The Convergence Properties of a New Type of Conjugate Gradient Methods

Mohd Rivaie¹, Abdelrhaman Abashar², Mustafa Mamat³, Ismail Mohd²

¹Department of Computer Science and Mathematics, Univesiti Teknologi MARA (UiTM) Terengganu, Campus Kuala Terengganu, Malaysia
²Department of Mathematics, Faculty of Science and Technology Universiti Malaysia Terengganu (UMT), Malaysia
³Department of Informatics and Computing Universiti Sultan Zainal Abidin (Unisza), Kampus Tembila, Malaysia

Abstract

Conjugate gradient (CG) methods are widely used for solving large scale unconstrained optimization problems. Many studies have been devoted to develop and improve these methods. In this paper, we compare our new CG coefficient ($\beta_k$) with the other classical formulas of CG methods under the exact line search. Numerical results have shown that our new $\beta_k$ performs better than these classical formulas. Our method also possesses global convergence properties.

Mathematics Subject Classification: 65K10, 49M37

Keywords: Conjugate gradient methods, exact line search, global convergence properties
1. Introduction

The general formula for unconstrained optimization problem can be expressed as

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

(1)

where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is continuously differentiable function. There are many known methods for solving (1), but the CG methods are the most popular. These CG methods are in the form of iterations given by

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \ldots$$

(2)

where \( x_k \) is a current iterate point and \( \alpha_k > 0 \) is a step size. The \( \alpha_k > 0 \) is computed using exact line search by the formula

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k),$$

(3)

The search direction \( d_k \) of CG methods are defined by

$$d_k = \begin{cases} -g_k & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 1, \end{cases}$$

(4)

where \( g_k \) denotes gradient of \( f(x) \) at the point \( x_k \). The \( \beta_k \in \mathbb{R} \) is a scalar known as the conjugate gradient coefficient. Some classical formulas for \( \beta_k \in \mathbb{R} \) are given as follows:

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

(5)

$$\beta_k^{PR} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2},$$

(6)

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}},$$

(7)

$$\beta_k^{LS} = \frac{g_k^T (g_k - g_{k-1})}{-d_{k-1}^T g_{k-1}},$$

(8)

$$\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}},$$

(9)

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}.$$
where $g_k$ and $g_{k-1}$ are the gradient of $f(x)$ at the point $x_k$ and $x_{k-1}$ respectively. The above formula are known as Hestenes-Stiefel (HS) [8], Fletcher and Reeves (FR) [14], Polak, Ribiere and Polyak (PRP) [1], Conjugate Descent (CD) by Fletcher [13], Liu and Storey (LS) [17], and Dai and Yuan (DY) [16]. Usually, the PRP method performs much better than the FR method. There many research conducted to study the properties of CG, see for example Al-Baali [9], Zoutendijk [4], Powell [10], Touati-Ahmed and Storey [2], Gilbert and Nocedal [6], Rivaie et al. [11], Hager and Zhang [15], and Yuan et al. [5].

In this paper, we propose our new $\beta_k$ and compared its performance with classical formulas of FR, PRP and RMIL method (see Rivaie et al. [11]). The remaining sections of the paper are arranged as follows. In section 2, a new CG formula and a general algorithm for CG methods are presented. In section 3, we showed the sufficient descent condition and the global convergence proof of our new method. Section 4 deals with the Numerical results. Lastly, our discussion and conclusions based on these comparisons are presented in section 5 and section 6 respectively.

2. The algorithm and new proposed method

In this section, we propose our new $\beta_k$ denoted as $\beta_k^{RAMI}$, where RAMI represent Rivaie, Abashar, Mustafa and Ismail. $\beta_k^{RAMI}$ is defined by:

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

The following algorithm is the general algorithm for CG method

Algorithm 2.1

Step 1: Initialization. Given $x_0$, set $k = 0$.

Step 2: Compute $\beta_k$ based on (5), (6) and (11).

