A New Modification of Conjugate Gradient Method with Global Convergence and Sufficient Descent Properties for Unconstrained Optimization Problems

Mouiyad Bani Yousef1, Mustafa Mamat2 and Mohd Rivaie2

1Department of Mathematics, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia
2Department of Computer Sciences and Mathematics, Universiti Technology MARA (UiTM) Terengganu, Campus Kuala Terengganu, Malaysia

Conjugate gradient method (CG) is an important method for solving nonlinear unconstrained optimization problems. They are well-known for their global convergence properties and low memory requirement. This paper presents a modified CG method that is globally convergent under strong Wolfe-Powell (SWP) line search. The numerical results show that the new modification is more efficient than other CG methods tested.

Keywords: conjugate gradient method (CG); unconstrained optimization; large-scale optimization; sufficient descent property; Global convergence; strong Wolfe-Powell (SWP) line search

I. INTRODUCTION

The general unconstrained optimization problem is defined by the following,

$$\min_{x \in \mathbb{R}^n} f(x),$$  \hspace{1cm} (1)

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuously differentiable function.

The CG method is a type of iterative algorithm that generates a sequence $x_k$ by

$$x_{k+1} = x_k + \alpha_k d_k, \hspace{1cm} k = 0, 1, 2, 3, \ldots$$  \hspace{1cm} (2)

where $x_k$ is the current iteration point and $x_{k+1}$ is the next iteration point. The parameter $\alpha_k > 0$ is the step length obtained by SWP line search conditions as follows:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k,$$  \hspace{1cm} (3)

$$|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma |g_k^T g_k|,$$  \hspace{1cm} (4)

where $0 < \delta < \sigma < 1$, $g_k = g(x_k) = \nabla f(x)$ is the gradient of the function and $d_k$ is the search direction. The formula of $d_k$ for CG method is defined by

$$d_k = \begin{cases} -g_k, & k = 0 \\ -g_k + \beta_k d_{k-1}, & k \geq 1. \end{cases}$$  \hspace{1cm} (5)

The parameter $\beta_k \in \mathbb{R}$ is known as the CG coefficient that characterizes different CG methods. Examples of formulas for $\beta_k$ include the Polak-Ribiere-Polyak (PRP) method (Polak, 1969; Polak, 1969), the Fletcher-Reeves (FR) method (Fletcher et al., 1964), the Hestenes-Stiefel (HS) method (Hestenes & Stiefel, 1952), the ‘Aini Rivaie-Mustafa (ARM) method (Aini, et al., 2016), the Liu Storey (LS) method (Liu & Storey, 1991), the Conjugate Descent (CD) method (Fletcher, 2013) and the Wei-Yao-Liu (WYL) method (Wei et al., 2006). They are formulated as follows:

$$\beta_{k,PRP}^k = \frac{g_k^T y_{k-1}}{|g_k|^2},$$  \hspace{1cm} (6)

$$\beta_{k,FR}^k = \frac{|g_k|^2}{|g_{k-1}|^2};$$  \hspace{1cm} (7)

*Corresponding author’s e-mail: md.magableh@yahoo.com
II. NEW FORMULA FOR CG METHOD

In this section, we propose a new formula for $\beta_k$ known as $\beta_k^{MMR}$, where MMR denote Mouiyad, Mustafa and Rivaie. The new CG coefficient $\beta_k^{MMR}$ is constructed by applying an alternative denominator in $\beta_k^{MMR}$. The formula is as follows:

$$
\beta_k^{MMR} = \frac{m_k \| g_k \|}{m_k \| g_{k-1} \|} - \frac{\| g_k \|^2}{\| d_{k-1} \|},
$$

where

$$
m_k = \frac{\| d_{k-1} + g_k \|}{\| d_{k-1} \|}.
$$

Note that

$$
\| g_k \|^2 - \frac{\| g_k \|^2}{\| d_{k-1} \|} g_k^T g_{k-1} - \frac{\| g_{k-1} \|^2}{\| g_{k-1} \|} g_k^T g_{k-1} = \beta_k^{FR}.
$$

Therefore

$$
0 \leq \beta_k^{MMR} \leq \beta_k^{FR}.
$$

Hence, according to the argument presented in (Gilbert & Nocedal, 1992), $\beta_k^{MMR}$ should obtain all of the advantages and properties of $\beta_k^{FR}$.

