A new conjugate gradient algorithms using conjugacy condition for solving unconstrained optimization

Aseel M. Qasim¹, Zinah F. Salih², Basim A. Hassan³

¹Department of Mathematics, College of Education of Pure Sciences, University of Mosul, Nineveh, Iraq
²,³Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, Nineveh, Iraq

Article Info

ABSTRACT

The primarily objective of this paper which is indicated in the field of conjugate gradient algorithms for unconstrained optimization problems and algorithms is to show the advantage of the new proposed algorithm in comparison with the standard method which is denoted as. Hestenes Stiefel method, as we know the coefficient conjugate parameter is very crucial for this reason, we proposed a simple modification of the coefficient conjugate gradient which is used to derived the new formula for the conjugate gradient update parameter described in this paper. Our new modification is based on the conjugacy situation for nonlinear conjugate gradient methods which is given by the conjugacy condition for nonlinear conjugate gradient methods and added a nonnegative parameter to suggest the new extension of the method. Under mild Wolfe conditions, the global convergence theorem and lemmas are also defined and proved. The proposed method's efficiency is programming and demonstrated by the numerical instances, which were very encouraging.

Keywords:
Conjugacy condition
Conjugate gradient method
Global convergence
Sufficient descent property

1. INTRODUCTION

The conjugate gradient method is a group of very important smooth function minimization methods, which is have a large dimension. We are interesting with conjugate gradient method for locating the function’s local minimum,

\[ \min \{ f(x) | x \in \mathbb{R}^n \} \]  

where \( f : \mathbb{R}^n \rightarrow \mathbb{R}^1 \) is a nonlinear, smooth function whose gradient is denoted by \( g_k = \nabla f(x_k) \), the iteration method that the conjugate gradient (CG) method uses in it of the line search is:

\[ x_{0 \in \mathbb{R}^n}, \quad x_{k+1} = x_k + \alpha_k d_k \]  

where \( \alpha_k \) is the step size, in order to ensure the new method's global convergence, we will impose that the step length \( \alpha_k \) satisfies the Wolfe conditions which have a good property to prove convergence for more detail see [1]-[6]:

Journal homepage: http://ijeecs.iaescore.com
\[ f(x_k + \alpha_k d_k) - f(x_k) \leq \sigma_1 \alpha g_k^T d_k \] (3)

\[ g(x_k + \alpha_k d_k)^T \geq \sigma_2 g_k^T d_k \] (4)

where \(0 < \sigma_1 < \sigma_2 < 1\) see [7] and \(d_k\) is the search direction generation by:

\[ d_{k+1} = -g_{k+1} + \beta_k d_k \] (5)

where \(\beta_k\) is the CG update parameter, which is a scalar. It's worth noting that the formula specification of the update parameter is a critical component of any CG algorithm, which is why different CG algorithms have been proposed to conform to different choices of \(\beta_k\) in (5). In [8]-[13], the parameters are given by:

\[ \beta_{k}^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \] (6)

\[ \beta_{k}^{DY} = \frac{g_{k+1}^T g_{k+1}}{y_k^T d_k} \] (7)

\[ \beta_{k}^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2} \] (8)

\[ \beta_{k}^{HS} = \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k} \] (9)

\[ \beta_{k}^{BBSQ} = \frac{g_{k+1}^T g_{k+1}}{\alpha(g_k^T d_k)^2/2(f_k-f_{k+1})} \] (10)

Furthermore, some CG methods for unconstrained optimization problem are not globally convergent. As a result, the researchers have been working to improve CG methods that are both globally convergent and numerically efficient [14]. We present the motivation and new algorithm of the proposed method. The global convergence of the proposed algorithm is exhibited with distinct quite different computational efficiency.

