A New Scale in Conjugate Gradient Methods for Solving Unconstrained Optimization

Eman T. Hamed¹, Huda I. Ahmed², Eilaf S. Khaleel³

¹,²Department of Operation Research and Intelligent Techniques, College of Computer Sciences and Mathematics
³Department of mathematic, College of Computer Sciences and Mathematics. University of Mosul, Mosul, Iraq

ABSTRACT

Conjugate gradient methods are an important class methods for unconstrained optimization, especially for large-scale problems. The descent property for the suggested method is proved under some conditions. Also, we prove that for strongly convex functions the modified method is global convergent. Finally, the numerical results showed that the new method is very efficient for general problems.

1. INTRODUCTION

Conjugate gradient methods are a class of important methods for solving:

\[ \min \{f(x): x \in \mathbb{R}^n\} \] ⁷

where \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) is a continuously differentiable. This method generates a sequence \( x_i \) of the function \( f \). In general the method has the following form:

\[
\begin{align*}
  x_{i+1} &= x_i + \alpha_i d_i, & i = 0,1,2, \ldots, \\
  d_{i+1} &= -g_{i+1} + \beta_i d_i, & \text{if } i \geq 1
\end{align*}
\]

where \( g_i = \nabla f(x_i), \alpha_i \) is a step length obtained by a line search, and \( \beta_i \) is a scalar parameter. There are many ways to select \( \beta_i \), and some well-known formulas are given in [⁷, ⁸, ⁹, ¹³].

The step length \( \alpha_i \) in (2) is computed by carrying out line search. The Wolfe line search consists of finding a positive step-length \( \alpha_i \) such that:

\[ f(x_i + \alpha_i d_i) - f(x_i) \leq \delta \alpha_i g_i^T d_i \] ⁴

\[ g(x_i + \alpha_i d_i)^T d_i \leq -\sigma g_i^T d_i \] ⁵

where \( 0 < \delta < \sigma < 1 \) [¹⁵, ¹⁶].

2. A NEW SCALLED CONJUGATE GRADIENT METHOD \( \theta_{i+1} \)

The algorithm in this class of nonlinear conjugate gradient algorithms generates the sequence \( x_i \) in (2) and defined the search direction \( d_{i+1} \) by:

\[ d_{i+1} = -\theta_{i+1} g_{i+1} + \beta_i d_i \] ⁶

where \( \theta_{i+1} \) is a positive scalar, observe that if \( \theta_i = 1 \), then we get the classical conjugate gradient algorithm according to the value of formulas of \( \beta_i \).

Eilaf [⁶] proposed a parameter \( \beta_i \), which is defined by:

\[ \beta_i = \frac{\gamma_i + \gamma_{i+1}^T}{\gamma_i^T y_i} - \frac{\gamma_i^T d_i}{\gamma_i^T g_i} \] ⁷

HOW TO CITE THIS ARTICLE

where \( s_i = x_{i+1} - x_i \), \( y_i = g_{i+1} - g_i \). We used this \( \beta_i \) in (6) and in order to get the new scalar \( \theta_{i+1} \) in our method, multiply both sides of (6) by \( y_i \).

\[
d_i_{i+1}^T y_i = -\theta_{i+1} g_{i+1}^T y_i + \beta_i d_i^T y_i
\]

(8)

since \( d_i_{i+1}^T y_i = -s_i^T g_{i+1} \)

which is called Perry’s condition \(^{[10]}\). Substituting (7) and (9) in (8):

\[
-s_i^T g_{i+1} = -\theta_{i+1} g_{i+1}^T y_i + \left( g_{i+1}^T y_i - \frac{g_i^T y_i}{d_i^T g_i} - \frac{g_{i+1}^T d_i}{d_i^T g_i} \right) d_i^T y_i
\]

\[
= -\theta_{i+1} g_{i+1}^T y_i + g_{i+1}^T y_i - \frac{g_{i+1}^T d_i}{d_i^T g_i} d_i^T y_i
\]

\[
\theta_{i+1} = 1 - \frac{g_{i+1}^T d_i}{d_i^T g_i}
\]

(10)

then the new direction is defined by:

\[
d_{i+1} = -\theta_{i+1} g_{i+1} + \beta_i d_i
\]

\[
= -\left(1 - \frac{g_{i+1}^T d_i}{d_i^T g_i}\right) g_{i+1} + \left( \frac{g_{i+1}^T y_i}{d_i^T g_i} \right) g_{i+1} - \frac{g_{i+1}^T d_i}{d_i^T g_i} d_i
\]

(11)

If we used exact line search (ELS) then the new scalar \( \theta_{i+1} \) is equal to 1.

