A new modification nonlinear conjugate gradient method with strong wolf-powell line search

Chergui Ahmed¹, Bouali Taher²

¹Mathematics and computer science Department, Laboratory of Mathematics, informatics and systems (LAMIS), Larbi Tebessi University, Algeria
²Department of Mathematics, Khenchela University, Algeria

ABSTRACT

The conjugate gradient method has played a special role in solving large-scale unconstrained optimization problems. In this paper, we propose a new family of CG coefficients that possess sufficient descent conditions and global convergence properties this CG method is similar to (Wei et al) [10]. Global convergence result is established under Strong Wolf-Powell line search. Numerical results to find the optimum solution of some test functions show the new proposed formula has the best result in CPU time and the number of iterations, and the number of gradient evaluations when it comparing with FR, PRP, DY, and WYL.

Keywords: Global convergence, Strong wolf-powell line search, Unconstrained optimization, conjugate gradient

1. INTRODUCTION

The optimization problem finds application in several fields, such as classical continuum physics, theoretical, mathematical & computational physics, particle and nuclear physics, physical chemistry, pure mathematics, mathematical physics, fluid dynamics, actuarial science, applied information economics, astrophysics, biostatistics, business statistics, traffic routing in telecommunication systems (24.), cyber-physical security (25.), intelligent transportation systems (26.) and smart grids (27.). Consider the unconstrained optimization problem.

\[(p); \min \{ f(\chi); x \in R^n \} \]  

where \( f; R^n \rightarrow R \) continuously differentiable. The nonlinear conjugate gradient (CG) method usually takes the following iterative formula

\[ x_{k+1} = x_k + \alpha_k d_k, \quad \alpha_k > 0 \quad k = 0; 1; 2; 3 \]  

For solving (1), where \( x_k \) is the current iterate point, \( k > 0 \) is a step length, and \( d_k \) is a search direction defined by,

\[ d_k = \begin{cases} -\nabla f(x_k) & \text{if } k = 1 \\ -\nabla f(x_k) + \beta_k d_{k-1} & \text{if } k \geq 2 \end{cases} \]
Where $g_k$ is the gradient of $f(x)$, $\beta_k \in R$ is a scalar which determines the different conjugate gradient methods. (2) and (3). Well-known formulas for $B_k$ are called Conjugate-Descent (CD) (Fletcher 1997) [1], Fletcher-Reeves (FR) (Fletcher and Reeves 1964) [2], Hestenes.Stiefel (HS) (Hestenes and Stiefel 1952) [3], Liu.Storrey (LS) (Liu and Storey 1992) [4], and Polyak. Ribiére.Polyak (PRP) (Polyak and Ribiére 1969; Polyak 1969) [5], and some modified formulas (Dai 2002; Dai 2006b) [6]. The convergence behavior of the different conjugate gradient methods with some line search conditions (Armijo 1966, Al-baali 1985, Dai 2001, Dai et al. 1999, Dai and Yuan 1996, Grippo et al. 1986, Grippo and Lucid 1997, Liu and Han 1995). The well-known formulas for $\beta_k$ are,

\[ B_{k}^{CD} = \frac{\|\nabla f(x_k)\|^2}{\|\nabla f(x_{k-1})\|^2} \]  
\[ B_{k}^{FR} = \frac{\nabla f(x_k)^T (\nabla f(x_k) - \nabla f(x_{k-1}))}{\|\nabla f(x_k) - \nabla f(x_{k-1})\|^2} \]  
\[ B_{k}^{HS} = \frac{\nabla f(x_k)^T (\nabla f(x_k) - \nabla f(x_{k-1}))}{\|\nabla f(x_k) - \nabla f(x_{k-1})\|^2} \]  
\[ B_{k}^{PRP} = \frac{\nabla f(x_k)^T (\nabla f(x_k) - \nabla f(x_{k-1}))}{\|\nabla f(x_k) - \nabla f(x_{k-1})\|^2} \]  
\[ B_{k}^{DY} = \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T \nabla f(x_{k-1}) d_{k-1}} \]  
\[ B_{k}^{LS} = \frac{\nabla f(x_k)^T (\nabla f(x_k) - \nabla f(x_{k-1}))}{\|\nabla f(x_k) - \nabla f(x_{k-1})\|^2} \]  
\[ B_{k}^{WYL} = \frac{\nabla f(x_k)^T (\nabla f(x_k) - \nabla f(x_{k-1}))}{\|\nabla f(x_k) - \nabla f(x_{k-1})\|^2} \]  

