Abstract

In this paper, we present modified conjugancy coefficient for the conjugate gradient method based on the (Liu and Storey) method to solve non-linear programming problems. We proved the sufficient descent and the global convergence properties for the proposed algorithm for three cases and we get very good numerical results especially for the large scale optimization problem.

الملخص

تم في البحث اشتقاق معامل ترافق محسن لطريقة المتجهات المترافقة أساسه صيغة (sufficient descent) لحل مسائل البرمجة غير الخطية وتم إثبات خصائص التقارب الكافي وخصوصية التقارب الشامل للخوارزمية المقترحة بثلاث حالات، كما تم الحصول على نتائج عددية جيدة جدا وخاصة لمسائل الأمثلية ذات القياس العالي.

Keyword: conjugate gradient, conjugancy coefficient, nonlinear programming, unconstrained optimization.
1. Introduction

In unconstrained optimization, we minimize an objective function which depends on real variables with no restrictions on the values of these variables. The unconstrained optimization problem is:

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

where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is a continuously differentiable function, bounded from below. A nonlinear conjugate gradient method generates a sequence \( x_k, k \) is integer number, \( k \geq 0 \). Starting from an initial point \( x_0 \), the value of \( x_k \) calculate by the following equation:

\[ x_{k+1} = x_k + \lambda_k d_k, \tag{2} \]

where the positive step size \( \lambda_k > 0 \) is obtained by a line search, and the directions \( d_k \) are generated as:

\[ d_{k+1} = -g_{k+1} + \beta_k d_k, \tag{3} \]

where \( d_0 = -g_0 \), the value of \( \beta_k \) is determine according to the algorithm of Conjugate Gradient (CG), and its known as a conjugate gradient parameter, \( s_k = x_{k+1} - x_k \) and \( g_k = \nabla f(x_k), f(x_k) = f(x_k) \), consider \( \| \cdot \| \) is the Euclidean norm and \( y_k = g_{k+1} - g_k \). The termination conditions for the conjugate gradient line search are often based on some version of the Wolfe conditions. The standard Wolfe conditions:

\[ f(x_k + \lambda_k d_k) - f(x_k) \leq \rho \lambda_k \langle g_k, d_k \rangle \]
\[ g(x_k + \lambda_k d_k)^T d_k \geq \sigma \langle g_k, d_k \rangle \]  

where \( d_k \) is a descent search direction and \( 0 < \rho \leq \sigma < 1 \), where \( \beta_k \) is defined by one of the following formulas:

\[ \beta_k^{(HS)} = \frac{y_k^T g_{k+1}}{y_k^T d_k} \quad \text{(Hestenese and Stiefel [6])} \]  
\[ \beta_k^{(FR)} = \frac{g_k^T g_{k+1}}{g_k^T g_k} \quad \text{(Fletcher and Reeves[4])} \]  
\[ \beta_k^{(PRP)} = \frac{y_k^T g_{k+1}}{g_k^T g_k} \quad \text{(Polak – Ribiere [8]and Poljak [9])} \]  
\[ \beta_k^{(CD)} = -\frac{g_k^T g_{k+1}}{g_k^T d_k} \quad \text{(Conjugate descent [5])} \]  
\[ \beta_k^{(LS)} = -\frac{y_k^T g_{k+1}}{g_k^T d_k} \quad \text{(Liu and Stoery [10])} \]  
\[ \beta_k^{(DY)} = \frac{g_k^T g_{k+1}}{y_k^T s_k} \quad \text{(Dai and Yuan [2])} \]  

May not often have better computational performances. In order to exploit the attractive feature of each set, the so called new conjugate gradient method have been proposed as:
2. The Proposed Conjugate Gradient Algorithm:

The proposed algorithm generates the iterate \( x_0, x_1, x_3, \ldots, x_n \) compute by the equation (2), the step size \( \lambda_k > 0 \) is determined according to the Wolfe conditions (4) and (5), and the direction \( d_k \) are generated by the equation (3):

The Liu and Storey Conjugate Gradient algorithm is one of the better methods to solve the large scale non-linear optimization problems, we proposed the new algorithm by using the Liu and Storey formula as:

Since

\[
\beta_k^{(LS)} = - \frac{g_{k+1}^T y_k}{g_k^T d_k} \tag{12}
\]

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

\[
= - \frac{g_{k+1}^T g_{k+1} - g_{k+1}^T g_k}{g_k^T d_k}
\]