2. THE PROPOSED METHOD

The conjugacy condition for nonlinear conjugate gradient methods is given by (11).

\[ y_k^T d_{k+1} = 0 \] (11)

Perry decided to extend the deadline of (11) as:

\[ d_{k+1}^T y_k = -g_{k+1}^T s_k \] (12)

Dai and Liao in [15] took Perry's approach and added a nonnegative parameter \(t\) to suggest the following extension of (12).

\[ d_{k+1}^T y_k = -t g_{k+1}^T s_k \] (13)

We suggest this modification to the numerator of \(\beta_{k}^{BBSQ}\) [13], update the parameter to obtain the:

\[ \beta_{k}^{BBSQ} = \frac{\|g_{k+1}\|^2 + t g_{k+1}^T s_k}{a(g_k^T d_k)^2/2(f_k-f_{k+1})} \] (14)

Following the Li et al. approach in [16] and \(\beta_{k}^{BBSQ}\) we modified the following parameter extension:

\[ \beta_{k}^{MBSQ} = \frac{\|g_{k+1}\|^2 + t g_{k+1}^T s_k}{a(g_k^T d_k)^2/2(f_k-f_{k+1})} = \beta_{k}^{BBSQ} + t \frac{g_{k+1}^T s_k}{a(g_k^T d_k)^2/2(f_k-f_{k+1})} \] (15)
where \( t \) is a nonnegative parameter, whose values have been calculated by the use of the conjugacy condition analysis. More of than we compute \( t \) numerically by multiplying search direction by \( y_k \) and by using (14) we obtain:

\[
y^T_k d_{k+1} = -y^T_k g_{k+1} + \beta^MBSO_k y^T_k d_k
\]

(16)

Now we substitute \( y^T_k d_{k+1} = -t g^T_k s_k \) if the direction is in exact (ILS) and so we have:

\[
-t g^T_k s_k = -y^T_k g_{k+1} + \frac{g_{k+1}^T g_{k+1} + t g_{k+1}^T s_k}{a(g^T_k d_k)^2} y^T_k d_k
\]

\[
-t g^T_k s_k - \frac{(t g^T_k s_k) y^T_k d_k}{a(g^T_k d_k)^2} = -y^T_k g_{k+1} + \frac{(g_{k+1}^T g_{k+1} + t y^T_k d_k)}{a(g^T_k d_k)^2} y^T_k d_k
\]

\[
t = \frac{y^T_k g_{k+1} - \frac{g_{k+1}^T s_k}{a(g^T_k d_k)^2} y^T_k d_k}{g_{k+1}^T s_k + \frac{g_{k+1}^T s_k y^T_k d_k}{a(g^T_k d_k)^2}}
\]

(17)

Substituting the value of \( t \) from (17) into (15) yields:

\[
\beta^MBSO_k = \frac{g_{k+1}^T g_{k+1} + t^2 g_{k+1}^T s_k}{a(g^T_k d_k)^2} \frac{y^T_k d_k}{y^T_k d_k}
\]

By taking \( g_{k+1}^T s_k \) from the numerator and denominator as common factor:

\[
g_{k+1}^T g_{k+1} + \frac{(y^T_k g_{k+1})a(g^T_k d_k)^2}{a(g^T_k d_k)^2} = \frac{y^T_k d_k}{a(g^T_k d_k)^2} + \frac{y^T_k d_k}{a(g^T_k d_k)^2} = \frac{y^T_k d_k}{a(g^T_k d_k)^2} + \frac{y^T_k d_k}{a(g^T_k d_k)^2}
\]

\[
\beta^MBSO_k = \frac{g_{k+1}^T g_{k+1} + t^2 g_{k+1}^T s_k}{a(g^T_k d_k)^2} + \frac{y^T_k d_k}{a(g^T_k d_k)^2} + \frac{y^T_k d_k}{a(g^T_k d_k)^2}
\]

(18)

2.1. Outlines of the new algorithm

Define \( k=0 \) and choose an initial point \( x_0 \in \mathbb{R}^n \)

If \( \| g_k \| = 0 \), then terminate, otherwise proceed next step.

Calculate the descent \( d_k \) using (5).