In order to find the step length ($a_k$), we use Strong Wolf Powell (SWP) line search.

\[ f(x_k + \alpha_k d_k) - f(x_k) \leq \delta a_k \nabla f(x_k)^T d_k \]  
\[ |\nabla f(x_k + \alpha_k d_k)^T d_k| \leq -\sigma \nabla f(x_k)^T d_k \]  

Where \((0 < \delta < \frac{1}{3}) \) and \((0 < \sigma < 1)\)

From a bibliography point of view, the Nonlinear Conjugate Gradient methods can be improved by using novel techniques proposed in (20.), (21.), (22.), (23) R^T \nabla f(x_k) d_k \]

In this paper, we will present the new formula in section 2. In addition, the sufficient descent condition and the global convergence of the new method under the inexact line search (11) and (12), in the following theorem. Will be presented in section 3. Finally, we will discuss the numerical results and conclusion in sections 4 and five respectively.

2. THE NEW FORMULA

In this section, we propose the new $B_k$ which is extention of the $B_k^{WYL}$ [10] that we named $B_k^{wyjIM}$,

\[ B_k^{wyjIM} = \frac{\nabla f(x_k)^T \nabla f(x_k) - \nabla f(x_{k-1})^T \varphi_k d_{k-1}}{\|\nabla f(x_{k-1})\|^2} \]  

Where $\varphi_k = \frac{\|\nabla f(x_k)\|}{\|\nabla f(x_{k-1})\|}$ and $\|.\|$ means the Euclidean norm

3. CONVERGENT ANALYSIS OF WYLM METHOD

In this section, we will show the convergent properties of $B_k^{wyjIM}$ using inexact line searches.
3.1. Convergent Analysis Based on Inexact Line Search

In this section, we will show the convergent analysis based on the inexact line search by means of strong Wolfe line search. We will also show that these CG coefficients will possess sufficient descent conditions and global convergence properties. Under this inexact line search (11) and (12). In the following theorem; we discuss the sufficient condition,

\[ \nabla f(x_k)^T d_k \leq -C \|\nabla f(x_k)\|^2, \quad C > 0 \quad (14) \]

Where \( k > 0 \) and \( C \in (0, 1) \) under SWP line search.

3.2.1. Sufficient Descent Condition

For the sufficient descent condition, we present the following Theorem.

Theorem 1: If the sequences \( g_k \) and \( d_k \) are generated by the methods (2), (3) and (13) with step length \( \alpha_k \) determined by (10) and (11) if \( \sigma \in (0, \frac{1}{2}) \), then the sufficient descent condition holds.

Proof: We use proof by induction from (3). We know that for \( k = 0 \) it is hold. Suppose that it is:

\[ \nabla f(x_k)^T d_k = -\|\nabla f(x_k)\|^2 + B_k^{WYLM} \nabla f(x_k)^T d_{k-1} \]

Divide \( \|\nabla f(x_k)\|^2 \) indicated that;

\[
\frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|^2} = -1 + B_k^{WYLM} \frac{\nabla f(x_k)^T d_{k-1}}{\|\nabla f(x_k)\|^2} \\
= -1 + \frac{-\nabla f(x_{k-1})^T d_{k-1}}{\nabla f(x_k)^T d_{k-1}} \left( 1 - \frac{\nabla f(x_k)^T d_{k-1}}{\|\nabla f(x_k)\|\|\nabla f(x_{k-1})\|} \right) \quad (15)
\]

Using (12), we have,

\[
\frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|^2} \leq -1 + \frac{-\nabla f(x_{k-1})^T d_{k-1}}{\|\nabla f(x_k)\|^2} \left( 1 - \frac{\nabla f(x_k)^T d_{k-1}}{\|\nabla f(x_k)\|\|\nabla f(x_{k-1})\|} \right) \quad (16)
\]

\[
\frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|^2} \geq -1 + \frac{\nabla f(x_{k-1})^T d_{k-1}}{\|\nabla f(x_k)\|^2} \left( 1 - \frac{\nabla f(x_k)^T d_{k-1}}{\|\nabla f(x_k)\|\|\nabla f(x_{k-1})\|} \right) \quad (17)
\]

And applying the Cauchy-Schwartz we get,

\[ 0 \leq \frac{\nabla f(x_k)^T d_{k-1}}{\|\nabla f(x_k)\|\|\nabla f(x_{k-1})\|} \leq 2 \quad (18) \]

