New Hybrid Conjugate Gradient Method with Global Convergence Properties for Unconstrained Optimization

*Yasir Salih*a, Mohamed Ahmed Hamodab and Mohd Rivaiec

*aDepartment of Mathematics, Faculty of Education, Red Sea University, Sudan
bDepartment of Mathematics, Faculty of Science, Al-Asmara Islamic University, Libya
cDepartment of Computer Science and Mathematics, Universiti Teknologi MARA, (UiTM)Terengganu Campus Kuala Terengganu, Malaysia

yasirsalih2015@gmail.com

Received: 11/04/2018, Accepted: 15/06/2018

Abstract

Nonlinear conjugate gradient (CG) method holds an important role in solving large-scale unconstrained optimization problems. In this paper, we suggest a new modification of CG coefficient $\beta_k$ that satisfies sufficient descent condition and possesses global convergence property under strong Wolfe line search. The numerical results show that our new method is more efficient compared with other CG formulas tested.

Keywords: conjugate gradient method, large-scale, global convergence, strong Wolfe line search, unconstrained optimization.

Introduction

The general form of an unconstrained optimization problem is defined by

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

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuously differentiable function and its gradient $g \equiv \nabla f(x)$ is available.

The iterative formula of the CG method is given by

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

where $\alpha_k$ is the step-size computed by carrying out strong Wolfe line search procedure, defined as follows

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k$$

and

$$|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma |g_k^T d_k|$$

where $0 < \delta < \sigma < 1$ The parameter $d_k$ is the search direction defined by

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

$$\beta_k \text{satisfies sufficient descent condition and possesses global convergence property under strong Wolfe line search. The numerical results show that our new method is more efficient compared with other CG formulas tested.}$$
where $\beta_k \in \mathbb{R}$ is a scalar known as the CG coefficient. Examples of most well-known classical formulas for $\beta_k$ are Hestenes-Stiefel (HS) ([Hestenes and Stiefel, 1952]), Fletcher-Reeves (FR) ([Fletcher and Reeves, 1964]), Polak-Ribiere-Polyak (PRP) ([Polak and Ribiere, 1969]), Conjugate Descent (CD) ([Fletcher, 1980]), Liu-Storey (LS) ([Liu and Storey, 1991]), and Dai-Yuan (DY) ([Dai and Yuan, 1999]). The parameters of these $\beta_k$ are given as follows:

$$
\beta_{k^+}^{HS} = \frac{\beta_k^T (g_k - g_{k-1})}{\beta_k^T (g_k - g_{k-1})}, \quad \beta_{k^+}^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_{k^+}^{PRP} = \frac{\beta_k^T (g_k - g_{k-1})}{g_k^T g_{k-1}}.
$$

The convergence of CG method under different line searches has been studied by many authors such as Al-Baali (1985), Gilbert and Nocedal (1992), Liu and Storey (1991), Zoutendijk (1970), Touati-Ahmed and Storey (1990), and Andrei (2008). For further information, readers can refer to Abashar et al. (2017), Aini et al. (2017), Ghani et al. (2017a), Ghani et al. (2017b), Rivaie et al. (2017), Kamfa et al. (2017), Mohamed et al. (2017), Omer et al. (2015), Osman et al. (2017), Rivaie et al. (2015), and Zull et al. (2017).

**Modified Formula and Algorithm**

Recently, Wei et al. (2006) gave a variant of the PRP method which is called the WYL method, written as

$$
\beta_{k^+}^{WYL} = \frac{g_k^T (g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1})}{\|g_{k-1}\|^2}.
$$

The WYL method and PRP methods both come with restart properties. Zhang (2009) studied and improved WYL CG method and suggested the NPRP method, formulated as

$$
\beta_{k^+}^{NPRP} = \frac{\|g_{k+1}\|^2 - \|g_k\|^2}{\|g_{k+1}\|^2}.
$$

Zhang (2009) proved that the NPRP method satisfies descent condition under strong Wolfe line search. Later, Dai and Wen (2012) proposed a modified NPRP method as follows:

$$
\beta_{k^+}^{DPRP} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k+1}\|} |g_k^T g_{k+1}|}{\mu \|g_k\|^2 + \|g_{k-1}\|^2}, \quad \mu > 1
$$