The algorithm is given as follows:

**Algorithm 1**

Step 1: Given $x_0 \in R^n$, Select $\varepsilon \geq 0$ and, $d_0 = -g_0$, $k = 0$. If $\| g_0 \| \leq \varepsilon$ then stop.

Step 2: Compute $\alpha_k$ by (SWP) line search. (5) and (6)

Step 3: Let $x_{k+1} = x_k + \alpha_k d_k$, $g_{k+1} = g(x_{k+1})$ if $\| g_{k+1} \| < \varepsilon$ then stop.
Step 4: Compute $\beta_k$ by (6) and generate $d_{k+1}$ by (4).

Step 5: Set $k = k + 1$ and go to Step 2.

III. CONVERGENCE ANALYSIS OF MMR METHOD

The following basic assumption are always necessary in the analysis to present the global convergence properties of CG methods with (SWP) line search.

Assumption 1.

(i) $f(x)$ is bounded from below on the level set

$$\Omega = \{ x \in \mathbb{R}^n \mid f(x) \leq f(x_0) \}.$$  

(ii) In some neighborhood $N$ of $\Omega$, $f$ is continuously differentiable and its gradient $g(x)$ is Lipschitz continuous, hence there exists a constant $L > 0$ such that

$$\| g(x) - g(y) \| \leq L \| x - y \|, \quad x, y \in N.$$  

In 1992, Gilbert and Nocedal introduced the property (*) which plays an important role in the studies of CG methods. This property means that the next research direction approaches to the steepest direction automatically when a small step-size generated, and the step-sizes are not produced successively (Zhang et al., 2012).

Property (*). Consider a CG method of the form (2) and (5). Suppose that, for all $k \geq 0$.

$$0 < \gamma \leq \| g(x) \| \leq \bar{\gamma}.$$  

Where $\gamma$ and $\bar{\gamma}$ are two positive constants. We say that the method has the property (*) if there exist constants $b > 1$, $\lambda > 1$ such that for

all $k$,

$$|\beta_k| \leq b, \quad \| S_k \| \leq \lambda$$  

implies $\|g_k\| \leq \frac{1}{2b}$ where

$$S_k = \alpha_k \beta_k.$$  

The following lemma shows that the new method $\beta_k^{MMR}$ has the property (*).

Lemma 1. Consider the method of form (2) and (5), suppose that Assumptions 1 hold, then, the method $\beta_k^{MMR}$ has the property (*).

Proof. Set $b = \frac{\sqrt{2(\gamma + \bar{\gamma})}}{\gamma} > 1$, $\lambda = \frac{\gamma}{4L\bar{\gamma}b}$ by (13) and (17) we have

$$\| \beta_k^{MMR} \| \leq \frac{\| g_k \|^2}{\| g_k \|^2 - \| s_k^T g_{K-1} \|} \leq \frac{\| g_k \|^2}{\| g_k \|^2 - \| g_{k-1} \|^2} \leq \frac{\| g_k \|^2}{\| g_k \|^2 - \frac{\gamma}{\gamma^3}} = \frac{b}{\gamma^3}$$

From the Assumption ii, (16) holds. If $\| S_k \| \leq \lambda$ then,

$$\| \beta_k^{MMR} \| \leq \frac{\| S_k \|^2}{\| S_k \|^2 - \| s_k^T g_{K-1} \|} \leq \frac{\| S_k \|^2}{\| S_k \|^2 - \| S_{k-1} \|^2} \leq \frac{2L\| S_k \|^2}{\gamma^2} = \frac{1}{2b}$$

The proof is finished.

IV. SUFFICIENT DESCENT

The sufficient descent condition should be satisfied as follows:

$$g_k^T d_k \leq -c \| g_k \|^2$$  

for $k \geq 0$ and $c > 0$.

Note that the CG coefficient, $\beta_k^{MMR}$ satisfies
The following theorem shows that the formula MMR with SWP line search possess the sufficient descent condition

**Theorem 1**

Suppose that the sequences \( \{ g_k \} \) and \( \{ d_k \} \) are generated by the method of form (2), (5) and (13), and the step length \( \alpha_k \) is determined by the (SWP) line search (3) and (4), then the sequence \( \{ d_k \} \) possesses the sufficient descent condition (18).