This implies that,

\[ -1 + 2 \frac{\nabla f(x_{k-1})^T d_{k-1}}{\|\nabla f(x_k)\|^2} \leq \frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|^2} \leq -1 - 2 \frac{\nabla f(x_{k-1})^T d_{k-1}}{\|\nabla f(x_k)\|^2} \quad (19) \]

By repeating this process and the fact \( \nabla f(x_1)^T d_1 = -\|\nabla f(x_1)\|^2 \) we have,

\[ -\sum_{i=0}^{k-1} (2\sigma)^i \leq \frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|^2} \leq -2 + \sum_{i=0}^{k-1} (2\sigma)^i \quad (20) \]

Since \( \sum_{i=0}^{k-1} (2\sigma)^i < \sum_{i=0}^{\infty} (2\sigma)^i = \frac{1}{1-2\sigma} \)

As shown in (19) Can be written as:

\[ -\frac{1}{1-2\sigma} \leq \frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|^2} \leq -2 + \frac{1}{1-2\sigma} \quad (21) \]

By making the restriction \( \sigma \in (0, \frac{1}{4}) \) we have \( \nabla f(x_k)^T d_k < 0 \) so by induction, \( \forall k \in N, \nabla f(x_k)^T d_k < 0 \) holds.

Now, we prove the sufficient descent property of \( d_k \) if \( \sigma \in (0, \frac{1}{4}) \)
Set \( \lambda = 2 - \frac{1}{1-\sigma} \) then \( 0 < \lambda < 1 \), and (18) turns out to be

\[
\lambda - 2 \leq \frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|^2} \leq -\lambda
\]  

(22)

Thus we obtain \( \nabla f(x_k)^T d_k \leq -\gamma \|\nabla f(x_k)\|^2 \) or \( \gamma = -2 + \frac{1}{1-\sigma} \) where \( \lambda \in (0, 1) \)

The proof is completed.

3.2.2. Global Convergent Analysis

The following assumption is needed in order to proceed with the proof of global convergence properties.

**Assumption 1**

(i) The function \( f \) is bounded below on the level set \( R^n \) and is continuous and differentiable in neighbourhood \( N \) of the level set \( \Omega = \{ x \in R^n; f(x) < f(x_0) \} \) at the initial point \( x_0 \)

(ii) The gradient \( g(x) \) is Lipschitz continuous in \( N \), so a constant \( L \geq 0 \) exists, such that that

\[
\|g(x) - g(y)\| \leq L\|x - y\| : \text{For all } x, y \in N
\]

**Theorem 2**: Suppose Assumption 1 is true, consider any CG method of form (2) and (3), where \( \alpha_k \) satisfied. SWP line search and the sufficient descent condition holds, then \( \lim_{k \to \infty} \|\nabla f(x_k)\| = 0 \).

**Proof**

Subtracting \( \nabla f(x_k)^T d_k \) from both sides of (12) and using the Lipschitz condition we have

\[
-(1-\sigma)\nabla f(x_k)^T d_k \leq (\nabla f(x_k) - \nabla f(x_{k-1}))^T d_k \leq L \alpha_k \|d_k\|^2
\]

(23)

Therefore,

\[
\frac{(1-\sigma)\nabla f(x_k)^T d_k}{\alpha_k} \leq L \|d_k\|^2
\]

(24)

With (11) we obtain:

\[
f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k \nabla f(x_k)^T d_k \geq -\delta \frac{(1-\sigma)\|\nabla f(x_k)^T d_k\|^2}{\|d_k\|^2}
\]

(25)

Moreover, from the hypothesis (1), we have that \( \{ f(x_k) \} \) is a decreasing sequence and has a limit in, which shows that \( \lim_{k \to \infty} f(x_k) = f(x_{k+1}) < +\infty \) and after (25) we have,

\[
+\infty > f(x_1) - \lim_{k \to \infty} f(x_{k+1}) = \sum |f(x_k) - f(x_{k+1})| \geq \delta \frac{(1-\sigma)\|\nabla f(x_k)^T d_k\|^2}{\|d_k\|^2}
\]

(26)

Then \( \sum \frac{(\nabla f(x_k)^T d_k)^2}{\|d_k\|^2} \leq +\infty \)

Hence, \( \lim_{k \to \infty} \|\nabla f(x_k)\| = 0 \) Then proof is completed.

4. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we present the results of our proposed method WYLM on comparison with CG methods FR, DY, WYL, and PRP. We will use some of test problems considered in Andrei [18] to analyze the efficient of \( B_k^{WYLM} \). We considered \( \epsilon = 10^{-6} \); \( \sigma = 0.1 \); \( \delta = 0.01 \) and the gradient value as stopping criteria. The tolerance \( \epsilon = 10^{-6} \) is selected for all algorithms to investigate the rapidity of iteration of these algorithms towards the optimal solution. We used Matlab R2010 the performance results are shown in Figures 1-6. We use the following algorithm,

**Algorithm 1: \( B_k^{WYLM} \) method**
Step 1: Choose an initial point \( x_0 \in R^n \) set \( k = 1 \), \( \epsilon = 10^{-6} \) set \( d_0 = -g_0 = -\nabla f(x_0) \)
Step 2: Compute \( B_k \) based on (13), (15), (7), (8), or (10)
Step 3: Compute $d_k$ based on (3): if $\| \nabla f(x_k) \| = 0$, then stop.
Step 4: Compute step length $\alpha_k$ by one line search technique, let $x_{k+1} = x_k + \alpha_k \, d_k$
Step 5: Updating new point based on (2).
Step 6: Convergent test and stopping criteria, if $(x_k) < f(x_{k-1})$ and $\| \nabla f(x_k) \| < \epsilon$, then stop, otherwise go to Step 1 with $k = k + 1$.

Table 1. Test Problems used with the Strong Wolf Condition $\varepsilon = 10^{-6}; \sigma = 0.1; \delta = 0.01$

<table>
<thead>
<tr>
<th>#</th>
<th>Functions</th>
<th>N</th>
<th>Initial points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Booth</td>
<td>2</td>
<td>(10, 10), (14,14), (40,40)</td>
</tr>
<tr>
<td>2</td>
<td>Rosenbrock</td>
<td>2</td>
<td>(5.5), (25,25), (12,12)</td>
</tr>
<tr>
<td>3</td>
<td>Goldstein-price</td>
<td>2</td>
<td>(2,2), (-11,-11), (-13,-13)</td>
</tr>
<tr>
<td>4</td>
<td>Extended Powell</td>
<td>4</td>
<td>(7,7,7,7), (15,15,15,15)</td>
</tr>
<tr>
<td>5</td>
<td>Tridiagonal 1</td>
<td>2</td>
<td>(2,2), (17,17), (100,100)</td>
</tr>
<tr>
<td>6</td>
<td>Strait</td>
<td>2</td>
<td>(100,100), (50,50)</td>
</tr>
<tr>
<td>7</td>
<td>Generalized Quartic</td>
<td>2</td>
<td>(10,10), (200,200), (17,17)</td>
</tr>
</tbody>
</table>

Figures 1-3 list the performance of the above methods relative to iterations number, the number of gradient evaluations and the CPU time, respectively.

Figure 1. Performance based on the number of iterations

Figure 2. Performance based on the number of gradient evaluations

Figure 3. Performance based CPU time

Remark 1: Figures 1-3, shows that “WYLM” method has best performance since it solves about 99% of the test problems successfully.

Example: Extended Rosenbrock function,

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2, \ x optimal = (1, 1).$$
\[ \nabla f(x, y) = \left( -2 + 2x - 400xy + 400x^3 \right) \]

Table 2. Extended Rosenbrock function \( f(x, y) = (1-x)^2 + 100(y-x^2)^2 \). \( x \) optimal = (1, 1) Numerical results for \( B_{\text{WYLM}} \), WY L, DY, FR and PRP in terms of number iterations (NI) and CPU time.