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

By using the relation \( u^T v = \|u\| \|v\| \cos(\theta) \); where \( \theta \) is the angle between the vectors \( u \) and \( v \) we get:

\[
= - \frac{\|g_{k+1}\|^2}{g_k^T d_k} + \frac{\|g_{k+1}\| \|g_k\| \cos(\theta_1)}{g_k^T d_k} \frac{\|g_{k+1}\| \|g_k\| \cos(\theta_2)}{g_k^T d_k}
\]

where \( \theta_1, \theta_2 \) are the angles between \( g_{k+1}, g_k \) and \( d_k, g_k \) respectively.

\[
= \beta_k^{(CD)} + \frac{\|g_{k+1}\| \|g_k\|}{\|g_k\| \|d_k\|} \omega, \tag{13}
\]

where:

\[
\omega = \frac{\cos(\theta_1)}{\cos(\theta_2)}, \quad \cos(\theta_2) \neq 0
\]

(since \( \beta_k^{(CD)} = - \frac{\|g_{k+1}\|^2}{g_k^T d_k} \)).

Then we have three cases:

**Case 1:** If \( \cos(\theta_1) = 0 \) then

\[
\beta_k^{(MLS)} = \beta_k^{(CD)} \tag{14a}
\]

**Case 2:** If \( \cos(\theta_1), \cos(\theta_2) > 0 \) or \( \cos(\theta_1), \cos(\theta_2) < 0 \) then

\[
\beta_k^{(MLS)} = \beta_k^{(CD)} + \left[ \frac{\|g_{k+1}\|}{\|d_k\|} \omega^+ \right] \tag{14b}
\]
where \( \omega^* = \frac{\cos(\theta_1)}{\cos(\theta_2)} > 0 \)

**Case 3:** If \( \cos(\theta_1) < 0 \) or \( \cos(\theta_2) < 0 \) then

\[
\beta_k^{(MLS)} = \beta_k^{(CD)} + \frac{\|g_{k+1}\|}{\|d_k\|} \omega^{-} \quad (14c)
\]

where \( \omega^{-} = \frac{\cos(\theta_1)}{\cos(\theta_2)} < 0 \)

3. **Outlines of the Proposed Algorithm:**

   **Step (1):** The initial step: We select starting point \( x_0 \in \mathbb{R}^n \), and we select the accuracy solution \( \varepsilon > 0 \) is a small positive real number and we find \( d_0 = -g_0 \), \( \lambda_0 = \text{Minary}(g_0) \), and we set \( k = 0 \).

   **Step (2):** The convergence test: If \( \|g_k\| \leq \varepsilon \) then stop and set the optimal solution is \( x_k \).

   Else, go to step (3).

   **Step (3):** The line search: We compute the value of \( \lambda_k \) by Cubic method and that satisfy the Wolfe conditions in Eqs. (4),(5) and go to step (4).

   **Step (4):** Update the variables: \( x_{k+1} = x_k + \lambda_k d_k \) and compute \( f(x_{k+1}), g_{k+1} \) and \( s_k = x_{k+1} - x_k, y_k = g_{k+1} - g_k \).

   **Step (5):** Check: if \( \|g_{k+1}\| \leq \varepsilon \) then stop. Else continue.

**Step (6):** The search direction: We compute the scalar \( \beta_k^{(MLS)} \) by use the equations (14) and (3) and set \( k = k + 1 \), and go to step (4).

4. **The Convergence Analysis:**

4.1 **Sufficient Descent Property:**

We will show that in this section the proposed algorithm which defined in the equations (14) and (3) satisfy the sufficient descent property which satisfy the convergence property.

**Theorem (4.1.1):**

The search direction \( d_k \) that generated by the proposed algorithm of modified CG satisfy the descent property for all \( k \), when the step size \( \lambda_k \) satisfied the Wolfe conditions (4),(5).

**Proof:** we will use the indication to prove the descent property, for \( k = 0 \), \( d_0 = -g_0 \Rightarrow d_0^T g_0 = -\|g_0\| < 0 \), then we proved that the theorem is true for \( k = 0 \), now assume that the theorem is true for any \( k \) i.e \( d_k^T g_k < 0 \) or \( s_k^T g_k < 0 \) since \( s_k = \lambda_k d_k \), now we will prove that the theorem is true for \( k + 1 \) then