Based on the above ideas, we present a new $\beta_k$ known as $\overline{\beta}_{k^+}^{YHM}$, where YHM denotes Yasir, Hamoda, and Mamat. The formula for $\overline{\beta}_{k^+}^{YHM}$ is defined by

$$
\overline{\beta}_{k^+}^{YHM} = \begin{cases} 
\frac{g_k^T (g_k - g_{k-1})}{\|g_k\|^2} & \text{if } 0 \leq g_k^T g_{k-1} \leq \|g_k\|^2 \\
\frac{g_k^T (g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1})}{\|g_{k-1}\|^2} & \text{otherwise}
\end{cases}
$$

The following algorithm is a general algorithm for solving optimization by CG methods.

**Algorithm 2.1:**

**Step1:** Given an initial point $x_0 \in \mathbb{R}^n$, $\epsilon > 0$, set $d_0 = -g_0$, $k = 0$

**Step2:** Compute $\beta_k$ by formula (9)

**Step3:** Compute $d_k$ based on (3). If $g_k = 0$, then stop.

**Step4:** Compute $\alpha_k$ by inexact line search.

**Step5:** Update new point based on (2)
Step 6: Convergence test and stopping criteria. If \( f(x_{k+1}) < f(x_k) \) and \( \|g_k\| \leq \epsilon \), then stop. Otherwise, set \( k = k + 1 \) and go to Step 1.

Global Convergence analysis
In this section, we study the global convergence properties of \( \beta_k^{YHM} \), starting with the sufficient descent condition. Firstly, we need to simplify \( \beta_k^{YHM} \) so that the proving steps will be easier. From (9), we know that:

\[
\beta_k^{YHM} = \begin{cases} 
\frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} & \text{if } 0 \leq g_k^T g_{k-1} \leq \|g_k\|^2 \\
\frac{g_k^T (g_k - \|g_k\| g_{k-1})}{\|g_{k-1}\|^2} & \text{otherwise}
\end{cases}
\]

If \( 0 \leq g_k^T g_{k-1} \leq \|g_k\|^2 \), then,

\[
\beta_k^{YHM} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} = \frac{\|g_k\|^2 - g_k^T g_{k-1}}{\|g_{k-1}\|^2} \geq 0.
\]

Otherwise,

\[
\beta_k^{YHM} = \frac{g_k^T (g_k - \|g_k\| g_{k-1})}{\|g_{k-1}\|^2} \geq \frac{\|g_k\|^2 - \|g_k\| \|g_{k-1}\| g_k^T g_{k-1}}{\|g_{k-1}\|^2}.
\]

By Cauchy - Schwarz inequality, it is implied that

\[
\beta_k^{YHM} \geq \frac{\|g_k\|^2 - \|g_k\| \|g_{k-1}\| g_k^T g_{k-1}}{\|g_{k-1}\|^2} = 0.
\]

Hence, we can deduce that for both cases of \( 0 \leq g_k^T g_{k-1} \leq \|g_k\|^2 \) and otherwise,

\[
\beta_k^{YHM} \geq 0.
\]

Sufficient descent condition
The sufficient descent condition is defined by:

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

The following theorem shows that YHM with inexact line search possesses the sufficient descent property.

**Theorem 1.** Suppose that the sequence \( \{g_k\} \) and \( \{d_k\} \) are generated by Algorithm (2.1) and the step-length \( \alpha_k \) is determined by strong Wolfe line search. If \( g_k \neq 0 \), then the sequence \( \{d_k\} \) satisfies the sufficient descent condition for all \( k \geq 0 \).

**Proof.** The proof of the descent property of \( \{d_k\} \) is by induction. Firstly, we prove the theorem for the case of \( k = 0 \).

**Case (1):** If \( 0 \leq g_k^T g_{k-1} \leq \|g_k\|^2 \), then

\[
\beta_k^{YHM} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}.
\]

Since \( g_0^T d_0 = -\|g_0\|^2 < 0 \), then condition (11) is fulfilled for \( k = 0 \).