<table>
<thead>
<tr>
<th>Initial point</th>
<th>WYLM</th>
<th>WYL</th>
<th>PRP</th>
<th>FR</th>
<th>DY</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1000, 1000)</td>
<td>2900/8948</td>
<td>Failed</td>
<td>8377/42.896843</td>
<td>Failed</td>
<td>Failed</td>
</tr>
<tr>
<td>(30, 30)</td>
<td>119/0.29190</td>
<td>1511/2.613863</td>
<td>8377/42.896843</td>
<td>3944/16.488</td>
<td>Failed</td>
</tr>
<tr>
<td>(-1, 1000)</td>
<td>316/1.2419</td>
<td>2324/7.481274</td>
<td>Failed</td>
<td>Failed</td>
<td>Failed</td>
</tr>
<tr>
<td>(-1, 1.3)</td>
<td>88/1.993664</td>
<td>14532/75.441580</td>
<td>466/2.602825</td>
<td>420/2.6787</td>
<td>170/0.21942</td>
</tr>
<tr>
<td>(0.1, 1)</td>
<td>1240/425383</td>
<td>20000/118.46089</td>
<td>510/3.010145</td>
<td>88.2/21504</td>
<td>100/0.376978</td>
</tr>
<tr>
<td>(1, 4)</td>
<td>224/0.809022</td>
<td>2000/113.847652</td>
<td>456/2.338696</td>
<td>3944/16.488</td>
<td>100/0.376978</td>
</tr>
<tr>
<td>(100, 100)</td>
<td>210/0.607405</td>
<td>Failed</td>
<td>Failed</td>
<td>5840/25.106</td>
<td>Failed</td>
</tr>
<tr>
<td>(-1, 7)</td>
<td>201/0.754304</td>
<td>20000/113.847652</td>
<td>527/6.118101</td>
<td>130/2.086794</td>
<td>219/0.442945</td>
</tr>
</tbody>
</table>

**Remark 2:** In table 2, the WYLM method was successful in all attempts to achieve the optimal solution, while the other methods failed.

Table 3. Summary of Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Ranking</th>
<th>The success rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>WYLM</td>
<td>1</td>
<td>100%</td>
</tr>
<tr>
<td>PRP</td>
<td>2</td>
<td>75%</td>
</tr>
<tr>
<td>FR</td>
<td>3</td>
<td>75%</td>
</tr>
<tr>
<td>WYL</td>
<td>4</td>
<td>75%</td>
</tr>
<tr>
<td>DY</td>
<td>5</td>
<td>50%</td>
</tr>
</tbody>
</table>

**Remark 3:** Table 3, shows that “WYLM” has best performance since it solves about 100% of the test problems successfully.

Table 4. Rosenbrock Function, Initial Point [-3 100]

<table>
<thead>
<tr>
<th></th>
<th>WYLM</th>
<th>WYL</th>
<th>DY</th>
<th>PRP</th>
<th>FR</th>
</tr>
</thead>
<tbody>
<tr>
<td>NI</td>
<td>309</td>
<td>Failed</td>
<td>3505</td>
<td>Failed</td>
<td>2456</td>
</tr>
<tr>
<td>CPU(s)</td>
<td>1.0048</td>
<td>Failed</td>
<td>13.2389</td>
<td>Failed</td>
<td>85297</td>
</tr>
<tr>
<td>X optimal</td>
<td>(1, 1)</td>
<td>(1, 1)</td>
<td>(1, 1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figures 4 and 5, list the comparison of WYLM method and DY, WYL, PRP, FR methods \( x_0 = [-3 100] \).

![Figure 4](image1.png) ![Figure 5](image2.png)

Figure 4. Performance profiles based on the number of function evaluations

Figure 5. Performance profiles based on the number of gradient evaluations
Remark 4: From the Figures 4-5. The WYLM method performs better than other methods by selecting a starting point with the Resenbrook function $f(x,y) = (1-x)^2 + 100(y-x^2)^2$

And she is best performance in terms of values gradients and functions and the number of iterations.

5. CONCLUSION

We present a new CG formula $B_{WYLM}^k$ that similar to WYL method and the global convergence properties are presented with several line searches. Results showed that the new coefficient satisfies the sufficient descent conditions and converge globally under strong Wolfe-Powell line search. It is proved that the algorithm is practical and effective to be used. Our numerical results have shown that our new method has the best performance compared to the FR, PRP, DY and WYL Methods

ACKNOWLEDGEMENTS

We are grateful to anonymous referees for their useful suggestions and comment on this paper.

REFERENCES


A new modification nonlinear conjugate gradient method with strong wolf-powell... (Chergui Ahmed)


**BIOGRAPHIES OF AUTHORS**

Ahmed Chergui Professor of Mathematics, Department of Mathematics and Computer Science, Larbi Tebessi University-Tebessa, Algeria. Email: ahmed.chergui@univ-tebessa.dz, University of Khensela Department of Mathematics and Computer Science, Algeria Email: Ahmedc.48@gmail.com

Taher Bouali Dr Mathematics, Department of Mathematics and Computer Science, Larbi Tebessi University - Tebessa, Algeria, Email: tahar.bouali@univ-tebessa.dz