Step 3: Compute $d_k$ based on (4). If $g_k = 0$, then stop.

Step 4: Compute $\alpha_k$ based on exact line search (3).

Step 5: Updating new point based on iterative formula (2).

Step 6: Convergent test and stopping criteria.

If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| \leq \varepsilon$ then stop.
Otherwise go to Step 1 with \( k = k + 1 \).

3. Global Convergence analysis

In this section, we study the global convergent properties of \( \beta_k^{RAMI} \) and we begin with the sufficient descent condition.

The sufficient descent condition is defined as follows

\[
g_k^T d_k \leq -C\|g_k\|^2 \quad \text{for} \quad k \geq 0, \ C > 0.
\]  

The following theorem shows that our new formula with exact line search will possess the sufficient descent condition.

**Theorem 1**

Suppose that the \( x_k \) and \( d_k \) are generated by the method of form (2), (4) and (11), and the step size \( \alpha_k > 0 \) determined by the exact line search then, condition (12) holds for all \( k \geq 0 \).

**Proof.** We proves Theorem 1 by Induction. If \( k = 0 \) then we already have \( g_0^T d_0 = -C\|g_0\|^2 \). Hence condition (12) holds true. Then we need to show that for \( k \geq 1 \), condition (12) will also holds true. From (4) multiply by \( g_{k+1}^T \) then

\[
g_{k+1}^T d_{k+1} = g_{k+1}^T (-g_{k+1} + \beta_{k+1} d_k) = -\|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k.
\]  

For exact line search, we know that \( g_{k+1}^T d_k = 0 \). Thus

\[
g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2.
\]

Hence, this condition holds true for \( k + 1 \). The proof is completed. To study the global convergence properties, we need to show that \( \beta_k^{RAMI} \) are always not less than zero

\[
\beta_{k+1}^{RAMI} \geq \frac{g_{k+1}^T (g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k)}{d_k^T (d_k - g_{k+1})} \geq \frac{\|g_{k+1}\|^2 - \|g_{k+1}\| \|g_{k+1}\| \|g_k\|}{\|d_k\|^2} = 0
\]  

We can simplified \( \beta_{k+1}^{RAMI} \).
Convergence properties of a new type of conjugate gradient methods

\[
\beta_{k+1}^{RAMI} = \frac{\|g_{k+1}\|^2 - \left\|g_{k+1}\right\| g_k^T g_{k+1} g_k}{\|d_k\|^2} \leq \frac{\left\|g_{k+1}\right\|^2 + \|g_{k+1}\| \left\|g_{k+1}\right\|}{\|d_k\|^2} \leq \frac{2\|g_{k+1}\|^2}{\|d_k\|^2}
\]  

(15)

We also needed the following assumption

**Assumption 1**

(i) The level set \( \Omega = \left\{ x \in \mathbb{R}^n \mid f(x) \leq f(x_0) \right\} \) is bounded, where \( x_0 \) is the starting point.

(ii) In some neighborhood \( N \) of \( \Omega \) the objective function is continuously differentiable, and its gradient is Lipschitz continuous, namely, there exists a constant \( l > 0 \) such that \( \|g(x) - g(y)\| \leq L\|x - y\| \) for any \( x, y \in N \).

Under this assumption, we have the following lemma, which was proved by Zoutendijk [4].

**Lemma 1**

Suppose Assumption 1 holds, let \( x_k \) be generated by Algorithm 2.1 and \( d_k \) satisfy

(12) then the following condition, known as the Zoutendijk condition, holds

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

(16)

The following theorem is based on Lemma 1.