**Proof.** By the formulae (13), we have

\[
\beta_{k+1}^{MMR} g_{k+1} \leq \frac{\| s_k \|^2}{\| g_{k+1} \|^2} \left( \| d_{k+1} \|^2 + s_k \right)
\]

Hence, we obtain

\[
0 \leq \beta_{k+1}^{MMR} \leq \frac{\| s_k \|^2}{\| g_{k+1} \|^2}
\]

Using (4) and (19), we get

\[
\beta_{k+1}^{MMR} g_{k+1}d_k \leq \frac{\| s_k \|^2}{\| g_{k+1} \|^2} \| g_k \| \| -g_k \| \| d_k \| (20)
\]

By (5), we have

\[
g_{k+1}d_k = g_{k+1} + \beta_{k+1}d_k \]

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

We prove the descent property of \( \{ d_k \} \) by induction. Since \( g_0^T d_0 < 0 \)

By (20), we get

\[
\beta_{k+1}^{MMR} g_{k+1}d_k \leq \frac{\| s_k \|^2}{\| g_{k+1} \|^2} \sigma (-g_k d_k) (22)
\]

That is,

\[
\frac{\| s_k \|^2}{\| g_{k+1} \|^2} \sigma g_k^T d_k \leq \beta_{k+1}^{MMR} g_{k+1}d_k \leq -\frac{\| s_k \|^2}{\| g_{k+1} \|^2} \sigma g_k^T d_k (23)
\]

(21) and (23) deduce,

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

By repeating this process and the fact

\[
g_0^T d_0 = -\| g_0 \|^2\]

we have,

\[
-\sum_{j=0}^{\infty} \sigma^j g_{k+1}^T d_{k+1} \leq -2 + \sum_{j=0}^{\infty} \sigma^j (24)
\]

Since

\[
\sum_{j=0}^{\infty} \sigma^j (24) \text{ Can be written as}
\]

\[
- \frac{1}{1-\sigma} \leq -2 + \frac{1}{1-\sigma} (25)
\]

By making the restriction \( \sigma \in (0, \frac{1}{4}) \) we have

\[
g_{k+1}^T d_{k+1} < 0.
\]

So, by induction, \( g_k^T d_k < 0 \) holds for all \( k \geq 0 \).

Denote \( c = 2 - \frac{1}{1-\sigma} \) then, \( 0 < c < 1 \), and (25) turns out to be

\[
(c-2)\| g_k \|^2 \leq g_k^T d_k \leq -c \| g_k \|^2 (26)
\]

this implies that (18) holds. The proof is complete.
V. GLOBAL CONVERGENCE

The following condition known as Zoutendijk condition is used to prove the global convergence of nonlinear CG methods (Zoutendijk, 1970; Wolfe, 1969).

Lemma 2. Suppose that Assumptions 1 hold. Consider a CG method of the form (2) and (5), where \( d_k \) satisfies \( g_k^T d_k < 0 \), for all \( k \), and \( \alpha_k \) is obtained by (SWP) line search (3) and (4).

Then,

\[
\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \tag{27}
\]

The proof had been given by Wei et al. 2006 and Yuan et al. 2010. Gilbert and Nocedal (1992) introduced the following important theorem.

Theorem 2

Consider any CG method of form (2) and (5), that satisfies the following conditions:

1. \( \beta_k \geq 0 \)
2. The search directions satisfy the sufficient descent condition.
3. The Zoutendijk condition holds.
4. Property (*) holds. If the Lipschitz and boundedness Assumptions hold, then the iterates are globally convergent.

From (16), (18), (27) and Lemma 1, we found that the MMR method with the parameter \( 0 < \delta < \sigma < 1/4 \) satisfies all four conditions in theorem 2 under the strong Wolfe-Powell line search, so the method is globally convergent.