**Algorithm 2.1**

**Step 1:** given \( x_1 \in \mathbb{R}^n \), \( \epsilon > 0 \) set \( k=1 \)

**Step 2:** set \( d_1 = -g_1 \), if \( \|g_1\| < \epsilon \) then stop

**Step 3:** find \( \alpha_k > 0 \) satisfying (4) \(^{[4]}\), (5)

**Step 4:** let \( x_{i+1} = x_i + \alpha_k d_i \), if \( \|g_k\| < \epsilon \) then stop, otherwise continue

**Step 5:** compute \( \beta_i \), \( \theta_{i+1} \) by the formula (7), (10) respectively and generate the new search direction \( d_{i+1} \) by (11)

**Step 6:** if \( k=n \) or \( |g_{k+1}^T g_k| \geq 0.2 \|g_{k+1}\|^2 \) (powell restarting \([13]\)) is satisfy, go to step (2) else set \( k=k+1 \) and go to step 3

### 3- CONVERGENCE ANALYSIS

Assume the following basic on the objective function.

**Assumption 3.1** \(^{[17]}\)

i) The function \( f(x) \) is bounded in the level set \( S \) i.e. there exists positive constant \( z > 0 \), such that

\[
\|x\| \leq z, \forall x \in S
\]

(12)

ii) If there is a neighborhood \( W \) of \( S \), and it is gradient is Lipschitz continuous, i.e. there exists a constant \( L > 0 \) such that

\[
\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|, \forall x, y \in W
\]

(13)

Under these assumptions on \( f \), there exists a constant \( \Gamma \geq 0 \) such that

\[
\|\nabla f(x)\| \leq \Gamma, \forall x \in S
\]

(14)

**Theorem 3.2** \(^{[4 and 17]}\)

Suppose that the function \( f(x) \) is a uniformly convex then there exists a constant \( M > 0 \) such that

\[
(g(x) - g(y))^T (x - y) \geq M \|x - y\|^2, \forall x, y \in S
\]

(15)

we can rewrite (15) in the following manner

\[
y_i^T s_i \geq M \|s_i\|^2
\]

(16)

Then (16) with (13), implies that:

\[
M \|s_i\|^2 \leq y_i^T s_i \leq L \|s_i\|^2
\]

(17)

**Theorem 3.3**

Suppose that \( d_i \) satisfies the Wolfe condition (4) and (5), then the direction \( d_{i+1} \) given by (11) is a sufficient descent direction.

Proof:

since \( d_0 = -g_0 \), we have \( g_0^T d_0 = -\|g_0\|^2 < 0 \)

now, multiplying (11) by \( g_{i+1} \) we have:
\[ d_{i+1}^T g_{i+1} = - \left( 1 - \frac{g_{i+1}^T d_i}{d_i^T g_i} \right) \| g_{i+1} \|^2 + \left( \frac{g_{i+1}^T y_i - s_i^T g_{i+1}}{d_i^T y_i} \right) d_i^T g_{i+1} \]