**Theorem 2**

Suppose that Assumption 1 holds true \( x_k \) is generated by Algorithm 2.1, \( \alpha_k \) is obtained by the exact line search (3) and the sufficient descent condition hold true. Then either

\[
\lim_{k \to \infty} \|g_k\| = 0 \quad \text{or} \quad \sum_{k=0}^{\infty} \left( \frac{g_k^T d_k}{\|d_k\|} \right)^2 < \infty
\]

(17)

**Proof.** To prove theorem 2, we use contradiction. That is, if theorem 2 is not true, then a constant \( c > 0 \) exists, such that

\[
\|g_k\| \geq c
\]

Rewriting (4) as

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

And squaring both sides of the equation, we get

\[
\|d_{k+1}\|^2 = (\beta_{k+1})^2 \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2
\]

(18)

Dividing both side by \( (g_{k+1}^T d_{k+1})^2 \) then,
Applying (14) we get
\[
\frac{\|d_{k+1}\|^2}{(g_k^T d_{k+1})^2} = \left(\frac{\beta_{k+1}}{g_k^T d_{k+1}}\right)^2 - \frac{2}{g_k^T d_{k+1}} - \frac{\|g_{k+1}\|^2}{(g_k^T d_{k+1})^2}
\]
\[
= \left(\frac{\beta_{k+1}}{g_k^T d_{k+1}}\right)^2 - \left(\frac{1}{\|g_{k+1}\|} + \frac{g_k^T d_{k+1}}{\|g_{k+1}\|^2}\right)^2 + \frac{1}{\|g_{k+1}\|^2}
\]
\[
\leq \left(\frac{\beta_{k+1}}{g_k^T d_{k+1}}\right)^2 + \frac{1}{\|g_{k+1}\|^2}
\]
Applying (14) we get
\[
\leq \frac{1}{\|g_{k+1}\|^2}
\] (19)
Hence
\[
\frac{\|d_{k}\|^2}{(g_k^T d_{k})^2} \leq \sum_{i=0}^{k} \frac{1}{\|g_i\|^2}
\]
\[
\frac{(g_k^T d_{k})^2}{\|d_{k}\|^2} \geq \frac{c^2}{k}
\] (20)
There from (19) and (16), it follows that
\[
\sum_{k=0}^{\infty} \left(\frac{g_k^T d_{k}}{\|d_{k}\|}\right)^2 = \infty
\]
This contradicts the Zoutendijk condition in Lemma 1. The proof is completed.

**Theorem 3**
Suppose that Assumptions 1 holds, consider any CG the methods of form (2) and (4), \(\alpha_k\) obtained by the exact line search and \(\beta_k\) is determined by (11). Then either
\[
\lim_{k \to \infty} \|g_k\| = 0 \quad \text{or} \quad \sum_{k=0}^{\infty} \left(\frac{g_k^T d_{k}}{\|d_{k}\|}\right)^2 < \infty.
\]
**Proof.** From (17) and (15)
\[
\|d_{k+1}\|^2 = \left(\frac{2\|g_{k+1}\|^2}{\|d_{k}\|^2}\right)\|d_{k}\|^2 - 2g_k^T d_{k+1} - \|g_{k+1}\|^2
\]
\[
\|d_{k+1}\|^2 = \frac{4\|g_{k+1}\|^4}{\|d_{k}\|^2} - 2g^T_{k+1} d_{k+1} - \|g_{k+1}\|^2
\] (21)
We have already proven that sufficient descent condition holds. Therefore, we know that
Convergence properties of a new type of conjugate gradient methods

\[ g_k^T d_k \leq -C\|g_k\|^2. \]

Hence from (21),

\[ \|d_k\|^2 = \frac{4\|g_k\|^4}{\|d_k\|^2} + 2c\|g_k\|^2 - \|g_k\|^2 \]
\[ \|d_k\|^2 = \frac{4\|g_k\|^4}{\|d_k\|^2} - \|g_k\|^2(1 - 2c) \] (22)

