A New Version Coefficient of Three-Term Conjugate Gradient Method to Solve Unconstrained Optimization

Alaa Luqman Ibrahim and Banaz Hamza Jahwar

1Department of Mathematics, Faculty of Science, University of Zakho, Zakho, Kurdistan Region, Iraq.
2Department of Mathematics, College of Science, University of Duhok, Duhok, Kurdistan Region, Iraq.

Received: 9 February 2022, Accepted: 1 July 2022
Published online: 31 October 2022.

Abstract: This paper presents the new three-term conjugate gradient method for solving unconstrained optimization problem. The main aim is to upgrade the search direction of conjugate gradient method to present a more active new three term method. Our new method satisfies the descent and the sufficient descent conditions and global convergent property. Furthermore, the numerical results show that the new method has a better numerical performance in comparison with the standard (PRP) method from an implementation of our new method on some test functions of unconstrained optimization according to number of iterations (NOI) and the number of functions evaluation (NOF).

Keywords: Unconstrained optimization, three-term Conjugate gradient method, descent and sufficient descent conditions and global convergent.

1 Introduction

The main idea of the unconstrained optimization problems is minimizing an objective function that depends on real variable, with no restrictions at all on the values of these variables. The mathematical formulation is

$$\min f(x) \quad \forall x \in \mathbb{R}^n$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a smooth function and its gradient $g$ is available, we start from the initial guess point $x_0$ to solve the problem (1). A nonlinear conjugate gradient (CG) method is an iterative scheme that usually generated a sequence $\{x_i\}$ of an approximation to the solution of the problem (1), using the repetition:

$$x_{i+1} = x_i + \lambda_i d_i, \quad i = 0, 1, 2, \ldots, \quad \lambda_i > 0$$

where $\lambda_i$ is determined by some line searches (Goldstein condition, wolfe condition, curvature condition and sufficient decent condition) and the search direction $d_i$ is designed by:

$$d_{i+1} = \begin{cases} -g_{i+1} & \text{if } i = 0 \\ -g_{i+1} + \beta_i d_i & \text{if } i \geq 1 \end{cases}$$

where $g_i = \nabla f(x_i)$ and $\beta_i$ is an important parameter. The different choices for the parameter $\beta_i$ correspond to different CG methods. For examples, Polak-Ribiere-Polyak (PRP) [8], Hestenes and Stiefel (HS) [5], Fletcher and Reeves (FR) [4], Alaa et al. method [2], Some well-known formulas for $\beta_i$ given below:

$$d_{i+1} = -g_{i+1} + \beta^{PRP}_i \frac{g_i^T y_i}{g_i^T g_i} d_i$$

* Corresponding author e-mail: alaa.ibrahim@uoz.edu.krd, banaz.jahwar@uod.ac

© 2022 BISKA Bilisim Technology
\[ d_{i+1} = -g_{i+1} + \beta_t^{HS} \frac{g_i^T y_i}{d_i^T y_i} d_i \] (5)
\[ d_{i+1} = -g_{i+1} + \beta_t^{FR} \frac{g_i^T g_{i+1}}{g_i^T g_{i+1}} d_i \] (6)
\[ d_{i+1} = -g_{i+1} + \beta_t \left( \frac{g_i^T y_i}{g_i^T g_i} - \frac{g_{i+1}^T y_{i+1}}{g_{i+1}^T g_{i+1}} \right) d_i \] (7)

where \( y_i = g_{i+1} - g_i \) and \( v_i = \lambda_i d_i = x_{i+1} - x_i \). In the convergence analysis and implementation of conjugate gradient method, one often requires the exact and inexact line search such as the Wolfe conditions or the strong Wolfe conditions. The Wolfe line search is to find \( \lambda_i \) such that

\[ f(x_i + \lambda_i d_i) \leq f(x_i) + \theta \lambda_i g_i^T d_i \] (8)
\[ d_i^T g(x_i + \lambda_i d_i) \geq \sigma g_i^T d_i \] (9)

with \( 0 < \theta < \sigma \). The strong Wolfe line search is to find \( \lambda_i \) such that

\[ f(x_i + \lambda_i d_i) \leq f(x_i) + \theta \lambda_i g_i^T d_i \] (10)
\[ |d_i^T g(x_i + \lambda_i d_i)| \leq \sigma g_i^T d_i \] (11)

where \( 0 < \theta < \sigma < 1 \) are constants, by Li and Weijun, [6]. There are other types of conjugate gradient methods called three-term conjugate gradient methods. That it is one of the most reliable and attractive method and this method numerically stronger than classical conjugate gradient method. Among the generated three-term conjugate gradient methods in the literature we have the three-term conjugate methods proposed by Nazareth [7] present a computationally efficient three-term nonlinear conjugate gradient method with the search direction:

\[ d_{i-1} = -y_i + \frac{\gamma^T y_i}{\gamma^T d_i} d_i - \frac{y_{i-1}^T y_i}{y_{i-1}^T d_{i-1}} d_{i-1} \] (12)