Now suppose that \( d_i, i = 1, 2, 3, ..., k \) are all descent directions that \( s g_i^T d_i < 0 \). From the strong Wolfe condition, and (10)

\[
|\beta_{k+1}^{YHM} g_{k+1}^T d_k| \leq \sigma \frac{\|g_{k+1}\|^2}{\|g_k\|^2} |g_k^T d_k|.
\]

Now, we multiply \( d_{k+1} = -g_{k+1} + \beta_{k+1}^{YHM} g_{k+1}^T d_k \) with \( g_{k+1}^T \) to get

\[
g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_{k+1}^{YHM} g_{k+1}^T g_{k+1} d_k
\]
We divide both sides by \(\|\mathbf{g}_{k+1}\|^2\), which gives us
\[
\frac{\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}}{\|\mathbf{g}_{k+1}\|^2} = -1 + \beta_{k+1}^\text{YHM} \frac{\mathbf{g}_{k+1}^T \mathbf{d}_k}{\|\mathbf{g}_{k+1}\|^2}.
\] (13)
Since \(\mathbf{g}_{k+1}^T \mathbf{d}_{k+1} < 0\), then from (12), we have
\[
\left| \beta_{k+1}^\text{YHM} \frac{\mathbf{g}_{k+1}^T \mathbf{d}_k}{\|\mathbf{g}_{k+1}\|^2} \right| \leq \sigma \frac{\|\mathbf{g}_{k+1}\|^2}{\|\mathbf{g}_k\|^2} (-\mathbf{g}_k^T \mathbf{d}_k).
\]
Hence,
\[
\frac{\|\mathbf{g}_{k+1}\|^2}{\|\mathbf{g}_k\|^2} \sigma \mathbf{g}_k^T \mathbf{d}_k \leq \beta_{k+1}^\text{YHM} \frac{\mathbf{g}_{k+1}^T \mathbf{d}_k}{\|\mathbf{g}_{k+1}\|^2} \leq -\frac{\|\mathbf{g}_{k+1}\|^2}{\|\mathbf{g}_k\|^2} \sigma \mathbf{g}_k^T \mathbf{d}_k
\] (14)
Substitute (13) into (14), then
\[
-1 + \sigma \frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\|^2} \leq -1 - \frac{\|\mathbf{g}_{k+1}\|^2}{\|\mathbf{g}_k\|^2} \sigma \mathbf{g}_k^T \mathbf{d}_k
\]
By repeating this process and taking into account that \(\mathbf{g}_0^T \mathbf{d}_0 = -\|\mathbf{g}_0\|^2\), we get
\[
-\sum_{j=0}^{k} \sigma^j \leq \frac{\mathbf{g}_{k+1}^T \mathbf{d}_k}{\|\mathbf{g}_{k+1}\|^2} \leq -2 + \sum_{j=0}^{k} \sigma^j
\] (15)
Since \(\sum_{j=0}^{K} \sigma^j < K\), equation (15) can be written as
\[
-\frac{1}{1-\sigma} \leq \frac{\mathbf{g}_{k+1}^T \mathbf{d}_k}{\|\mathbf{g}_{k+1}\|^2} \leq -2 + \frac{1}{1-\sigma}
\] (16)
By making the restriction \(\sigma \in (0, \frac{1}{2})\), we can see that \(\mathbf{g}_{k+1}^T \mathbf{d}_{k+1} < 0\). Therefore, by induction, \(\mathbf{g}_{k+1}^T \mathbf{d}_k < 0\) holds for all \(k \geq 0\). Substitute \(c = 2 - \frac{1}{1-\sigma}\), \(0 < c < 1\) into (16) and we get \(c - 2\) \(\|\mathbf{g}_k\|^2 \leq \mathbf{g}_k^T \mathbf{d}_k \leq -c \|\mathbf{g}_k\|^2\).
This implies that condition (11) holds. The proof is completed.