\( d_{k+1} = -g_{k+1} + \beta_k^{(MLS)} d_k \)
Case 1: Multiply the both sides of the above equation by $g_{k+1}$ and we set the value of the scalar $\beta_k^{\text{MLS}}$ in equation (14a) and we get

$$d_k^T g_{k+1} = -g_k^T g_{k+1} + (\beta_k^{\text{MLS}}) g_{k+1}^T d_k$$

(15)

$$d_k^T g_{k+1} = \left\| g_{k+1} \right\|^2 - \frac{g_{k+1}^T g_{k+1}^T}{d_k^T} g_{k+1}^T d_k$$

$$d_k^T g_{k+1} = -\left\| g_{k+1} \right\|^2 (1 + \frac{g_{k+1}^T d_k}{d_k^T})$$

Then by using the Wolfe condition we get:

$$d_k^T g_{k+1} \leq -\left\| g_{k+1} \right\|^2 (1 + \rho)$$

(16)

$$= -\left\| g_{k+1} \right\|^2 (\rho + 1), \ c = (1 + \rho) > 0 \ \text{sufficient descent satisfied.}$$

Case 2: Multiply the both sides of the above equation by $g_{k+1}$ and we set the value of the scalar $\beta_k^{\text{MLS}}$ in equation (14b) and we get

$$d_k^T g_{k+1} = -g_k^T g_{k+1} + (\beta_k^{\text{MLS}}) g_{k+1}^T d_k$$

(17)

$$\beta_k^{\text{MLS}} = \beta_k^{\text{(CD)}} + \frac{\left\| g_{k+1} \right\|}{d_k} \omega^+$$

(18)

$$d_k^T g_{k+1} = -\left\| g_{k+1} \right\|^2 - \frac{g_{k+1}^T d_k}{g_k^T d_k} g_{k+1}^T d_k + \frac{\left\| g_{k+1} \right\|}{d_k} g_{k+1}^T d_k \omega^+$$

$$= -\left\| g_{k+1} \right\|^2 (1 + \frac{d_k^T g_{k+1}}{g_k^T d_k} - \frac{1}{\left\| d_k \left\| g_{k+1} \right\|}} g_{k+1}^T d_k \omega^+)$$

Then by using the Wolfe condition we get:

$$\leq -\left\| g_{k+1} \right\|^2 (1 + \rho - \frac{\rho g_k^T d_k}{d_k \left\| g_{k+1} \right\|} \omega^+)$$

$$= -\left\| g_{k+1} \right\|^2 (1 + \rho - \frac{\rho g_k^T d_k}{d_k \left\| g_{k+1} \right\|} \omega^+)$$

(19)

Since $g_k^T d_k \leq 0, \ c = \left( 1 + \rho - \frac{\rho g_k^T d_k}{d_k \left\| g_{k+1} \right\|} \left( \frac{\cos(\theta_1)}{\cos(\theta_2)} \right) \right) > 1$. Then the sufficient descent is satisfied.

Case 3: Multiply the both sides of the above equation by $g_{k+1}$ and we set the value of the scalar $\beta_k^{\text{MLS}}$ in equation (14c) and we get

$$d_k^T g_{k+1} = -g_k^T g_{k+1} + (\beta_k^{\text{MLS}}) g_{k+1}^T d_k$$

(20)

$$\beta_k^{\text{MLS}} = \beta_k^{\text{(CD)}} + \frac{\left\| g_{k+1} \right\|}{d_k} \omega^-$$

(21)
\[ d_{k+1}^T g_{k+1} = -\frac{||g_{k+1}||^2}{g_k^T d_k} + \frac{||g_{k+1}||}{d_k} g_k^T d_k \omega^- \]

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

(22)

Then by using the Wolfe condition and by using the relation 
\[ u^T v = ||u|| ||v|| \cos \theta \Rightarrow u^T v \leq ||u|| ||v|| \] (where \( \theta \) is the angle between the vectors \( u \) and \( v \)), we have:

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

\[ \text{sufficient descent condition is satisfied.} \]

Where \( c = \left(1 + \rho - \left(\frac{\cos(\theta)}{\cos(\theta)}\right)\right) \geq 1 \)

### 4.3 Global Convergence property:

**Assumption:**

We assume that:

(i) The level set \( S = x \in R^n : f(x) \leq f(x_0) \) is bounded.

(ii) In a neighborhood \( N \) of \( S \), the function \( f \) is continuously differentiable and its gradient is Lipschitz continuous, i.e. there exists a constant \( L > 0 \) such that

\[ ||g(x) - g(y)|| \leq L||x - y||, \text{ for all } x, y \in N. \]

Under these assumptions on \( f \), there exists a constant \( \Gamma \geq 0 \) such that \( ||g(x)|| \leq \Gamma \).

In [1,3] it is proved that, for any conjugate gradient method with strong Wolfe line search, the following general result holds.