Alaa L. I., and Salah G. Sh., in (2019) [1], submit a new class of three-term conjugate gradient method as:

\[ d_{i+1} = \begin{cases} -g_{i+1} & \text{if } i = 0 \\ -g_{i+1} + \beta_t^{HS} d_i - t_i (\frac{g_i^T d_i}{d_i^T y_i}) y_i & \text{if } i \geq 1 \end{cases} \] (13)

where \( t_i = \gamma \frac{||\gamma||}{||\gamma||} + (1-\gamma) \frac{g_i^T y_i}{||\gamma||} \) and the parameter \( \beta_t = \beta_t^{PRP} , \beta_t^{HS} \) or \( \beta_t^{FR} \).

Another important class of the three-term conjugate gradient method proposed by Zhang et al. [9, 10] by considering a descent modified PRP and also a descent modified HS conjugate gradient method as

\[ d_{i+1}^{PRP} = \begin{cases} -g_{i+1} & \text{if } i = 0 \\ -g_{i+1} + \beta_t^{PRP} d_i - t_i (\frac{g_i^T d_i}{d_i^T y_i}) y_i & \text{if } i \geq 1 \end{cases} \] (14)

and

\[ d_{i+1}^{HS} = \begin{cases} -g_{i+1} & \text{if } i = 0 \\ -g_{i+1} + \beta_t^{HS} d_i - t_i (\frac{g_i^T d_i}{d_i^T y_i}) y_i & \text{if } i \geq 1 \end{cases} \] (15)

This paper is organized as follow: in Section 2, we will present a new three-term conjugate gradient method as we prove the descent condition and sufficient descent condition of new method then we will study the global convergence property of our new three term method. In section 3, some numerical experiments to this new three term conjugate gradient method are reported. In section 4, we will present the conclusion.
2 A New Three-Term Conjugate Gradient Method

In this section, we will derive our suggestion based on the search direction that is present by Alaa et al. [2]. We suggested a three-term conjugate gradient method, in which the search direction is:

\[ d_{i+1}^{NEW} = -g_{i+1} + \beta_i^{PRP} d_i + \beta_i^{NEW} y_i \]  

(16)

Now, from equation (7) and equation (16), we have

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

(17)

Multiplying both sides of above equation by \( y_i \), we obtain

\[ \beta_i^{NEW} y_i = -i \frac{g_i^T y_i}{g_i^T g_i} d_i y_i \]

(18)

This implies that,

\[ \beta_i^{NEW} = -i \frac{g_i^T y_i}{g_i^T g_i} \]

(19)

To achieve balance, we will suppose that the value of parameter \( \gamma \) is

\[ \gamma = \frac{g_i^T y_i}{g_i^T g_i} \]

(18)

where \( \gamma > 0 \), so,

\[ d_{i+1}^{NEW} = -g_{i+1} + \beta_i^{PRP} d_i + \beta_i^{NEW} y_i \]

or

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

(20)

2.2 Algorithm of New Method

1. Given an initial point \( x_0 \in R^n \)
2. Set \( d_0 = -g_0 \), \( i = 0 \).
3. If \( \|g_i\| = 0 \) then stop, otherwise go to Step 4.
4. Compute the step size \( \lambda_i \) by using minimize \( f(x_i + \lambda_i d_i) \).
5. Set \( x_{i+1} = x_i + \lambda_i d_i \).
6. Determine \( g_{i+1} \), if \( \|g_{i+1}\| \leq 10^{-5} \) stop, else go to Step 7.
7. Compute \( d_{i+1} \) by (20).
8. If \( \|g_{i+1}\| \leq \frac{\|g_{i+1}\|}{\|g_i\|} \) is satisfied go to step 3,

else \( i = i + 1 \) and go to step 4.

**Theorem 2.1:** Assume that the sequence \( \{x_i\} \) is generated by (2), then the search direction in (20) satisfy the descent condition, i.e

\[ g_{i+1}^T d_{i+1} \leq 0 \]

(21)

**Proof:** Multiply both sides of above equation by \( g_{i+1} \), to obtain

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

(22)
If the above search direction is exact, then it satisfies the descent condition i.e.

\[ g^T_{i+1}d^\text{NEW}_{i+1} = -\|g_{i+1}\|^2 \leq 0. \]

However, if the search direction (22) is inexact (i.e.) \( g^T_{i+1}d_i \neq 0 \) we concludes that the first two terms of equation (22) are satisfies the descent condition i.e.

\[ -g^T_{i+1}g_{i+1} + \frac{g^T_{i+1}y_i}{g_i^Tg_i} g^T_{i+1}d_i \leq 0, \]  

(23)

because the PRP method satisfies the (21) condition.