Multiply both sides of (22) with \( \frac{\|g_k\|^2}{\|d_k\|^2} \) then,

\[ \|d_k\|^2 \|g_k\|^2 = \frac{\|g_k\|^2}{\|d_k\|^2} \left( \frac{4\|g_k\|^4}{\|d_k\|^2} - \|g_k\|^2(1 - 2c) \right) \]
\[ \frac{\|d_k\|^2 \|g_k\|^2}{\|d_k\|^2} = \frac{\|g_k\|^4}{\|d_k\|^2} \left( 2c \|g_k\|^2 + 4\|g_k\|^2 \|d_k\|^2 \right) \]
\[ \frac{\|d_k\|^2 \|g_k\|^2}{\|d_k\|^2} \leq \frac{\|g_k\|^4}{\|d_k\|^2} \] (23)

Based on theorem 2, we know that

\[ \lim_{k \to \infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < 0. \]

This will imply that if theorem 3 is not true, then we have

\[ \lim_{k \to \infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \infty \]

and from (23) we get \( \infty \leq \frac{\|g_k\|^2}{\|d_k\|^2} \). Hence, theorem 3 holds true for sufficiently large \( k \).

4. Numerical results

In this section, we carried out some numerical experiments to test Algorithm 2.1. We use some of the test problem considered in Andrei [12] as shown in table 1 to analyze the efficiency of our new formula as compared with the FR, PRP and RMIL method. The comparisons are based on the number of iterations and CPU time. We considered \( \varepsilon = 10^{-5} \) and \( \|g_k\| \leq \varepsilon \) as stopping criteria as mentioned by Hillstrom [7]. All problems mentioned in table 1 are solved by Maple 13 subroutine programming. The CPU processor used was
Core™ i3-2328M (2.2GHZ, 3MB L3 Cache), with 6GB DDR3 RAM. The performance results are shown in Figs.1 and 2, respectively, based on the performance profile introduced by Dolan and More [3]. This performance profile provided the means to evaluate and compare the performance of the set solvers $S$ on a test set $P$. Assuming $n_s$ solvers and $n_p$ problems exist, for each problem $p$ and solver $s$, they defined

$$t_{p,s} = \text{Computing time (the number of iterations, CPU time or others)}$$

required to solve problems $p$ by solver $s$.

Requiring a baseline for comparisons, they compared the performance on problem $p$ by solver $s$ with the best performance by any solver on this problem using the performance ratio

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s} : s \in S\}}.$$  

Suppose that a parameter $r_M \geq r_{p,s}$ for all $p,s$ is chosen, and $r_{p,s} = r_M$ if and only if solver $s$ does not solve problem $p$. The performance of solvers $s$ on any given problem might be of interest, but because we would like to obtain an overall assessment of the performance of the solver, then the following was defined:

$$\rho_s(t) = \frac{1}{n_p} \text{size}\{p \in P : r_{p,s} \leq t\}.$$  

Thus $\rho_s(t)$ was the probability for solver $s \in S$ that a performance ratio $r_{p,s}$ was within a factor $t \in R$ of the best possible ratio. Function $\rho_s$ was then the cumulative distribution function for the performance ratio. The performance profile $p_s : R \mapsto [0,1]$ for a solver was non-decreasing, piecewise, and continuous from the right. The value of $p_s(1)$ is the probability that the solver will win over the rest of the solvers. In general, a solver with high values of $p(t)$ or at the top right of the figure are preferable or represent the best solver.
**TABLE 1: A list of problem functions**