Divided both said by \( \| g_{i+1} \|^2 \)

\[ \frac{d_{i+1}^T g_{i+1}}{\| g_{i+1} \|^2} = 1 - \left( 1 - \frac{g_{i+1}^T d_i}{d_i^T g_i} \right) \| g_{i+1} \|^2 \]

since \( g_{i+1}^T d_i = d_i^T y_i + g_{i+1}^T d_i \)

\[ \therefore g_{i+1}^T d_i < d_i^T y_i \text{ also } s_i^T g_{i+1} \leq y_i \text{ and from (5) we have } \]

\[ \frac{d_{i+1}^T g_{i+1}}{\| g_{i+1} \|^2} + 1 \leq -\sigma g_{i+1}^T d_i \frac{d_i^T y_i}{g_i^T d_i} + \left( \frac{g_{i+1}^T y_i - s_i^T g_{i+1}}{d_i^T y_i} \right) d_i^T y_i \]

\[ \leq -\sigma d_i^T y_i + \frac{g_{i+1}^T y_i}{\| g_{i+1} \|^2} \frac{s_i^T y_i}{\| g_{i+1} \|^2} + \sigma d_i^T y_i \]

\[ \text{since } g_{i+1}^T y_i \leq \| g_{i+1} \| \| y_i \| \]

and from (13) we have \( s_i^T y_i \leq L\| s_i \|^2 \). Then \( d_i^T y_i \leq \alpha L \| d_i \|^2 \)

then we have

\[ \leq -\sigma \alpha L \| d_i \|^2 \frac{\| g_{i+1} \| \| y_i \|}{\| g_{i+1} \|^2} + \left( \frac{L\| s_i \|^2}{\| g_{i+1} \|^2} \right) \frac{\| g_{i+1} \| \| d_i \|^2}{\| g_{i+1} \|^2} \]

\[ \leq \frac{\| y_i \|}{\| g_{i+1} \|^2} \frac{\alpha \alpha L \| d_i \|^2}{\| g_{i+1} \|^2} + \left( \frac{L\| s_i \|^2}{\| g_{i+1} \|^2} \right) \frac{\| g_{i+1} \| \| d_i \|^2}{\| g_{i+1} \|^2} \]

Let \( c_1 = \frac{\| y_i \|}{\| g_{i+1} \|^2} + \frac{\alpha \alpha L \| d_i \|^2}{\| g_{i+1} \|^2} \)

\[ \frac{d_{i+1}^T g_{i+1}}{\| g_{i+1} \|^2} + 1 \leq c_1 \]

\[ \frac{d_{i+1}^T g_{i+1}}{\| g_{i+1} \|^2} \leq c_1 - 1 \]

\[ \frac{d_{i+1}^T g_{i+1}}{\| g_{i+1} \|^2} \leq -(1 - c_1) \| g_{i+1} \|^2 \]

Where \( c_1 = 1 - c_1 \) the proof is complete.

**Lemma 3.4 (see [5])**

If the Assumption 3.1 holds, and suppose that any conjugate gradient method of the form (2), (3) where \( d_i \) is satisfying the descent condition and \( a_i \) is satisfied the strong Wolfe line search.

If \( \sum_{i=1}^{\infty} \| d_i \|^2 = \infty \) we have that 

\[ \lim_{i\to\infty} \inf \| g_i \| = 0 \]

(19)

**Theorem 3.5**

Suppose that Assumption 3.1 hold and the function is a uniformly convex. Consider the conjugate gradient method in the form (2) , (11) where \( d_{i+1} \) satisfy the sufficient descent condition then the new method satisfy the global convergence (i.e. \( \lim_{i\to\infty} \| g_i \| = 0 \) ).

Proof:

\[ d_{i+1} = -\theta_{i+1} g_{i+1} + \beta_i d_i \]

\[ \| d_{i+1} \| = \| \theta_{i+1} g_{i+1} + \beta_i d_i \| \]

\[ | \beta_i | \leq \frac{g_{i+1}^T y_i}{d_i^T y_i} + \frac{s_i g_{i+1}^T d_i}{d_i^T d_i} \]

since \( s_i g_{i+1} \leq s_i^T y_i \)

\[ \leq \frac{\| g_{i+1} \| \| y_i \|}{\alpha M \| d_i \|^2} + \frac{s_i^T y_i}{\alpha M \| d_i \|^2} + \frac{\sigma g_{i+1}^T d_i}{g_{i+1}^T d_i} \]

\[ \leq \frac{\| g_{i+1} \| \| y_i \|}{\alpha M \| d_i \|^2} + \frac{\| g_{i+1} \| \| y_i \|}{\alpha M \| d_i \|^2} + \frac{\sigma g_{i+1}^T d_i}{g_{i+1}^T d_i} \]