A new conjugate gradient algorithms using conjugacy condition for solving ... (Aseel M. Qasim)
Determine a step size $\alpha_k$ by Wolfe line search condition (3), (4).
Let $x_{k+1} = x_k + \alpha_k d_k$.
Estimate $\beta_k$ which defined in (18).
Set $k = k+1$ and proceed to step (2).

3. GLOBAL CONVERGENCE

The following assumption is needed to analyze the algorithm's global convergence.
a) The level set $\xi = \{x \in \mathbb{R}^m | f(x) \leq f(x_0)\}$ is bounded.
b) $f(x)$ is continuously differentiable and its gradient is Lipchitz continuous in some neighborhood $u$ of $\xi$.

3.1. Lemma (1)

Assume that assumptions a) and b) are true, and consider any iteration of the form (5), where $d_k$ is the path direction and the step size $\alpha_k$ satisfies the condition (3), (4), then the zoutendijk condition:

$$\sum_{k=1}^{\infty} \left( \frac{\|g_k\|^2}{\|d_k\|^2} \right)^2 < \infty$$

(19)
holds.

3.2. Lemma (2)

If $d_{k+1}$ is given in (5) and $\rho_k^{MSQ}$ is given in (18), the following result holds.

$$g_k^T d_{k+1} < -c \|g_{k+1}\|^2 \forall k$$

(20)

Proof:

We have $d_1 = -g_1$ then $d_1^T g_1 < 0$ this is by induction for $k=1$, then we conclude that $d_k^T g_k < 0, \forall k \geq 2$.

$$g_k^T d_{k+1} = -\|g_{k+1}\|^2 + \frac{g_k^T g_{k+1} + y_k^T g_{k+1}}{\alpha (g_k^T d_k)^2/2(f_k - f_{k+1}) + y_k^T d_k} g_k^T d_k$$

In [8], it follows from $|g_k^T d_k| \leq |\sigma g_k^T d_k|$ that $g_k^T d_k \leq -\sigma g_k^T d_k$ and $d_k^T y_k = d_k^T (g_k - g_k) \geq (\sigma - 1) g_k^T d_k$ and from $d_k^T g_k \leq c \|g_k\|^2$ by using powell restart equation (i.e $g_k^T g_k \geq 0.2 g_{k+1}^T g_{k+1}$)

$$g_k^T d_{k+1} \leq -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 + \|g_{k+1}\|^2 (-\sigma g_k^T d_k)}{\sigma (g_k^T d_k)^2/2(f_k - f_{k+1}) + (\sigma - 1) g_k^T d_k}$$

$$\leq -\|g_{k+1}\|^2 + \frac{-2\|g_{k+1}\|^2}{\sigma (g_k^T d_k)^2/2(f_k - f_{k+1}) + (\sigma - 1) g_k^T d_k}$$

From (3) and (4) we have

$$g_k^T d_{k+1} \leq -\|g_{k+1}\|^2 + \frac{-2\|g_{k+1}\|^2/2(f_k - f_{k+1})}{\alpha (g_k^T d_k)^2/2(f_k - f_{k+1}) + (\sigma - 1) g_k^T d_k}$$

$$\leq -\|g_{k+1}\|^2 + \frac{-\sigma g_{k+1}^2/2(f_k - f_{k+1})}{\alpha (g_k^T d_k)^2/2(f_k - f_{k+1}) + (\sigma - 1) g_k^T d_k}$$

$$\leq -\|g_{k+1}\|^2 + \frac{\sigma g_{k+1}^2/2(f_k - f_{k+1})}{\alpha (g_k^T d_k)^2/2(f_k - f_{k+1}) + (\sigma - 1) g_k^T d_k}$$

$$\leq -\|g_{k+1}\|^2 + \frac{2\|g_{k+1}\|^2}{1 + (\sigma - 1) g_k^T d_k}$$
Let \( \epsilon = \frac{2\sigma(2\delta)}{1+\sigma(1)(-2\delta)} \) and \( 0<\epsilon<1 \), then we have \( g_k^T d_{k+1} \leq -\|g_{k+1}\|^2 + \epsilon \|g_{k+1}\|^2 \)
\[ \therefore g_k^T d_{k+1} \leq -c \|g_{k+1}\|^2 \] (21)

By this mathematical induction, we reach the result.