**Case (2):** When \(\beta_{k+1}^\text{YHM} = \frac{\mathbf{g}_k^T (\mathbf{g}_k - \mathbf{g}_{k-1})/\|\mathbf{g}_{k-1}\|^2}{\|\mathbf{g}_{k-1}\|^2} = \frac{\mathbf{g}_k^T \mathbf{d}_{k-1}}{\|\mathbf{g}_{k-1}\|^2}\), the proof of this theorem can be seen in (Wei et al., 2006).

**Global Convergence Properties**
The following assumptions are often used in the studies of the CG method.

**Assumption 1**
A. \(f(x)\) is bounded from below on the level set \(\Omega = \{x \in \mathbb{R}^n, f(x) \leq f(x_0)\}\) where \(x_0\) is the starting point.
B. In some neighbourhood \(\Omega\) of \(\Omega\), the objective function is continuously differentiable and its gradient is Lipschitz continuous, that is, there exists constant \(L > 0\) such that:
\[
\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in \Omega
\]
In 1992, Gilbert and Nocedal introduced property (*) which plays an important role in the studies of CG method. This property means that the following search direction automatically approaches the steepest direction when a small step-length is generated, and the step-length are not produced successively (Zhang et al., 2012).

**Property (*)**
Consider a CG method of the form (2) and (3). Suppose that for all \(k \geq 1\),
\[
0 < \gamma \leq \|\mathbf{g}_k\| \leq \gamma^-
\] (18)
where \(\gamma\) and \(\gamma^-\) are two positive constants. The method has property (*) if there exist constants \(b > 1\) and \(\lambda > 0\) such that for all \(k\): \(|\beta_k| \leq b\), \(|s_k| \leq \lambda\) implies \(||\beta_k\| \leq \frac{1}{2b}\) where \(s_k = \alpha_k d_k\).
The following lemma shows that the new parameter \(\beta_{k+1}^\text{YHM}\) possesses property (*).
Lemma 1. Consider the method of form (2) and (3), and suppose that Assumption 1 holds, then the CG method with $\beta_{k}^{YHM}$ has property (*).

Proof: Case 1: If $0 \leq \frac{g_{k}^{T}g_{k-1}}{\|g_{k}\|^2} \leq \|g_{k}\|^2$, then $\beta_{k}^{YHM} = \frac{g_{k}^{T}(g_{k-1}-g_{k})}{\|g_{k-1}\|^2}$.

Set $b = \frac{y^2}{y^2} > 1$, $\lambda = \frac{y^2}{2y^2-b}$.

By (8) and (18) $|\beta_{k}^{YHM}| = \frac{|g_{k}^{T}(g_{k-1}-g_{k})|}{\|g_{k-1}\|^2} \leq \frac{\|g_{k}\|^2}{\|g_{k-1}\|^2} \leq \frac{y^2}{y^2} = b$.

By Assumption 1, if $\|s_{k}\| \leq \lambda$, then $|\beta_{k}^{YHM}| = \frac{\|g_{k}\||g_{k-1}\|}{\|g_{k-1}\|^2} \leq \frac{LY^{-2}}{\gamma^2} = \frac{1}{2b}$.

The proof is complete.

Case 2: When $\beta_{k}^{YHM} = \frac{g_{k}^{T}(g_{k-1}-g_{k})}{\|g_{k-1}\|^2}$, the proof of this theorem can be seen in (Wei et al., 2006).

Lemma 2. Suppose that Assumption 1 holds, and $x_{k}$ is generated by Algorithm 2.1 where $d_{k}$ satisfies $g_{k}^{T}d_{k} < 0$ for all $k$. The step size $\alpha_{k}$ is obtained by (SWP) line search (4) and (5), then,

$$\sum_{k=1}^{\infty} \frac{(g_{k}^{T}d_{k})^2}{\|d_{k}\|^2} < \infty$$

(19)

Proof. By Assumption 1 and the strong Wolfe line search, we obtain

$$1 - \sigma|g_{k}^{T}d_{k} \leq (g_{k+1} - g_{k})^{T}d_{k} \leq L\alpha_{k}\|d_{k}\|^2$$

Hence,

$$\alpha_{k} \geq \frac{-(1-\sigma)g_{k}^{T}d_{k}}{L\|d_{k}\|^2}$$

(20)

We combine (20) with (12), which then results to

$$\sum_{k=1}^{\infty} \frac{(g_{k}^{T}d_{k})^2}{\|d_{k}\|^2} < \frac{L}{1-\sigma} \sum_{k=1}^{\infty} (-\alpha_{k} g_{k}^{T}d_{k}) < \infty$$

The proof is complete.