**Lemma 4.3.1:**

Let assumptions (i) and (ii) hold and consider any conjugate gradient method (2) and (3), where \( d_k \) is a descent direction and \( \lambda_k \) is obtained by the strong Wolfe line search. If

\[ \sum_{k \geq 1} \frac{1}{||d_k||^2} = \infty \]  

(23)

Then

\[ \lim \inf_{k \to \infty} ||g_k|| = 0 \]  

(24)

For uniformly convex functions which satisfy the above assumptions, we can prove that the norm of \( d_{k+1} \) given by (3) is bounded above. Assume that the function \( f \) is a uniformly convex function.

Using lemma 4.3.1 the following result can be proved.
Theorem 4.3.2:  
Suppose that the assumptions (i) and (ii) hold. Consider the algorithm (2), (15).  
If $\|d_k\|$ tends to zero and there exists nonnegative constants $\eta_1$ and $\eta_2$ such that:

$$\|g_k\|^2 \geq \eta_1 \|d_k\|^2,$$
$$\|g_{k+1}\|^2 \geq \eta_2 \|d_k\|$$

and $f$ is a uniformly convex function, then

$$\lim \inf_{k \to \infty} \|g_k\| = 0$$

Proof:

Case 1: In this case we have:

$$\beta_k^{(MLS)} = \beta_k^{(CD)}$$

From eq.(25), we get:

$$\left| \beta_k^{MLS} \right| \leq \frac{\|g_{k+1}\|^2}{\eta_1 \|d_k\|} \leq \frac{\eta_2 \|d_k\|}{\eta_1 \|d_k\|}$$

since

$$\|d_{k+1}\| \leq \|g_{k+1}\| + \beta_k^{MLS} \|d_k\|$$

$$\|d_{k+1}\| \leq \Gamma + \eta_2 \eta_1$$

$$\sum_{k \geq l} \frac{1}{\|d_{k+1}\|^2} = \infty$$

$$\frac{1}{\left( \Gamma + \frac{\eta_2}{\eta_1} \right)^2} \sum_{k \geq l} 1 = \infty$$

Then:

$$\lim \inf_{k \to \infty} \|g_{k+1}\| = 0$$

Case 2: In this case we have:

$$\beta_k^{MLS} = \frac{\|g_{k+1}\|^2}{g_k^T d_k} + \frac{\|g_{k+1}\|_{w^+}}{\|d_k\|}$$

$$\left| \beta_k^{MLS} \right| \leq \frac{\|g_{k+1}\|^2}{g_k^T d_k} + \frac{\|g_{k+1}\|_{w^+}}{\|d_k\|}$$

From eq.(25), we get:

$$\left| \beta_k^{MLS} \right| \leq \frac{\eta_2 \|d_k\|}{\eta_1 \|d_k\|^2} + \frac{\Gamma w^+}{\|d_k\|} = \frac{\eta_2}{\eta_1 \|d_k\|} + \frac{\Gamma w^+}{\|d_k\|}$$

since
\[ \begin{align*}
& \|d_{k+1}\| \leq \|g_{k+1}\| + \beta_k^{\text{MLS}} \|d_k\| \\
& \|d_{k+1}\| \leq \Gamma + \left( \frac{\eta^2}{\eta^1} \|d_k\| + \Gamma W^+ \right) \|d_k\| = \Gamma + \left( \frac{\eta^2}{\eta^1} + \Gamma W^+ \right) \\
& \sum_{k \geq l} \frac{1}{\|d_{k+1}\|^2} = \infty \\
& \left( \Gamma + \left( \frac{\eta^2}{\eta^1} + \Gamma W^+ \right) \right)^2 \sum_{k \geq l} 1 = \infty \\
& \text{Then:} \\
& \lim \inf_{k \to \infty} \|g_{k+1}\| = 0 \\
\end{align*} \]

**Case 3:**

\[ \beta_k^{\text{MLS}} = -\frac{\|g_{k+1}\|^2}{g_k^T d_k} + \frac{\|g_{k+1}\| W^-}{\|d_k\|} \]

\[ |\beta_k^{\text{MLS}}| \leq \frac{\|g_{k+1}\|^2}{g_k^T d_k} + \frac{\|g_{k+1}\| W^-}{\|d_k\|} \]

By the similar way and by use the absolute value:

\[ \sum_{k \geq l} \frac{1}{\|d_{k+1}\|^2} = \infty \\
& \left( \Gamma + \left( \frac{\eta^2}{\eta^1} + \Gamma W^- \right) \right)^2 \sum_{k \geq l} 1 = \infty \\
& \text{Then:} \\
& \lim \inf_{k \to \infty} \|g_{k+1}\| = 0 \\
\]

**5. Computational and Results:**

In this section, we reported some numerical results obtained with the implementation of the new algorithm on a set of unconstrained optimization test problems. We have selected (10) large scale unconstrained optimization problems in extended or generalized form, for each test function we have considered numerical experiment with the number of variable \( n=1000,5000,10000 \).