(21)
\theta_{n+1} = 1 - \frac{g_{i+1}^T d_i}{\nabla^2 g_i} d_i \frac{y_i}{g_{i+1}^T y_i} \\
| \theta_{n+1} | \leq 1 + \left| \frac{\sigma g_i^T d_i - \alpha L \|d_i\|^2}{g_{i+1}^T y_i} \right| \\
\leq 1 + \frac{\sigma a L \|d_i\|^2}{|g_{i+1}^T y_i|}

since \( g_{i+1}^T y_i = \|g_{i+1}\|^2 - g_{i+1}^T g_i \)
and since \( |g_{i+1}^T g_i| \geq 0.2 \|g_{i+1}\|^2 \) (Powell restart \footnote{13})
\( g_{i+1}^T y_i \geq \|g_{i+1}\|^2 - 0.2 \|g_{i+1}\|^2 \)
i.e. \( \frac{g_{i+1}^T y_i}{|g_{i+1}^T y_i|} \leq \frac{0.8 \|g_{i+1}\|^2}{|g_{i+1}^T y_i|} \)
\|d_{i+1}\| = c_4

use (21) and (22) in (20) then we get:
\[ \frac{1}{\|d_{i+1}\|^2} \geq \frac{1}{c_5} \sum_{i \geq i} 1 = \infty \]

by lemma (3.4), it follows that (19) is true, which for uniformly convex function is equivalent to (23).

5. NUMERICAL RESULTS

In order to perform demonstrate the analytical results and effectiveness of such computational algorithm each of these method reached the optimal results. All program are rewritten in FORTRAN language by using a set of well known unconstrained optimization test functions these test function are contributed in CUTE\footnote{15}. In Table(1) we have compared our new algorithm with Perry method, take the dimension(n=1000,10000) where Table(2) using the total dimension (n=1000,2000,\ldots,10000) for every test function. The comparative performance of the algorithm is evaluated by considering both the (NOF) which is number of function evaluations and the (NOI) which is the number of iterations and all these methods terminate when the following stopping criterion when \( \|g_i\| \leq \epsilon, \epsilon = 10^{-7} \)

<table>
<thead>
<tr>
<th>N</th>
<th>Test fu.</th>
<th>Dim</th>
<th>Perry</th>
<th>Newmethod</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Extended freudentein &amp; roth</td>
<td>1000</td>
<td>13(8)</td>
<td>68(60)</td>
</tr>
<tr>
<td>2</td>
<td>Trigonometric</td>
<td>1000</td>
<td>29(28)</td>
<td>29(18)</td>
</tr>
<tr>
<td>3</td>
<td>Rosenbrock</td>
<td>1000</td>
<td>39(22)</td>
<td>34(18)</td>
</tr>
<tr>
<td>4</td>
<td>White &amp; holst</td>
<td>1000</td>
<td>32(27)</td>
<td>35(20)</td>
</tr>
<tr>
<td>5</td>
<td>Beale</td>
<td>1000</td>
<td>12(7)</td>
<td>35(20)</td>
</tr>
<tr>
<td>6</td>
<td>Penalty</td>
<td>1000</td>
<td>2(2)</td>
<td>14(8)</td>
</tr>
<tr>
<td>7</td>
<td>perturbed quadratic</td>
<td>1000</td>
<td>320(84)</td>
<td>314(89)</td>
</tr>
<tr>
<td>8</td>
<td>Raydon 2</td>
<td>1000</td>
<td>4(4)</td>
<td>4(4)</td>
</tr>
<tr>
<td>9</td>
<td>Diagonal 2</td>
<td>1000</td>
<td>183(56)</td>
<td>209(74)</td>
</tr>
<tr>
<td>10</td>
<td>Generalized tridiagonal 1</td>
<td>1000</td>
<td>24(5)</td>
<td>27(14)</td>
</tr>
<tr>
<td>11</td>
<td>Extended three exponential terms</td>
<td>1000</td>
<td>10(6)</td>
<td>8(4)</td>
</tr>
<tr>
<td>12</td>
<td>Extended Himmeblau</td>
<td>1000</td>
<td>19(10)</td>
<td>10(6)</td>
</tr>
</tbody>
</table>

Table(1): Numerical Comparisons between the Perry method and New method.
Table (2): The total Numerical Comparisons between the Perry method and New method.

