A NEW CONJUGATE GRADIENT COEFFICIENT FOR
UNCONSTRAINED OPTIMIZATION BASED ON DAI-LIAO

Alaa Luqman Ibrahim\textsuperscript{a,*}, Muhammad Amin Sadiq \textsuperscript{b}, Salah Gazi Shareef \textsuperscript{c}

\textsuperscript{a} Department of Mathematics, College of Science, University of Duhok, Duhok, Kurdistan Region, Iraq (alaa.ibrahim@uod.ac)
\textsuperscript{b} College of Administration and Economics, Cihan University Duhok, Kurdistan Region, Iraq (muhammad.math@uod.ac)
\textsuperscript{c} Department of Mathematics, Faculty of Science, University of Zakho, Kurdistan Region, Iraq (salah.shareef@uoz.edu.krd)

Received: Dec. 20-18 / Accepted: Feb., 2019 / Published: Mar., 2019

ABSTRACT:
Conjugate gradient method plays an enormous role in resolving unconstrained optimization problem, particularly for large scale. In this paper, a new conjugate gradient method for unconstrained optimization based on Dai-Liao (DL) formula by using Barzilai and Borwein step size. Our new method satisfies both descent and sufficient descent conditions. The numerical results show that the proposed algorithm is potentially efficient and performs better than with Polak and Ribiere (PR) algorithm, depending on number of iterations (NOI) and the number of functions evaluation (NOF).

KEYWORDS: conjugate gradient, unconstrained optimization, Barzilai and Borwein step size, descent and sufficient descent conditions.

1. INTRODUCTION
The gradient method has great role in developing optimization techniques, and it has been associated with some well-known methods such steepest method for the unconstrained minimization (Dai and Liao, 2001). One of the important methods which is designed to solve unconstrained optimization problem is called conjugate gradient method (CG).

\[
\min f(x) \quad \forall x \in \mathbb{R}^n \tag{1.1}
\]

Where \( f : \mathbb{R}^n \rightarrow \mathbb{R}, \) \( f \in C^1 \) and its gradient at point \( x_i \) denoted by \( g_i = \nabla f(x_i) \neq 0 \), especially when our dimension is large. Therefore, the nonlinear CG-method formula is given by:

\[
x_{i+1} = x_i + \lambda_i d_i, \quad i = 0, 1, 2, \ldots \tag{1.2}
\]

where \( \lambda_i > 0 \) is a step size, and \( d_i \) is a search direction which is determined by:

\[
d_i = \begin{cases} 
-g_0 & \text{if } i = 0 \\
-g_i + \beta_i d_i & \text{if } i \geq 1
\end{cases} \tag{1.3}
\]

where \( \beta_i \) represents the conjugate gradient coefficient parameter. See ([Hestenes and Stiefel, 1952], [Polak and Ribiere, 1969], [Polyak, 1969], [Fletcher and Reeves, 1964], [Liu, and Storey, 1991], [Fletcher, 1987], [Dai and Yuan, 1996] and [Dai and Liao, 2001]) to be familiar with some well-known classical formulas for \( \beta_i \). The parameter \( \beta_i \) of the classical formula is determined as follows:

\[
\beta_i^{PD} = \frac{g_i^T g_{i+1}}{g_i^T d_i} \tag{1.8}
\]

\[
\beta_i^{DY} = \frac{g_i^T g_{i+1}}{d_i^T d_i} \tag{1.9}
\]

\[
\beta_i^{DL} = \frac{g_i^T g_{i+1}}{d_i^T d_i}, \quad \text{where } t > 0 \tag{1.10}
\]

Consider \( || \cdot || \) denotes the Euclidean norm and \( y_i = g_{i+1} - g_i \).

The global convergence of above conjugate gradient methods is studied by many researchers see (Hager, and Zhang, 2006), (Liu, and Storey, 1991) and (Wolfe, 1969). To prove the convergence condition of these methods, we need the step size \( \lambda_i \) satisfies the following strong Wolfe conditions:

\[
f(x_i + \lambda_i d_i) \leq f(x_i) + \delta \lambda_i g_i^T d_i \tag{1.11}
\]

\[
g(x_i + \lambda_i d_i)^T d_i \leq -\rho g_i^T d_i \tag{1.12}
\]

Where \( 0 < \delta < \rho < 1 \).

However, the standard Wolfe conditions are employed to prove the convergence of numerical methods such as Quasi-Newton method (Wolfe, 1969):

\[
f(x_i + \lambda_i d_i) \leq f(x_i) + \delta \lambda_i g_i^T d_i \tag{1.13}
\]

\[
g(x_i + \lambda_i d_i)^T d_i \leq \rho g_i^T d_i \tag{1.14}
\]

The majority studied characteristics of CG are its global convergence properties. The global convergence of Fletcher and Reeves (FR) is proved by (Zoutendijk, 1970). Analyzed the global convergence of algorithm regarding the Fletcher and Reeves method by Wolfe condition (Al-Baali, 1985). Further, FR as a superior method was proved by (Powell, 1986). Furthermore, Neculai Andrei presented a collection of unconstrained optimization functions in 2008. This paper organized as follows: In section two, the new conjugate gradient (CG) algorithm is presented with its algorithm. In section three, the new conjugate gradient satisfies the descent condition and sufficient descent condition. In section four the numerical results, percentages, graphs and discussion are presented. Finally, we conclude in section five.