Therefore, we can present the equation (22) as

\[ g^T_{i+1}d^\text{NEW}_{i+1} \leq -\gamma g^T_{i+1}y_i d^T_{i+1}y_i (g^T_{i+1}y_i)^2 \]

(24)

Since from wolfe condition we have \( -\sigma d^T_i g_i \leq g^T_{i+1}d_i \), by multiplying both side by \(-1\) we get \( -g^T_{i+1}d_i \leq d^T_i g_i \). So, the equation (24), can be written as

\[ g^T_{i+1}d^\text{NEW}_{i+1} \leq \gamma \frac{\sigma \lambda_i d^T_i g_i d^T_{i+1}y_i (g^T_{i+1}y_i)^2}{g_i^Tg_i y_i^Ty_i} \]

Clearly, \( \gamma, \lambda_i, \sigma, d^T_i y_i, (g^T_{i+1}y_i)^2 \), \( g_i^Tg_i \) and \( y_i^Ty_i \) are positive. But \( d^T_i g_i \leq 0 \)

So, we have

\[ g^T_{i+1}d^\text{NEW}_{i+1} \leq 0 \]

**Theorem 2.2:** Suppose \( \{x_i\} \) is a sequence generated by (2). Then the search directions defined by (20), satisfy the sufficient descent condition,

\[ g^T_{i+1}d_{i+1} \leq -\gamma \|g_{i+1}\|^2 \]  

(25)

**Proof:** It is obvious from theorem 2.1, after multiplying the new search direction (20) by \( g^T_{i+1} \) that the first two terms of equation (20) are less than or equal to zero and from equation (24), we have

\[ g^T_{i+1}d^\text{NEW}_{i+1} \leq -[\gamma \frac{\lambda_i (d^T_i y_i)^2 (g^T_{i+1}y_i)^2}{g_i^Tg_i y_i^Ty_i \|g_{i+1}\|^2} \]  

(26)

Let \( C = \gamma \frac{\lambda_i (d^T_i y_i)^2 (g^T_{i+1}y_i)^2}{g_i^Tg_i y_i^Ty_i \|g_{i+1}\|^2} \) which is positive, then

\[ g^T_{i+1}d^\text{NEW}_{i+1} \leq -C \|g_{i+1}\|^2 \]

Some necessary assumptions for the objective function to prove the global convergence property are given as follows.

**Assumption 2.1:** The level set \( \delta = \{x \mid f(x) \leq f(x_i)\} \) at \( x_i \) is bounded, there exists a constant \( a > 0 \) such that

\[ \|x\| \leq a, \forall x \in \delta. \]  

(27)

In some neighborhood \( N \) of \( \delta \), \( f \) is continuously differentiable, and its gradient is Lipschitz continuous with Lipschitz constant \( L > 0 \), i.e.

\[ \|g(r)-g(t)\| \leq \delta \|r-t\| \forall r, t \in \delta \]  

(28)

From the above assumptions, that there exists a positive constant \( b \) such that

\[ \|g(b)\| \leq b \forall x \in \delta. \]  

(29)
If function \( f \) is strongly convex, then there exists a constant \( \mu > 0 \) such that

\[
\mu \| x-y \|^2 \leq (\nabla f (x) - \nabla f (y))^T (x - y), \text{forall} \ x, y \in \delta
\] (30).

**Lemma 2.1:** [3, 5] Let assumption (2.1) hold. Consider the methods (1) and (2), where \( d_i \) is a descent direction and \( \lambda_i \) satisfies the e standard Wolfe line search. If

\[
\sum_{i=1}^{\infty} \frac{1}{\| d_i \|^2} = \infty.
\] (31)

Then,

\[
\liminf_{i \to \infty} \| g_i \| = 0.
\] (32)

**Theorem 2.3:** Suppose the assumption (2.1) hold and consider the new method in (2) and (20) where \( \lambda_i \) is computed by the strong Wolfe conditions satisfies the global convergence.

**Proof:**

\[
d_{i+1}^{NEW} = -g_{i+1} + \frac{g_{i+1}^T y_i}{g_i^T g_i} d_i - \gamma \frac{g_{i+1}^T v_i d_i g_{i+1}^T y_i}{g_i^T g_i y_i} y_i;
\]

\[
\| d_{i+1}^{NEW} \| \leq \| -g_{i+1} \| + \frac{g_{i+1}^T y_i}{g_i^T g_i} \| d_i \| + \gamma \frac{g_{i+1}^T v_i d_i g_{i+1}^T y_i}{g_i^T g_i y_i} \| y_i \| \] (33)

Since \( g_{i+1}^T v_i \leq \lambda_i d_i^T y_i \), so, we have

\[
\| d_{i+1}^{NEW} \| \leq \| -g_{i+1} \| + \frac{g_{i+1}^T y_i}{g_i^T g_i} \| d_i \| + \gamma \lambda_i \| d_i \|^2 \| g_{i+1} \| \| y_i \| \] (34)

Hence, form Lipschitz condition we have \( \| y_i \| \leq \delta \| v_i \| \) and from equation (3) we have \( d_i = -g_i \), also we have this form \( v_i = \lambda_i d_i = -\lambda_i g_i \),

\[
\| d_{i+1}^{NEW} \| \leq \| -g_{i+1} \| + \frac{\delta \lambda_i \| g_i \|}{\| g_i \|^2} \| d_i \| + \gamma \lambda_i \| d_i \|^2 \delta \lambda_i^2 \| g_{i+1} \| \] (35)

Hence

\[
\| d_{i+1}^{NEW} \| \leq \| -g_{i+1} \| + \frac{\delta \lambda_i \| g_i \|}{\| g_i \|^2} \| d_i \| + \gamma \lambda_i \| d_i \|^2 \delta \lambda_i^2 \| g_{i+1} \| \] (36)

\[
\therefore \sum_{i=1}^{\infty} \frac{1}{\| d_{i+1}^{NEW} \|^2} \geq \sum_{i=1}^{\infty} \frac{1}{M} = \infty
\]

\[
\Rightarrow \sum_{i=1}^{\infty} \frac{1}{\| d_{i+1}^{NEW} \|^2} = \infty
\]

By using lemma 2.1, we get

\[
\liminf_{i \to \infty} \| g_{i+1} \| = 0. \blacksquare
\]

### 3 Numerical results and discussion

In this section, we will study the implementation of the new three term conjugate gradient method. The tests include familiar nonlinear problems with different dimension. The code of the method is written by FORTRAN 95 language with given initial points. Table (1) illustrate that the numerical results of our new method is more effective than PRP method in the same field with respect to the number of iterations (NOI) and the number of functions evaluation (NOF).
Table 1: The results of the new three term method with PRP method

<table>
<thead>
<tr>
<th>Test function</th>
<th>i</th>
<th>Standard form PRP</th>
<th>New Formula TT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOI</td>
<td>NOF</td>
<td>NOI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Powell</td>
<td>5</td>
<td>40</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>40</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>43</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>46</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>46</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>50</td>
<td>180</td>
</tr>
<tr>
<td>Mile</td>
<td>5</td>
<td>37</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>37</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>44</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>44</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>50</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>50</td>
<td>180</td>
</tr>
<tr>
<td>Central</td>
<td>5</td>
<td>22</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>22</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>22</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>23</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>23</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>30</td>
<td>270</td>
</tr>
<tr>
<td>Cubic</td>
<td>5</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>15</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>16</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>16</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>16</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>16</td>
<td>47</td>
</tr>
<tr>
<td>Wolfe</td>
<td>5</td>
<td>14</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>32</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>49</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>58</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>64</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>99</td>
<td>214</td>
<td>55</td>
</tr>
<tr>
<td>Sum</td>
<td>5</td>
<td>6</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>6</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>14</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>21</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>23</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>31</td>
<td>145</td>
</tr>
<tr>
<td>Wood</td>
<td>5</td>
<td>29</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>29</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>30</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>30</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>30</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>30</td>
<td>69</td>
</tr>
<tr>
<td>Total</td>
<td>1539</td>
<td>5233</td>
<td>1309</td>
</tr>
</tbody>
</table>
Table 2 shows the improve percentage of new three term method. The new method has improvement as compared to PRP method with 14.9448% in NOI and 16.224% in NOF and in general the rate of improvement in the new three term method is 15.5844% in NOI and NOF.

4 Conclusion

In this paper, we present a new method of three-term conjugate gradient method for solving unconstrained optimization. The main idea is improving the search direction of conjugate gradient method to new search direction. The descent and sufficient descent conditions of our new method are proved. Also, we study the global convergent property under standard assumptions. The numerical results on low and high dimensionality problems reported that the new method is more efficient than the PRP method.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

All authors have contributed to all parts of the article. All authors read and approved the final manuscript.

References