Theorem 2. Consider any CG method of the form (2) and (3) that satisfies the following conditions:

1. $\beta_{k} \geq 0$
2. The search directions fulfil the sufficient descent condition.
3. The Zoutendijk condition holds.
4. Property(\text{\text{\textit{\text{*}}}}) holds.

If Assumptions 1 and 2 hold, then the iteration is globally convergent. From equations (11), (16), and (17) and Lemma 2, we found that the YHM method satisfies all four conditions in Theorem 2 under the strong Wolfe line search, so the method is globally convergent.

Numerical results and discussions

In this section, we present the results of the numerical tests conducted on our new parameter. The test problems used are taken from Andrei (2008), as shown in Table 1. We measure the performance of the proposed method by comparing it with other, well-established CG methods; FR, PRP, WYL and DPRP. A laptop with Intel(R) Core(TM) i5-M520 (2.40GHz) CPU processor and 4GB RAM in addition to MATLAB software version 8.3.0.532 (R2014a) are used to execute the optimization algorithms. We consider $\|g_{k}\| \leq \varepsilon$ as the stopping criteria as suggested by Hillstrom (1977) with $\varepsilon = 10^{-6}$. The dimensions of the test problems lay in the range of 2 to 10000. For
each test function, we use four initial points, starting from a point close to the solution to another point far from it. In some cases, the computation is stopped due to the line search failing to find a positive step-size, thus it is considered a failure. The performance results are shown in Figures 1 and 2, respectively, based on the performance profile introduced by Dolan and More (2002).

