th>Test problems</th>
<th>FR</th>
<th>SFR</th>
<th>SDY</th>
<th>SCD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOI</td>
<td>NOF</td>
<td>NOI</td>
<td>NOF</td>
</tr>
<tr>
<td>Trigonometric</td>
<td>42</td>
<td>67</td>
<td>33</td>
<td>58</td>
</tr>
<tr>
<td>Extended Rosenbrock</td>
<td>60</td>
<td>113</td>
<td>65</td>
<td>126</td>
</tr>
<tr>
<td>Extended White &amp; Holst</td>
<td>237</td>
<td>299</td>
<td>58</td>
<td>115</td>
</tr>
<tr>
<td>Penalty</td>
<td>26</td>
<td>209</td>
<td>16</td>
<td>43</td>
</tr>
<tr>
<td>Generalized Tridiagonal 1</td>
<td>43</td>
<td>390</td>
<td>30</td>
<td>57</td>
</tr>
<tr>
<td>Extended ThreeExpo Terms</td>
<td>12</td>
<td>21</td>
<td>38</td>
<td>583</td>
</tr>
<tr>
<td>Generalized Tridiagonal 2</td>
<td>121</td>
<td>231</td>
<td>72</td>
<td>119</td>
</tr>
<tr>
<td>Diagonal 4</td>
<td>32</td>
<td>50</td>
<td>17</td>
<td>32</td>
</tr>
<tr>
<td>Extended Himmelblau</td>
<td>21</td>
<td>37</td>
<td>16</td>
<td>30</td>
</tr>
<tr>
<td>Broyden Tridiagonal</td>
<td>47</td>
<td>78</td>
<td>42</td>
<td>69</td>
</tr>
<tr>
<td>DENSCHNA (CUTE)</td>
<td>58</td>
<td>110</td>
<td>18</td>
<td>30</td>
</tr>
<tr>
<td>DENSCHNF (CUTE)</td>
<td>24</td>
<td>41</td>
<td>22</td>
<td>39</td>
</tr>
<tr>
<td>Extended Block-Diagonal</td>
<td>31</td>
<td>50</td>
<td>25</td>
<td>48</td>
</tr>
<tr>
<td>Generalized quartic GQ1</td>
<td>8</td>
<td>20</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>Generalized quartic GQ2</td>
<td>53</td>
<td>83</td>
<td>39</td>
<td>63</td>
</tr>
<tr>
<td>Total</td>
<td>742</td>
<td>1665</td>
<td>396</td>
<td>731</td>
</tr>
</tbody>
</table>

From the above numerical experiments, it is shown that the new algorithms in this paper is promising.

**CONCLUSIONS AND DISCUSSIONS**

In this paper, a new spectral conjugate gradient algorithm has been developed for solving unconstrained minimization problems. Under some mild conditions, the global convergence has been proved. Compared with the other similar algorithm, the numerical performance of the developed algorithm is promising.

Table (5.1) gives a comparison between the new-algorithm (2.1) and the Fletcher and Reeves (FR)-algorithm for convex optimization, this table indicates that the new algorithm (2.1) saves \((58 - 66)\)% NOI and \((30 - 43)\)% NOF, overall against the standard Fletcher and Reeves (FR)-algorithm, especially for our selected test problems. Relative Efficiency of the Different Methods Discussed in the Paper.
<table>
<thead>
<tr>
<th>Tools</th>
<th>NOI</th>
<th>NOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR- algorithm</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>SFR- algorithm</td>
<td>41.85%</td>
<td>69.53%</td>
</tr>
<tr>
<td>SDY- algorithm</td>
<td>33.97%</td>
<td>43.50%</td>
</tr>
<tr>
<td>SCD- algorithm</td>
<td>36.65%</td>
<td>61.00%</td>
</tr>
</tbody>
</table>

REFERENCES