\* Corresponding author
This is an open access under a CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by-nc-sa/4.0/)
2. NEW CONJUGATE GRADIENT COEFFICIENT

The aim of this section is to derive a new conjugate gradient coefficient known as $\beta^{\text{New}}$ for unconstrained optimization from the Dai-Liao (DL) formula.

2.1 The new CG coefficient ($\beta^{\text{New}}$)

Barzilai and Borwein (1988), proposed a new value of $\lambda_i$ defined by

$$\lambda_i = \frac{\|x_i\|}{\|x_i - x_{i-1}\|}$$  \hspace{1cm} (2.1)

Consider $s_i = x_{i+1} - x_i = \lambda_i d_i = -\lambda_i g_i$ \hspace{1cm} (2.2)

By putting (2.2) in (2.1), we get

$$\lambda_i' = \lambda_i \frac{\|s_i\|}{\|s_i - s_{i-1}\|}$$

Let, $\beta^{\text{New}}_i = \frac{\lambda_i'}{\lambda_i} \beta^{\text{DL}}_i$  \hspace{1cm} (2.3)

Now, we replace $\lambda_i$ by $\lambda_i'$ in denominator of equation (2.3), and get new parameter as follows:

$$\beta^{\text{New}}_i = \frac{\lambda_i'}{\lambda_i} \beta^{\text{DL}}_i$$

After some algebraic operations, we obtain

$$\beta^{\text{New}}_i = \theta \frac{\beta^{\text{DL}}_i}{\beta^{\text{DL}}_i} + t \frac{\beta^{\text{DL}}_i}{\beta^{\text{DL}}_i}$$ \hspace{1cm} (2.4)

2.2 Outline of new CG algorithm (\(\beta^{\text{New}}\))

Step 1: Given $x_0 \in \mathbb{R}^n$.
Step 2: $i = 0$, $g_0 = \nabla f(x_0)$, $d_0 = -g_0 $ if $g_0 = 0$, stop. 
Step 3: Compute $\lambda_i$ by using cubic line search to minimize $f(x_i + \lambda d_i)$, (i.e.) $f_{i+1} \leq f_i$.
Step 4: Updating new point based on Eq. (1.2)
Step 5: Compute $g_{i+1}$, if $\|g_{i+1}\| \leq 10^{-5}$ stop, else determine $s_i = x_{i+1} - x_i$ and $y_i = g_{i+1} - g_i$.
Step 6: Evaluate $d_{i+1}$ by Eq. (1.3), where $\beta_i$ is computed by (2.4).
Step 7: If $\|g_{i+1}\| \leq \frac{\|g_i\|}{0.2}$ is satisfied go to step 2, else $i = i + 1$ and go to step 3


Clearly, any conjugate gradient coefficient should satisfy both the descent and sufficient descent conditions. Therefore, in this section, we are going to show that new conjugate gradient $\beta^{\text{New}}_i$ satisfies the descent and the sufficient descent conditions.

Theorem 3.1 Suppose that the sequence $\{x_i\}$ is generated by (1.2), then the search direction (1.3) with $\beta^{\text{New}}_i$ given as (2.4), satisfies the descent condition. i.e.

$$d_{i+1}^T g_{i+1} \leq 0$$  \hspace{1cm} (3.1)

Proof: By combining the equations (1.3) and (2.4), we obtain

$$d_{i+1} = -g_{i+1} + (\beta^{\text{PR}}_i - t \frac{\beta^{\text{DL}}_i}{\beta^{\text{DL}}_i}) d_i$$

After multiplying both sides of the equation (3.2) by $g_{i+1}^T$, the following formula is obtained

$$g_{i+1}^T d_{i+1} = -\|g_{i+1}\|^2 + \beta^{\text{PR}}_i g_{i+1}^T g_k + t \frac{\beta^{\text{DL}}_i}{\beta^{\text{DL}}_i} g_{i+1}^T d_i$$

Since $s_i = \lambda_i d_i$

$$g_{i+1}^T d_{i+1} = -\|g_{i+1}\|^2 + \beta^{\text{PR}}_i g_{i+1}^T g_k + t \frac{\beta^{\text{DL}}_i}{\beta^{\text{DL}}_i} g_{i+1}^T d_i$$  \hspace{1cm} (3.3)

If $d_{i+1}^T g_{i+1} = 0$, then the equation (3.3) gives $g_{i+1}^T d_{i+1} = -\|g_{i+1}\|^2 \leq 0$. Then the proof is completed.