3.3. Lemma (3)

Assume that \( x_t \) is a starting point for which assumption a) is hold. If the new method generates \( x_1, x_2, x_3 \), and so on then,
\[ \lim_{k \to \infty} \inf \|g_k\| = 0 \] (22)

Proof: Assume that the inference is false, that is, there is an appositive constant such that \( \|g_k\| \geq \rho \) for all \( k \), since
\[ d_{k+1} = -g_{k+1} + \beta_k d_k \] which can be written as
\[ d_{k+1} + g_{k+1} = \beta_k d_k \] (23)

and since
\[ \beta_k = \frac{g_k^T g_{k+1} \gamma_k g_{k+1}}{a(g_k^T d_k)^2 + 2 \gamma_k g_k^T d_k - (\delta \alpha g_k^T d_k)} \] and from Wolfe condition (3), (4)
\[ f_k - f_{k+1} \geq -\delta \alpha g_k^T d_k \] and from \( d_k^T y_k \geq (\delta - 1)g_k^T d_k \) [8]
\[ \therefore \beta_k \leq \frac{2 \|g_{k+1}\|^2 - 2(f_k - f_{k+1})}{a(g_k^T d_k)^2 + 2 \gamma_k g_k^T d_k - (\delta \alpha g_k^T d_k)} \]
\[ \leq \frac{2 \|g_{k+1}\|^2 - 2(f_k - f_{k+1})}{a(g_k^T d_k)^2 + 2 \gamma_k g_k^T d_k - (\delta \alpha g_k^T d_k)} \]
\[ \leq \frac{2 \|g_{k+1}\|^2 - 2(\alpha \delta g_k^T d_k)}{a(g_k^T d_k)^2 + 2(\sigma - 1) g_k^T d_k - (\delta \alpha g_k^T d_k)} \]
\[ \leq \frac{-\|g_{k+1}\|^2 - \alpha \delta}{g_k^T d_k \alpha + 2(\sigma - 1) - \delta \alpha} \]
\[ \leq \frac{-\|g_{k+1}\|^2 - \alpha \delta}{\|g_k\| \|g_k\| \|g_{k+1}\| - \alpha \delta} \] (24)

\[ \therefore |\beta_k| < \ell \] such that \( j \) is constant
\[ \|d_{k+1}\| = \|-g_{k+1}\| + |\beta_k|d_k\| \]
\[ \leq \|-g_{k+1}\| + |j|d_k\| \]
\[ \leq \ell + |j|\ell \] (25)

and we concluding the proof with this contradiction that is:
\[ \sum_{i=1}^{k} \frac{1}{\|d_k\|^2} \geq \frac{1}{(\ell + |j|\ell)^2} \sum 1 = \infty \] (26)

Which is contradiction with zoutendijk theorem therefore the algorithm is globally convergent.
4. RESULT AND DISCUSSION

Now, we preset the arithmetical experiments we test and compare our method with $B_k^{HS}$, the code is written in using Fortran 90 to apply these methods and the test function are selected from [20]. The stopping state is defined as $\|g_{k+1}\| \leq 10^{-6}$ as recommended by [21]. We used 14 test problems with dimension 100 and 1000. The computation results shown in Table 1 have the following meaning: the total number of iterations (NOI), the total number of restarts (NOR), and the total number of function evaluations (NOF).

Also, Table 2 shows the average efficiency of the new algorithm with respect to harmony search (HS) method. Furthermore, optimization problems used in many papers for example, see [22]-[25].