<table>
<thead>
<tr>
<th>No</th>
<th>Functions</th>
<th>n</th>
<th>Initial Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rosenbrock</td>
<td>10,100,500,1000</td>
<td>(13,13,…,13), (16,16,…,16), (20,20,…,20), (30,30,…,30)</td>
</tr>
<tr>
<td>2</td>
<td>Shallow</td>
<td>10,100,500,1000</td>
<td>(10,10,…,10),(25,25,…,25),(50,50,…,50), (100,100,…,100)</td>
</tr>
<tr>
<td>3</td>
<td>Tridiagonal -1</td>
<td>10,100,500,1000</td>
<td>(2,2,…,2), (4,4,…,4), (20,20,…,20), (100,100,…,100)</td>
</tr>
<tr>
<td>4</td>
<td>Strait</td>
<td>10,100,500,1000</td>
<td>(10,10,…,10),(50,50,…,50),(100,100,…,100), (200,200,…,200)</td>
</tr>
<tr>
<td>5</td>
<td>White and Holst</td>
<td>10,100,500,1000</td>
<td>(30,30,…,30),(35,35,…,35),(40,40,…,40), (73,73,…,73)</td>
</tr>
<tr>
<td>6</td>
<td>Quartic</td>
<td>10,100,500,1000</td>
<td>(80,80,…,80), (100,100,…,100), (200,200,…,200), (400,400,…,400)</td>
</tr>
<tr>
<td>7</td>
<td>Powell</td>
<td>10,100,500,1000</td>
<td>(5,5,…,5), (-5,-5,…,-5),(6,6,…,6),(7,7,…,7)</td>
</tr>
<tr>
<td>8</td>
<td>Denschnb</td>
<td>10,100,500,1000</td>
<td>(5,5,…,5), (8,8,…,8), (12,12,…,12), (30,30,…,30)</td>
</tr>
<tr>
<td>9</td>
<td>Denschnf</td>
<td>10,100,500,1000</td>
<td>(15,15,…,15),(20,20,…,20),(30,30,…,30), (100,100,…,100)</td>
</tr>
<tr>
<td>10</td>
<td>Wood</td>
<td>10,100,500,1000</td>
<td>(28,28,…,28), (40,40,…,40), (50,50,…,50), (100,100,…,100)</td>
</tr>
<tr>
<td>11</td>
<td>Maratos</td>
<td>10,100,500,1000</td>
<td>(13,13,…,13),(70,70,…,70),(150,150,…,150), (250,250,…,250)</td>
</tr>
<tr>
<td>12</td>
<td>Freudenstein and Roth</td>
<td>10,100,500,1000</td>
<td>(1,1,…,1), (-1,-1,…,-1), (-2,-2,…,-2), (-4,-4,…,-4)</td>
</tr>
<tr>
<td>13</td>
<td>Beale</td>
<td>10,100,500,1000</td>
<td>(13,13,…,13), (16,16,…,16), (20,20,…,20), (30,30,…,30)</td>
</tr>
<tr>
<td>14</td>
<td>Himmelblau</td>
<td>10,100,500,1000</td>
<td>(13,13,…,13),(30,30,…,30),(100,100,…,100), (180,180,…,180)</td>
</tr>
<tr>
<td>15</td>
<td>Tridiagonal -2</td>
<td>10,100,500,1000</td>
<td>(13,13,…,13), (16,16,…,16), (20,20,…,20), (30,30,…,30)</td>
</tr>
<tr>
<td>16</td>
<td>Engvali</td>
<td>10,100,500,1000</td>
<td>(50,50,…,50), (100,100,…,100), (200,200,…,200), (-2,-2,…,-2)</td>
</tr>
</tbody>
</table>
Figure 1 Performance profile based on the number of iteration

Figure 2 Performance profile based on the CPU time
5. Discussion

Figure 1 and 2 show that the performance of these methods is relative to the number of iteration and CPU time respectively. We show that our proposed method is better when compared with FR which solves 69% of test problem, RMIL method solve 88% of test problem. Although the PRP seems to be better than our new method but it can only solve 95% of test problem. Therefore, we rate our new formula as the best since it can solve all the test problem functions.

6. Conclusion

In this paper, we proposed a new and simple $\beta_k$ that has global convergence properties. Numerical results have shown that this new $\beta_k$ performs better than FR, PRP and RMIL. In near future, we intend to test our new formula using inexact line search, new scaling factor and apply it to three terms CG methods.

Acknowledgements. The authors would like to thank the University of Malaysia Terengganu (UMT) for funding under FRGS Vot 59256 and Red Sea University in Sudan.

References


**Received:** October 17, 2013