<table>
<thead>
<tr>
<th>N</th>
<th>Test fu.</th>
<th>Perry</th>
<th>Newmethod</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Extended freudentein &amp; roth</td>
<td>629(563)</td>
<td>194(133)</td>
</tr>
<tr>
<td>2</td>
<td>Trigonometric</td>
<td>204(168)</td>
<td>203(168)</td>
</tr>
<tr>
<td>3</td>
<td>Rosenbrock</td>
<td>326(177)</td>
<td>350(200)</td>
</tr>
<tr>
<td>4</td>
<td>White &amp; holst</td>
<td>120(79)</td>
<td>139(79)</td>
</tr>
<tr>
<td>5</td>
<td>Beale</td>
<td>42(42)</td>
<td>42(42)</td>
</tr>
<tr>
<td>6</td>
<td>Penalty</td>
<td>8042(2204)</td>
<td>7713(2102)</td>
</tr>
<tr>
<td>7</td>
<td>perturbed quadratic</td>
<td>4113(1194)</td>
<td>4188(1314)</td>
</tr>
<tr>
<td>8</td>
<td>Raydon 2</td>
<td>306(149)</td>
<td>341(189)</td>
</tr>
<tr>
<td>9</td>
<td>Diagonal 2</td>
<td>109(65)</td>
<td>81(42)</td>
</tr>
<tr>
<td>10</td>
<td>Generalized tridiagonal1</td>
<td>190(100)</td>
<td>172(92)</td>
</tr>
<tr>
<td>11</td>
<td>Extended three exponential</td>
<td>67(50)</td>
<td>67(50)</td>
</tr>
<tr>
<td>12</td>
<td>Himmelblau</td>
<td>668(333)</td>
<td>664(303)</td>
</tr>
<tr>
<td>13</td>
<td>quadratic Diagonal</td>
<td>3444(615)</td>
<td>3439(607)</td>
</tr>
<tr>
<td>14</td>
<td>Extended maratos</td>
<td>8000(2186)</td>
<td>7705(2157)</td>
</tr>
<tr>
<td>15</td>
<td>Extended psc1</td>
<td>415(209)</td>
<td>408(208)</td>
</tr>
<tr>
<td>16</td>
<td>Quadratic function qf1</td>
<td>13896(2609)</td>
<td>13866(2436)</td>
</tr>
<tr>
<td>17</td>
<td>Extended quadratic penalty QP2</td>
<td>5842(3076)</td>
<td>5599(2899)</td>
</tr>
<tr>
<td>18</td>
<td>Nondquar</td>
<td>1771(1655)</td>
<td>1347(1240)</td>
</tr>
</tbody>
</table>

CONCLUSION

The research concludes that the proposed new scalar $\theta_{i+1}$ of conjugate gradient methods is very effective and increase ability when applied on problem of high dimensional. Under some conditions, we establish that the new proposed method is globally convergent for uniformly convex.

Table (1) and Table (2), showing that the new algorithm is superior to reach of the optimal solution depend on NOF and NOI.

Page | 17
APPENDIX

The specifics the test functions that used in search can be find in [2].

1. Extended Freudentein and Roth Function.
2. Trigonometric Function.
5. Beale Function.
7. Perturbed quadratic Function.
10. Generalization tridiagonal function.
11. Extended three exponential terms.
12. Extended Himmelblau Function.
14. Extended Psc1 Function.
15. Quadrati Diagonal Perturbed Function.
16. Quadrati function QF.
17. Extended quadratic penalty QP2 function.
20. Fletcher Function.

REFERENCES