### Table 1. Numerical results of new algorithms and HS-CG algorithm

<table>
<thead>
<tr>
<th>Test Function</th>
<th>n</th>
<th>NOI</th>
<th>NOR</th>
<th>NOF</th>
<th>NOI</th>
<th>NOR</th>
<th>NOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadratic Diagonal Perturbed</td>
<td>100</td>
<td>48</td>
<td>11</td>
<td>86</td>
<td>52</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>167</td>
<td>31</td>
<td>291</td>
<td>171</td>
<td>36</td>
<td>302</td>
</tr>
<tr>
<td>Trigonometric</td>
<td>1000</td>
<td>38</td>
<td>22</td>
<td>68</td>
<td>33</td>
<td>19</td>
<td>64</td>
</tr>
<tr>
<td>Extended Beale</td>
<td>1000</td>
<td>12</td>
<td>7</td>
<td>24</td>
<td>14</td>
<td>8</td>
<td>27</td>
</tr>
<tr>
<td>Extended Himmelblau</td>
<td>1000</td>
<td>10</td>
<td>6</td>
<td>19</td>
<td>10</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>HIMMELBH (CUTE)</td>
<td>1000</td>
<td>10</td>
<td>6</td>
<td>19</td>
<td>11</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>Himmelblau</td>
<td>1000</td>
<td>6</td>
<td>3</td>
<td>13</td>
<td>6</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>Broyden Tridiagonal</td>
<td>1000</td>
<td>29</td>
<td>6</td>
<td>50</td>
<td>28</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>Almost Perturbed Quadratic</td>
<td>1000</td>
<td>90</td>
<td>25</td>
<td>141</td>
<td>112</td>
<td>40</td>
<td>175</td>
</tr>
<tr>
<td>EDENSCH (CUTE)</td>
<td>1000</td>
<td>283</td>
<td>78</td>
<td>448</td>
<td>346</td>
<td>106</td>
<td>547</td>
</tr>
<tr>
<td>DIXON3DQ (CUTE)</td>
<td>1000</td>
<td>28</td>
<td>10</td>
<td>51</td>
<td>38</td>
<td>22</td>
<td>466</td>
</tr>
<tr>
<td>DENSCHNA (CUTE)</td>
<td>1000</td>
<td>34</td>
<td>18</td>
<td>230</td>
<td>44</td>
<td>27</td>
<td>617</td>
</tr>
<tr>
<td>DENSCHNC (CUTE)</td>
<td>1000</td>
<td>343</td>
<td>120</td>
<td>689</td>
<td>487</td>
<td>139</td>
<td>762</td>
</tr>
<tr>
<td>DENSCHNB (CUTE)</td>
<td>1000</td>
<td>551</td>
<td>318</td>
<td>2001</td>
<td>543</td>
<td>3160</td>
<td></td>
</tr>
<tr>
<td>DENSCHNF (CUTE)</td>
<td>1000</td>
<td>2001</td>
<td>552</td>
<td>3155</td>
<td>579</td>
<td>617</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5780</td>
<td>1675</td>
<td>9469</td>
<td>6092</td>
<td>1827</td>
<td>10676</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Average efficiency of the new algorithm

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>NOI</th>
<th>New Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS algorithm</td>
<td>100%</td>
<td>94.88%</td>
</tr>
<tr>
<td>New Formula algorithm</td>
<td>100%</td>
<td>88.69%</td>
</tr>
</tbody>
</table>

5. CONCLUSION

A new CG-method based on the conjugacy condition have been derived with a new update parameter, which is demonstrated the theory of global convergence for one of the proposed methods under reasonable assumptions. According to numerical results, the proposed algorithms has outperformed the regular HS method on average. Furthermore, the efficiency of new algorithm has been explained clearly based on arithmetic results.

ACKNOWLEDGEMENTS

The authors are grateful to the University of Mosul/College of Computer Sciences and Mathematics for providing facilities that aided in the production of this paper.

REFERENCES

A new conjugate gradient algorithms using conjugacy condition for solving ...

(Aseel M. Qasim)