VI. NUMERICAL RESULTS

In this section, we compare the computational performance of MMR method with PRP, FR, WYL and ARM under SWP line search. These comparisons are based on number of iteration and CPU time. To perform the test, a set of twenty-five test functions with varying number of variables (2 ≤ n ≤ 10000) have been selected from Andrei (2008). For each test problems, we take four different initial points in order to study the global convergence properties of the new CG formula. All functions and initial points used are listed in Table 1.

Table 1. A list of problem functions

<table>
<thead>
<tr>
<th>No</th>
<th>Functions</th>
<th>Initial points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Six Hummp 2 (-3, -3) (21,21)</td>
<td>(-43, -43) (77,77)</td>
</tr>
<tr>
<td>2</td>
<td>Three Hummp 2 (21,21) (25,25)</td>
<td>(71,71) (67,67)</td>
</tr>
<tr>
<td>3</td>
<td>Zettl 2 (7,7) (18,18)</td>
<td>(-116, -116)</td>
</tr>
<tr>
<td>4</td>
<td>FELETCHCR 2(-2, -27) (-112, -112)</td>
<td>(101,101) (17,17)</td>
</tr>
<tr>
<td>5</td>
<td>Colville(28,..,28) (199,..,199)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Dixon and Price 2,4(12,..,12) (19,..,19)</td>
<td>(-108,..,-108)</td>
</tr>
<tr>
<td>7</td>
<td>Hager 2,4 (6,..,6) (16,..,16)</td>
<td>(-78,..,-78)</td>
</tr>
<tr>
<td>8</td>
<td>Raydan1 2,4 (7,..,7) (18,..,18)</td>
<td>(-89,..,-89)</td>
</tr>
<tr>
<td>9</td>
<td>Raydan 2 2,4 (-7,..,-7) (-75,..,-75)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>ARWAHEAD 2,4,10 (4,..,4) (21,..,21)</td>
<td>(80,..,80)</td>
</tr>
<tr>
<td>11</td>
<td>Freudenstein2,4,10 (-19,..,-19)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Extended2,4,10 (-4,..,-4) (5,..,5)</td>
<td>(18,..,18) (4.5,..,4.5)</td>
</tr>
<tr>
<td>13</td>
<td>Maratos (18,..,18) (-84,..,-84)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Generalized 2,4,10 (3,..,3)</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Tridiagonal 1 (14,..,14) (70,..,70)</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Generalized2,4,10,100,500,1000(11,,..,11)</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Quartic 1 (29,..,29) (87,..,87) (-80,..,-80)</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Extended 2,4,10,100,500,1000(107,..,107)</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Beale (-1. 3,..,-1.3) (72,..,72)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Extended 2,4,10,100,500,1000(5,..,5)</td>
<td></td>
</tr>
</tbody>
</table>
The algorithm is set to stop when \( \| g_k \| \leq 10^{-6} \). All codes are written in MATLAB version R2015a subroutine programming and run on a PC computer with Intel(R) Core™ i3-4005U CPU @ 1.70GHz processor, 4GB RAM and Windows 10 Professional operating system. For result analysis, we use Sigma Plot 10 program to graph the data based on the performance profile proposed by Dolan and More (2002).

Figures 1 and 2 show that MMR has the best efficiency in terms of number of iteration and CPU time and plots the performance of our new method with respect to other four other CG methods, PRP, FR, WYL, and AMR. The curve at the top left signifies the fastest solver, while the curve at the top left is the most robust. The best method should be the one at the top left and right of the performance profile.

In Figure 1, the curve for MMR method is at the top of other curves, while in Figure 2, it is mostly at the top. In addition, MMR manages to solve 100% of the test problems. Compared to that, the PRP, FR, WYL, and AMR methods only solve 92%, 94%, 97%, and 94% of the test problems, respectively. Hence, our new formula is the most robust amongst all the methods tested.
VII. CONCLUSION

In this paper, we have considered a new CG method called the MMR method. We proved that it satisfies the sufficient descent condition and possesses global convergence properties when used with SWP line search. The numerical results demonstrate promising results for the MMR method, in which it has shown higher efficiency and robustness than other tested CG parameters.

VIII. ACKNOWLEDGEMENTS

The authors are grateful to Universiti Sultan Zainal Abidin, for financial support under the fundamental research Grant Scheme (FRGS)/1/2017/STG06/UniSZA/01/1.

IX. REFERENCES