On the other hand, if $d_{i+1}^T g_{i+1} \neq 0$. Hence, the first two terms of equation (3.3) are equal or less than zero because the PR algorithm satisfies the descent condition, i.e.

$$-\|g_{i+1}\|^2 + \frac{\beta^{\text{DL}}_i}{\beta^{\text{DL}}_i} g_{i+1}^T d_i \leq 0$$

Clearly $t, \lambda_i$, $g_{i+1}$, and $(g_{i+1}^T d_i)$ are positive. Here, we obtain the third term of equation (3.3) less than or equal to zero. Hence

$$g_{i+1}^T d_{i+1} = -\|g_{i+1}\|^2 + \beta^{\text{PR}}_i g_{i+1}^T g_k - t \lambda_i \frac{\beta^{\text{DL}}_i}{\beta^{\text{DL}}_i} g_{i+1}^T d_i \leq 0$$

**Theorem 3.2** Assume that $d_{i+1}$ given by (1.3) and (2.4) and the step size $\lambda_i$ is obtained by (1.11) and (1.12) then, the sufficient descent condition is satisfied, i.e.

$$d_{i+1}^T g_{i+1} \leq -\epsilon \|g_{i+1}\|^2$$  \hspace{1cm} (3.4)

Proof: It is clear that the first two terms of equation (3.3) are equal or less than zero. Therefore, the equation (3.3) can be written as follows:

$$g_{i+1}^T d_{i+1} \leq -(t \lambda_i \frac{\beta^{\text{DL}}_i}{\beta^{\text{DL}}_i} \|g_{i+1}\|^2 (3.5)$$

Which means (3.4) holds where $\epsilon = t \lambda_i \frac{\beta^{\text{DL}}_i}{\beta^{\text{DL}}_i}$.

4. NUMERICAL RESULTS

In this section, we examine the implementation of the new CG method. The tests include well-known nonlinear problems standard test functions (Andrei, 2008), with different dimensions $4 \leq n \leq 5000$. The FORTRAN 95 language is the program which is used to write the numerical algorithm.

<table>
<thead>
<tr>
<th>Test Function</th>
<th>Test No.</th>
<th>NOI</th>
<th>NOF</th>
<th>NOI</th>
<th>NOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wood</td>
<td>100</td>
<td>29</td>
<td>67</td>
<td>26</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>30</td>
<td>69</td>
<td>27</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>30</td>
<td>69</td>
<td>27</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>30</td>
<td>69</td>
<td>28</td>
<td>64</td>
</tr>
<tr>
<td>Central</td>
<td>100</td>
<td>22</td>
<td>159</td>
<td>22</td>
<td>158</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>22</td>
<td>159</td>
<td>22</td>
<td>158</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>23</td>
<td>171</td>
<td>23</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>23</td>
<td>171</td>
<td>23</td>
<td>171</td>
</tr>
<tr>
<td>Cubic</td>
<td>100</td>
<td>15</td>
<td>45</td>
<td>13</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>16</td>
<td>47</td>
<td>13</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>16</td>
<td>47</td>
<td>13</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>16</td>
<td>47</td>
<td>14</td>
<td>39</td>
</tr>
</tbody>
</table>
investigate the efficiency of its behavior. One can use
better than experimen
table 4 variables
methods. In this
Numerous studies of conjugate
Clearly, in Table 2 the NOI and NOF of the PR method are about 100%. That means, the new method has improvement of 10.7582% compared with standard method in NOI and 13.6596% in NOF. Generally, the New (CG) method was improved by 12.2089% compared with PR method.

5. CONCLUSION

Numerous studies of conjugate gradient lead to several of new methods. In this paper, a modified (CG) method for solving nonlinear unconstrained optimization in formula (2.4) based on Dai-Liao (DL) formula by Barzilai and Borwein step size is presented. We found that the new conjugate gradient satisfies both conditions (descent and sufficient descent) and it is reliable for functions up to 4 variables. Limited numerical experiments and comparisons show that, the new algorithm is better than CG (PR) according to the NOI and NOF. In future one can use our new CG in neural network in order to investigate the efficiency of its behavior.

The numerical results in Table (1) illustrate that, the new conjugate gradient method is more efficient than standard (PR) method with respect to NOI and NOF. Indeed, in central function we have not seen any deference in terms of NOI when \( n \leq 500 \), yet there is a small deference in terms of NOF. However, for the other functions the result of the new CG is more efficient starting from the very beginning iterations. Finally, it is clear that the new CG is reliable function compare with the standard one.

Table 2. Comparison improvement ratio between the \( \beta_1^{\text{New}} \) and PR methods

<table>
<thead>
<tr>
<th>Tools</th>
<th>CG (PR)</th>
<th>CG (New)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOI</td>
<td>100%</td>
<td>89.2418%</td>
</tr>
<tr>
<td>NOF</td>
<td>100%</td>
<td>86.3404%</td>
</tr>
</tbody>
</table>

Clearly,\( 84.82 \) % in NOF. Generally, the New (CG) method was improved by \( 12.2089 \) % compared with PR method.

REFERENCES