Using the standard Wolfe line search conditions (4), (5), the stopping criteria is the \( \|g_{k+1}\| \leq 10^{-6} \). The programs were written in FORTRAN 90.

We compare our method namely (MLS) with the FR method (7). The preliminary numerical results of our tests are reported in Tables (1),(2) and (3).
first column “test fun.” The names of test functions, the second column “NOI” denote the number iterations, the third column “NOF” denoted the number of calculated functions and the fourth column ”MIN” denoted the minimum value. We compute:

\[
\cos \theta_1 = \frac{-g_{k+1}^T g_k}{\|g_{k+1}\| \|g_k\|} \quad \text{and} \quad \cos \theta_2 = -\frac{g_k^T d_k}{\|g_k\| \|d_k\|}.
\]

<table>
<thead>
<tr>
<th>Test Fun.</th>
<th>FR-CG algorithm</th>
<th>MCG algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOI</td>
<td>NOF</td>
</tr>
<tr>
<td>Non-diagonal</td>
<td>148</td>
<td>345</td>
</tr>
<tr>
<td>Wolfe</td>
<td>192</td>
<td>385</td>
</tr>
<tr>
<td>Wood</td>
<td>3354</td>
<td>17377</td>
</tr>
<tr>
<td>Rosen</td>
<td>175</td>
<td>440</td>
</tr>
<tr>
<td>Cubic</td>
<td>62</td>
<td>132</td>
</tr>
<tr>
<td>Edgar</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>Sum</td>
<td>31</td>
<td>173</td>
</tr>
<tr>
<td>Powell3</td>
<td>1002</td>
<td>2009</td>
</tr>
<tr>
<td>Dixon</td>
<td>396</td>
<td>795</td>
</tr>
<tr>
<td>Reciep</td>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>5377</td>
<td>21700</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test Fun.</th>
<th>FR-CG algorithm</th>
<th>MCG algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOI</td>
<td>NOF</td>
</tr>
<tr>
<td>Non-diagonal</td>
<td>109</td>
<td>266</td>
</tr>
<tr>
<td>Wolfe</td>
<td>208</td>
<td>419</td>
</tr>
<tr>
<td>Wood</td>
<td>4887</td>
<td>27294</td>
</tr>
<tr>
<td>Rosen</td>
<td>182</td>
<td>454</td>
</tr>
<tr>
<td>Cubic</td>
<td>63</td>
<td>134</td>
</tr>
<tr>
<td>Edgar</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>Sum</td>
<td>44</td>
<td>252</td>
</tr>
<tr>
<td>Powell3</td>
<td>4915</td>
<td>9835</td>
</tr>
<tr>
<td>Dixon</td>
<td>1999</td>
<td>4000</td>
</tr>
<tr>
<td>Reciep</td>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>12424</td>
<td>42698</td>
</tr>
</tbody>
</table>
Comparative Performance of the Two Algorithms for Group of Test Functions at N=10000

<table>
<thead>
<tr>
<th>Test Fun.</th>
<th>FR-CG algorithm</th>
<th>MCG algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOI</td>
<td>NOF</td>
</tr>
<tr>
<td>Non-diagonal</td>
<td>138</td>
<td>325</td>
</tr>
<tr>
<td>Wolfe</td>
<td>239</td>
<td>481</td>
</tr>
<tr>
<td>Wood</td>
<td>7969</td>
<td>34587</td>
</tr>
<tr>
<td>Rosen</td>
<td>183</td>
<td>456</td>
</tr>
<tr>
<td>Cubic</td>
<td>63</td>
<td>134</td>
</tr>
<tr>
<td>Edgar</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>Sum</td>
<td>54</td>
<td>258</td>
</tr>
<tr>
<td>Powell3</td>
<td>5648</td>
<td>11301</td>
</tr>
<tr>
<td>Dixon</td>
<td>528</td>
<td>1058</td>
</tr>
<tr>
<td>Reciep</td>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>14839</td>
<td>48644</td>
</tr>
</tbody>
</table>

6. References: