A NEW THREE-TERM CONJUGATE GRADIENT ALGORITHM FOR SOLVING MINIMIZATION PROBLEMS

Dlovan Haji Omar a, Salah Gazi Shareef b, Bayda Ghanim Fathi c

a,b,c Faculty of Science, University of Zakho, Zakho, Kurdistan Region, Iraq

Received: 13 Apr., 2023 / Accepted: 11 Aug., 2023 / Published: 7 Dec., 2023. https://doi.org/10.25271/sjuoz.2023.11.3.1152

ABSTRACT: The method of optimization is used to determine the most precise value for certain functions within a certain domain; it is mostly studied and employed in the fields of mathematics, computer science, and physics. This work presents a novel three-term conjugate gradient (CG) approach for unconstrained optimization problems. Both the descending criteria and the sufficient descent criterion were met by the new approach. The novel method that has been proposed has been evaluated for global convergence. The outcomes of numerical trials on a few well-known test functions demonstrated how highly successful our new modified method is, depending on the number of iterations (NOI) and the number of functions to be evaluated (NOF).

KEYWORDS: Optimization, Conjugate Gradient Methods and Three Terms Conjugate Gradient.

1. INTRODUCTION

The field of applied mathematics known as numerical optimization seeks the best answer to a mathematical problem involving the maximization or minimization of a particular function. The function can represent a wide range of things, including the financial success of an enterprise, the effectiveness of a manufacturing procedure, or the precision of a statistical model.

The objective of numerical optimization is to identify the input values that, within any constraints or limitations, result in the best output value. Various optimization algorithms can be used, depending on the properties of the function being optimized, the kinds of constraints involved, and the computational resources available.

Gradient descent, Newton’s method, simulated annealing, genetic algorithms, and particle swarm optimization are a few of the frequently used optimization algorithms. These approaches employ iterative procedures to find the best solution, modifying the input values slightly at each iteration until the best solution is found or a stopping criterion is satisfied.

Iterative techniques such as conjugate gradient methods (CG) are used to solve linear systems of equations. They are particularly useful for solving large, sparse, symmetric positive definite (SPD) systems, which arise in many scientific and engineering applications.

CG methods can also be applied to nonlinear functions, but the approach is slightly different from the linear case. Nonlinear CG methods are used to find the minimum of a nonlinear function of several variables

\[
\text{Min } f(x), x \in \mathbb{R}^n
\]

where \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) is a real-valued continuously differentiable function. Starting from a first guess \( x_1 \in \mathbb{R}^n \), a nonlinear CG algorithm creates the sequence \( \{x_k\} \), defined as

\[
x_{k+1} = x_k + \alpha_k d_k
\]

The main idea of NCG methods is to iteratively create a sequence of search directions that are conjugate to the previous search directions. The previous search directions are linearly combined in each direction of the search, with certain coefficients that ensure that the search direction is a descent direction. The conjugacy condition ensures that the search directions are independent as possible, which leads to a faster convergence.

The popular CG methods are Fletcher-Reeves (FR) [1], the Hestenes-Stiefel (HS) [2], and the Polak-Rivière (PR) [3], Dai and Yuan (DY) [4], are respectively determined as follows:

\[
\beta_k^{FR} = \frac{g_{k+1}^T g_k}{g_k^T g_k}
\]

\[
\beta_k^{HS} = \frac{g_{k+1}^T y_k}{y_k^T y_k}
\]

\[
\beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k}
\]

\[
\beta_k^{DY} = \frac{\|y_k\|^2}{\|y_k\|^2}
\]

which is based on the following update rule for the search direction:

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

where \( g_k \) is a gradient of \( f \) at \( x_k \) and \( y_k = g_{k+1} - g_k \). The Euclidean norm of vectors is represented by the letter \( \| \| \). There are also many studies on this method see ([5,6,7,8,9]).

Three-term conjugate gradients (TT CG) are additional significant classes of CG which is itself an iterative algorithm used for solving linear systems of equations. The three-term CG method is a specific class of conjugate gradient methods, which are characterized by their use of conjugate search directions in each iteration.

The three-term CG method is an improvement over the classical CG method in terms of convergence rate, and is particularly useful for solving symmetric, positive-definite systems of linear equations. It is based on the idea of generating a sequence of three-term conjugate directions, which are linearly independent and orthogonal with respect to a symmetric positive-definite matrix. In each iteration of the algorithm, three terms are computed: the current solution vector, the residual vector, and the search direction vector.
The three-term CG method is a more general class of CG methods that includes other variants, such as, the HS and the PR methods. All of these approaches generate a sequence of conjugate directions, but they differ in how they compute the search direction vectors. The three-term CG method presented by Zhang et al. [10, 11] in the literature by taking into consideration a descent-modified PRP and also a descent-modified HS CG method as

\[ d_{k+1}^{\text{PRP}} = -g_{k+1} + \frac{\|g_{k+1}\|}{\|v_k\|} d_k - \frac{\|g_{k+1}\|}{\|v_k\|} y_k \]  

(1.8)

\[ d_{k+1}^{\text{HS}} = -g_{k+1} + \frac{\|g_{k+1}\|}{\|v_k\|} d_k - \frac{\|g_{k+1}\|}{\|v_k\|} y_k \]  

(1.9)

In the same way, Alaa and Salah in (2019) [12], proposed a new class of the three-term CG method that is computationally efficient and is defined in the following search direction:

\[ d_{k+1}^{\text{NTT-CG}} = -g_{k+1} + \beta_k d_k - t_k \frac{\partial \|v_k\|^2}{\partial y_k} y_k \]  

(1.10)

where \( t_k = \sqrt{\frac{\|v_k\|}{\|v_k\|} + 1 - \sqrt{\frac{\|v_k\|}{\|v_k\|}}} \), and the parameter \( \beta_k \) is given from normal CG methods (HS, PRP and FR).

This essay is organized as follows: In Section two, we'll submit a new three-term CG method suggestion. We demonstrate the descent and sufficient descent conditions of the new algorithm in section 3. Section 4 presents numerous numerical evaluations of our three-terms CG technique. Section 5 contains our concluding observations.

2. DERIVATION OF THE NEW DIRECTION OF THREE TERMS

In this part, we will create a new three terms direction. The main idea to drive the new search direction is replace the \( g_{k+1} \) in the third term of the search direction (1.8) by \( g_{k+1} \) that is defined below

\[ g_{k+1} = g_{k+1} + (1 - \delta)(\frac{\|v_k\|}{\gamma} - \|y_k\|) \]  

(2.1)

Where, \( \delta \in (0, 1) \), \( \mu = 0.1 \) and \( \gamma = \frac{2\mu}{\|v_k\|} (1 + \|x_k\|) \) and \( \omega \) is the machine error.

More specifically, the search direction of our method, named as (New 3-TT-CG) method, defined by:

\[ d_{k+1}^{\text{New}} = -g_{k+1} + \frac{\|g_{k+1}\|}{\|v_k\|} d_k - \frac{\|g_{k+1}\|}{\|v_k\|} y_k \]  

(2.2)

Where \( g_{k+1} = g_{k+1} + (1 - \delta)(\frac{\|v_k\|}{\gamma} - \|y_k\|) - \mu g_{k+1} \)

Or

\[ d_{k+1}^{\text{New}} = -g_{k+1} + \frac{\|g_{k+1}\|}{\|v_k\|} d_k - \frac{\|g_{k+1}\|}{\|v_k\|} y_k \]  

(2.3)

Where, \( \delta \in (0, 1) \), \( \mu = 0.1 \) and \( \gamma = \frac{2\mu}{\|v_k\|} (1 + \|x_k\|) \) and \( \omega \) is the machine error.

2.1 ALGORITHM OF (NEW TT-CG)

**Step 1:** Given an \( x_0 \in R^n \) and set \( d_0 = -g_0, k = 0 \).

**Step 2:** If \( \|g_k\| = 0 \) then stop, else go to Step 3.

**Step 3:** Set \( \alpha_k \) by using the cubic line search.

**Step 4:** Set \( x_{k+1} = x_k + v_k \).

**Step 5:** Compute \( g_{k+1} \), if \( \|g_{k+1}\| \leq 10^{-5} \) stop. Otherwise, go to Step 6.

**Step 6:** Compute \( d_{k+1} \) from the equation (3.3).

**Step 7:** If \( \|g_{k+1}\|^2 \leq \frac{\|v_{k+1}\|^2}{\alpha_k} \) is satisfied go to step 2.

Otherwise, set \( k = k + 1 \) and go to step 3.

3. GLOBAL CONVERGENCE AND (DESCENT AND SUFFICIENT DESCENT PROPERTIES) OF THE (NEW TT-CG).

**Theorem 3.1:** Suppose that the \( \{x_k\} \) is a sequence created by (1.2), then the \( d_{k+1}^{\text{NTT-CG}} \) satisfy the descent condition.

**Proof:** By multiplying both sides of equation(2.3) by \( g_{k+1} \) from right, we obtain

\[ d_{k+1}^{\text{New}} g_{k+1} = -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|}{\|v_k\|} d_k g_{k+1} - \frac{\|g_{k+1}\|}{\|v_k\|} y_k g_{k+1} \]  

(3.1)

If \( d_{k+1}^T g_{k+1} = 0 \), (which is mean \( \alpha_k \) is chosen by an exact line search), then we get \( d_{k+1}^{\text{New}} g_{k+1} = -\|g_{k+1}\|^2 \leq 0 \). If we have inexact line search which is \( d_{k+1} g_{k+1} \neq 0 \).

By mathematical induction, from the first search direction, we get

\[ d_{k+1}^T g_k = -\|g_k\|^2 \leq 0 \], and we assume that it is true for case \( k \) that is mean

\[ d_{k+1}^T g_k \leq 0 \]. To prove case \( k + 1 \)

Because the PR parameter satisfies the descent condition, the first two terms of the previous formula are less than or equal to zero.

We just need to demonstrate that the third term is less than or equal to zero at this point.

\[ -\frac{\|g_{k+1}\|^2}{\|v_k\|} y_k = -\frac{\|g_{k+1}\|^2}{\|v_k\|} \gamma - \|y_k\| \]  

from Lipschitz Condition \( \|y_k\| \leq L\|v_k\| \) and \( \|v_k\| \leq L\|x_k\|^T d_k \) where \( L > 0 \) (3.2)

Then,

\[ d_{k+1}^{\text{New}} g_{k+1} \leq -\frac{\|g_{k+1}\|^2}{\|v_k\|} \frac{d_k^T g_{k+1}}{\|v_k\|} \]  

\[ = -1 + (1 - \delta) - \mu \frac{d_k^T g_{k+1}}{\|v_k\|} \frac{(1 - \delta)}{\gamma} \]  

\[ < 0 \]  

(3.3)

Finally, we have \( d_{k+1}^{\text{New}} g_{k+1} \leq 0 \).

**Theorem 3.2:** If \( \{x_k\} \) is a sequence generated by (1.2), then the search direction in (2.3) satisfies the sufficient descent condition.

**Proof:** From equations (3.1) and (3.3), we have

\[ d_{k+1}^{\text{New}} g_{k+1} \leq -1 + (1 - \delta) - \mu \]  

\[ \frac{\|g_{k+1}\|^2}{\|v_{k+1}\|^2} \leq \frac{\|g_{k+1}\|^2}{\|v_{k+1}\|^2} \]  

\[ \|g_{k+1}\|^2 \]  

\[ \|g_{k+1}\|^2 \]
Let $C = \left(1 + \frac{(1 - \delta)x}{\gamma}\right) - (1 - \delta)\mu L \left\frac{d_{k}^{2}B_{k+1}}{d_{k}^{2}B_{k+1}}\right)^{2}$

Then, we have $d_{k+1}^{2}B_{k+1} \leq -C\|g_{k+1}\|^{2}.$

The global convergence of (New TT-CG) method will be presented by the following theorem.

**Assumption 3.1** [13]. The level set $S = \{x: x \in R^{n}, f(x) \leq f(x_{0})\}$ is bounded. i.e. $\exists B > 0$ such that $\|x\| \leq B, \forall x \in S$ \hspace{2cm} (3.4)

**Assumption 3.2** In a neighbourhood $\Omega \in S$, $f$ is differentiable and its gradient $g$ is Lipschitz continuous, i.e. $\exists L > 0$ such that $\|g(x) - g(x_{k})\| \leq L\|x - x_{k}\|, \forall x, x_{k} \in \Omega$ \hspace{2cm} (3.5)

From Assumptions (3.1) and (3.2), $\exists M > 0$ such that $\|g(x)\| \leq M, \forall x \in S$. \hspace{2cm} (3.6)

**Lemma 3.1** [14]. The sequence $\{x_{k}\}$ is produced by the equations (1.2) and (1.7), where $d_{k}$ satisfies the descent condition and $a_{k}$ is determined by strong Wolfe conditions and by assuming that Assumptions (3.5) and (3.5) are true. If

$$\sum_{k=1}^{\infty} \frac{1}{\|g_{k}\|} = \infty$$ \hspace{2cm} (3.7)

The $\lim \inf_{k \to \infty} \|g_{k}\| = 0.$ \hspace{2cm} (3.8)

**Theorem 3.3**: Suppose that assumptions (3.1) and (3.2), hold that if any iteration of the equations (1.2) and (2.3) and $\alpha_{k}$ satisfies the strong Wolfe line search conditions, then

$$\lim \inf_{k \to \infty} \|g_{k}\| = 0.$$

**Proof:** From equation (3.2), we have

$$\|d_{k+1}^{\text{New}}\| \leq \|g_{k+1}\| + \frac{\|g_{k+1}^{T}y_{k}\|}{\|d_{k}\|} \|g_{k+1}\| + \frac{\|g_{k+1}^{T}a_{k}\|}{\|d_{k}\|} \|y_{k}\|.$$ \hspace{2cm} (3.9)

by using (3.6) and (3.2)

$$\|d_{k+1}^{\text{New}}\| \leq M + \frac{1}{\|d_{k}\|} \|d_{k}\| + (1 + (1 - \delta)(\frac{1}{\gamma})) - \mu \|y_{k}\|.$$ \hspace{2cm} (3.10)

since, $d_{k} = -g_{k}$ and $\|g_{k+1}^{T}d_{k}\| \leq M\|d_{k}\|.$ So,

$$\|d_{k+1}^{\text{New}}\| \leq M + LM + (1 + (1 - \delta)\left(\frac{1}{\gamma}\right)) - \mu \|y_{k}\|.$$ \hspace{2cm} (3.10)

$$= \sum_{k=1}^{\infty} \frac{1}{\|d_{k}^{2}B_{k+1}\|^2} \geq \sum_{k=1}^{\infty} \frac{1}{\|d_{k}^{2}B_{k+1}\|^2} = \infty$$

$$= \sum_{k=1}^{\infty} \frac{1}{\|d_{k}^{2}B_{k+1}\|^2} = \infty$$

Now, by using lemma (3.1), we get $\|g_{k}\| = 0.$ \hspace{2cm} \boxed{\blacksquare}

4. **NUMERICAL RESULTS**

In this section, the numerical outcomes of the New Three Terms CG Method and the Traditional Method PR Method are compared, along with their performances. The comparative tests call for nine different functions with well-known non-linear problems where $4 \leq k \leq 5000$. The code was also written in the Fortran 95 programming language. The (NOI) and the (NOF) are shown in the comparative results shown in Table 4.1. Table (4.2) contains additional experimental findings that support the New TT-CG’s superiority to the traditional CG method in terms of the NOI and NOF.

<table>
<thead>
<tr>
<th>Test function</th>
<th>DIM.</th>
<th>PR</th>
<th>NOI</th>
<th>NOF</th>
<th>New TT-CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-Cantrel</td>
<td>1</td>
<td>10</td>
<td>22</td>
<td>159</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>22</td>
<td>159</td>
<td>21</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>23</td>
<td>171</td>
<td>22</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>23</td>
<td>171</td>
<td>22</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>30</td>
<td>270</td>
<td>23</td>
<td>175</td>
</tr>
<tr>
<td>G-Wolfe</td>
<td>4</td>
<td>15</td>
<td>45</td>
<td>14</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>16</td>
<td>47</td>
<td>15</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>16</td>
<td>47</td>
<td>15</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>16</td>
<td>47</td>
<td>15</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>16</td>
<td>47</td>
<td>15</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>16</td>
<td>47</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td>Cubic</td>
<td>4</td>
<td>37</td>
<td>116</td>
<td>39</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>37</td>
<td>116</td>
<td>40</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>44</td>
<td>148</td>
<td>40</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>44</td>
<td>148</td>
<td>43</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>50</td>
<td>180</td>
<td>46</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>50</td>
<td>180</td>
<td>46</td>
<td>160</td>
</tr>
<tr>
<td>Miele</td>
<td>4</td>
<td>29</td>
<td>67</td>
<td>26</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>29</td>
<td>67</td>
<td>26</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>30</td>
<td>69</td>
<td>26</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>30</td>
<td>69</td>
<td>26</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>30</td>
<td>69</td>
<td>26</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>30</td>
<td>69</td>
<td>27</td>
<td>63</td>
</tr>
<tr>
<td>G-Wood</td>
<td>4</td>
<td>30</td>
<td>85</td>
<td>28</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>30</td>
<td>85</td>
<td>28</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>30</td>
<td>85</td>
<td>28</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>30</td>
<td>85</td>
<td>28</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>30</td>
<td>85</td>
<td>28</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>30</td>
<td>85</td>
<td>28</td>
<td>66</td>
</tr>
<tr>
<td>Total</td>
<td>1170</td>
<td>3825</td>
<td>1053</td>
<td>3348</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Comparison Between Two Algorithms (PR. and New1 TT-CG)

<table>
<thead>
<tr>
<th></th>
<th>PR</th>
<th>New TT-CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOI</td>
<td>100%</td>
<td>90%</td>
</tr>
<tr>
<td>NOF</td>
<td>100%</td>
<td>89.524%</td>
</tr>
</tbody>
</table>

Table 3.2: The Percentage Of Improving Of New TT-CG Algorithm
The suggested strategy raises NOI and NOF by 10.476% and 10%, respectively. The modified new three term CG method has generally improved by 10.238% when compared to the standard PR approach.

CONCLUSIONS

For issues involving unconstrained optimization, we proposed a novel three-term CG approach in this study. The proofs using exact and approximate line searches must both be in decent condition and be sufficiently decent condition. We demonstrate that the suggested strategy converges globally for general functions. The numerical results clearly show that the new approach uses fewer iterations and more function evaluations per iteration than the old one.

REFERENCES