Table 1: A list of problem functions

<table>
<thead>
<tr>
<th>No</th>
<th>Function</th>
<th>Dimension</th>
<th>Initial points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Six hump camel</td>
<td>2</td>
<td>-10, -8, 8, 10</td>
</tr>
<tr>
<td>2</td>
<td>Booth</td>
<td>2</td>
<td>10, 25, 50, 100</td>
</tr>
<tr>
<td>3</td>
<td>Treccani</td>
<td>2</td>
<td>5, 10, 20, 50</td>
</tr>
<tr>
<td>4</td>
<td>Zettl</td>
<td>2</td>
<td>5, 10, 20, 30</td>
</tr>
<tr>
<td>5</td>
<td>Ex–rosenbrock</td>
<td>2, 4, 10, 100, 500, 1000, 10000</td>
<td>13, 25, 30, 50</td>
</tr>
<tr>
<td>6</td>
<td>Extended penalty</td>
<td>2, 4, 10, 100</td>
<td>50, 60, 70, 80</td>
</tr>
<tr>
<td>7</td>
<td>Generalized Tridiagonal 1</td>
<td>2, 4, 10, 100</td>
<td>30, 35, 40, 45</td>
</tr>
<tr>
<td>8</td>
<td>Shallow</td>
<td>2, 4, 10, 100, 500, 1000, 10000</td>
<td>10, 25, 50, 70</td>
</tr>
<tr>
<td>9</td>
<td>Ex-Tridiagonal 1</td>
<td>2, 4, 10, 100, 500, 1000, 10000</td>
<td>12, 17, 20, 30</td>
</tr>
<tr>
<td>10</td>
<td>Extended White and Holst</td>
<td>2, 4, 10, 100, 500, 1000, 10000</td>
<td>3, 10, 30, 50</td>
</tr>
<tr>
<td>11</td>
<td>Quadratic qf2</td>
<td>2, 4, 10, 100, 500, 1000, 10000</td>
<td>10, 30, 50, 100</td>
</tr>
<tr>
<td>12</td>
<td>Extended Denschnb</td>
<td>2, 4, 10, 100, 500, 1000, 10000</td>
<td>8, 13, 30, 50</td>
</tr>
<tr>
<td>13</td>
<td>Hager</td>
<td>2, 4, 10, 100</td>
<td>1, 3, 5, 7</td>
</tr>
<tr>
<td>14</td>
<td>Ex-Powell</td>
<td>4, 8, 20, 100, 500, 1000</td>
<td>-1, 1, 7, 11</td>
</tr>
<tr>
<td>15</td>
<td>Extended Beale</td>
<td>2, 4, 10, 100, 500, 1000, 10000</td>
<td>-19, 1, 13, 23</td>
</tr>
<tr>
<td>16</td>
<td>Ex–Himmelblau</td>
<td>100, 500, 1000, 10000</td>
<td>50, 70, 100, 125</td>
</tr>
<tr>
<td>17</td>
<td>Diagonal 2</td>
<td>2, 4, 10, 100, 500, 1000, 10000</td>
<td>-1, 1, 2, 3</td>
</tr>
<tr>
<td>18</td>
<td>Perturbed quadratic</td>
<td>2, 4, 10, 100, 500, 1000, 10000</td>
<td>1, 5, 10, 15</td>
</tr>
<tr>
<td>19</td>
<td>Sum Squares function</td>
<td>2, 4, 10, 100, 500, 1000, 10000</td>
<td>1, 10, 20, 30</td>
</tr>
<tr>
<td>20</td>
<td>Ex- quadratic penalty QP2</td>
<td>4, 10, 100, 500, 1000, 10000</td>
<td>17, 18, 19, 20</td>
</tr>
<tr>
<td>21</td>
<td>Raydan1 function</td>
<td>2, 4, 10, 100</td>
<td>1, 3, 5, 7</td>
</tr>
<tr>
<td>22</td>
<td>Generalized Tridiagonal 2</td>
<td>2, 4, 10, 100</td>
<td>1, 7, 8, 14</td>
</tr>
<tr>
<td>23</td>
<td>Quadratic QF1</td>
<td>2, 4, 10, 100, 500, 1000, 10000</td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td>24</td>
<td>Dixon and Price</td>
<td>2, 4, 10, 100</td>
<td>100, 125, 150, 175</td>
</tr>
<tr>
<td>25</td>
<td>Fletcher</td>
<td>4, 10, 100, 500, 1000</td>
<td>7, 11, 13, 15</td>
</tr>
<tr>
<td>26</td>
<td>Ex-Maratos</td>
<td>2, 4, 10, 100</td>
<td>5, 10, 12, 15</td>
</tr>
<tr>
<td>27</td>
<td>Leon function</td>
<td>2</td>
<td>2, 5, 8, 10</td>
</tr>
</tbody>
</table>
Table 1. Performance profile relative to the number of iteration.

<table>
<thead>
<tr>
<th></th>
<th>Function</th>
<th>n</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>Extended wood</td>
<td>4</td>
<td>3, 5, 20, 30</td>
</tr>
<tr>
<td>29</td>
<td>Quartic function</td>
<td>4</td>
<td>5, 10, 15, 20</td>
</tr>
<tr>
<td>30</td>
<td>Matyas function</td>
<td>2</td>
<td>5, 10, 15, 20</td>
</tr>
<tr>
<td>31</td>
<td>Colville function</td>
<td>4</td>
<td>2, 4, 7, 10</td>
</tr>
</tbody>
</table>

**Figure 1.** Performance profile relative to the number of iteration.
From Figures 1 and 2, we found that our proposed algorithm solves 100% of the test problems, followed by WYL which solves 99.4% and DPRP with 88.4% of problems solved. Older CG methods like FR and PRP solve about 57% and 49% of the test functions, respectively.

**Conclusion**

This paper gives a new $\beta_k$ formula for solving unconstrained optimization problems. Under strong Wolfe line search, this new $\beta_k$ possesses global convergence properties. Numerical results show that the YHM method is very efficient and has the best performance when compared with other tested CG methods.

**References**


