A new parameter in three-term conjugate gradient algorithms for unconstrained optimization

Alaa Saad Ahmed1, Hisham M. Khudhur2, Mohammed S. Najmuldeen3

1Computer Science Department, College of Education for pure sciences, University of Mosul, Mosul, Iraq
2Mathematics Department, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq
3Ministry of Education, Kirkuk, Iraq

ABSTRACT
In this study, we develop a different parameter of three term conjugate gradient kind, this scheme depends principally on pure conjugacy condition (PCC). Whereas, the conjugacy condition (PCC) is an important condition in unconstrained non-linear optimization in general and in conjugate gradient methods in particular. The proposed method becomes converged, and satisfy conditions descent property by assuming some hypothesis. The numerical results display the effectiveness of the new method for solving test unconstrained non-linear optimization problems compared to other conjugate gradient algorithms such as Fletcher and Reeves (FR) algorithm and three term Fletcher and Reeves (TTFR) algorithm. The numerical results demonstrate the efficacy of the suggested method for solving test unconstrained nonlinear optimization problems from where a number of iterations and evaluation of function and A comparison of the time taken to perform the functions.

Keywords:
Algorithm
Conjugate gradient
pure conjugacy
Unconstrained optimization

1. INTRODUCTION
Researchers have studied the problem of unrestricted improvement as a matter of finding a solution to the minimization of the real function \( f(x) \).

\[
\min_{x \in \mathbb{R}^n} f(x)
\]

Whereas \( f(x) \) a derivative function at least once.
Conjugate gradient (CG) algorithms are important to solve for (1) problem using the following iterative method:

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

Whereas \( \alpha_k \) calculates the step size in either exactly or inexactly line search using the following relation:

\[
f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha_k d_k)
\]

Corresponding Author:
Hisham M. Khudhur
Mathematics Department
College of Computer Science and Mathematics
University of Mosul, Iraq
Email: hisham892020@uomosul.edu.iq

Journal homepage: http://ijeecs.iaescore.com
$d_{k+1}$ is a search direction and it is known as the following formula:

$$
\begin{align*}
    d_1 &= -g_1 & k &= 1 \\
    d_{k+1} &= -g_{k+1} + \beta_k d_k & k &\geq 1
\end{align*}
$$

$g_{k+1}$ is a vector matrix of function $f$, and $\beta_k$ is a CG method parameter.

Below are parameters for some conjugate gradient algorithms:

$\beta_k$ is calculated with the search direction $d_{k+1}$ in the following formulas:

1. $d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T g_k}{g_k^T g_k} d_k$ [1]

2. $d_{k+1} = -g_{k+1} + \frac{y_k^T g_k}{g_k^T g_k} d_k$ [2]

See [3]-[8].

There are three-term conjugate gradient methods for three parameters (FR, PR, HS) proposed by Zhang [9]. These three methods always achieve regression property. Below is the search direction for some three-term conjugated gradient methods:

1) The conjugate gradient method (FR) with three-term is known as:

$$
    d_{k+1} = -g_{k+1} + \beta_{FR} d_k - \theta_k^{(1)} g_{k+1}
$$

Whereas

$$
    \theta_k^{(1)} = \frac{d_k^T g_{k+1}}{g_k^T g_k}
$$

2) The conjugate gradient method (PR) with three-term is known as:

$$
    d_{k+1} = -g_{k+1} + \beta_{PR} d_k - \theta_k^{(2)} y_k
$$

Whereas

$$
    \theta_k^{(2)} = \frac{g_{k+1}^T d_k}{g_k^T g_k}
$$

3) The conjugate gradient method (HS) with three-term is known as:

$$
    d_{k+1} = -g_{k+1} + \beta_{HS} d_k - \theta_k^{(3)} y_k
$$

Whereas

$$
    \theta_k^{(3)} = \frac{g_{k+1}^T d_k}{y_k^{T} d_k}
$$

We notice that these methods always achieve the following property:

$$
    d_k^T g_k = \|g_k\|^2 < 0 \quad \forall k
$$

Here the regression property is achieved with $c = 1$. 

A new parameter in three-term conjugate gradient algorithms for unconstrained… (Alaa Saad Ahmed)
Often the researcher needs either exactly or inexactly line search when studying convergence and applying the CG method. Like the strong Wolf conditions.

The strong Wolf conditions are to find $\alpha_k$

$$f(x_k + \alpha_k d_k) \leq f(x) + \delta \alpha_k \nabla f(x)^T d_k$$  \hspace{1cm} (5)

$$\|d_k^T g_k(x_k + \alpha_k d_k)\| \leq -\sigma \|d_k^T g_k\|$$  \hspace{1cm} (6)

$0 < \delta < \sigma < 1$ are constants according to Li and Weijun [3], [9], [10].

In Section 2, we present the derivation of the new method using the FR conjugate gradient method with three-term. In Section 3, we explain the regression of the new method. In Section 4, we explain the absolute convergence of the new improved algorithm. In Section 5, the numerical results of the proposed algorithm are presented, and the performance of the new improved algorithm is compared with other algorithms in the same field.

2. IMPROVING THE METHOD OF CONJUGATED GRADIENT FR WITH THREE-TERM

$$d_{k+1} = -\lambda g_{k+1} + \beta^F_k d_k - \theta_k g_{k+1}$$

Where $\lambda \in [0,1]$ and using pure conjugacy condition [11]

$$y_k^T d_{k+1} = -\lambda y_k^T g_{k+1} + \beta^F_k y_k^T d_k - \theta_k y_k^T g_{k+1} = 0$$

$$\theta_k y_k^T g_{k+1} = -\lambda y_k^T g_{k+1} + \beta^F_k y_k^T d_k$$

$$\theta_k^{NEW} = -\lambda + \beta^F_k \frac{y_k^T d_k}{y_k^T g_{k+1}}$$  \hspace{1cm} (7)

$$d_{k+1} = -g_{k+1} + \beta^F_k d_k - \theta_k^{NEW} g_{k+1}$$  \hspace{1cm} (8)

Algorithm:
The conjugate gradient method (FR) algorithm with improved three-term:

Step 1: Let $x_0$ is an initial value, put $d_0 = -g_0, \varepsilon > 0, k = 0$.

Step 2: Determine the length of the step $\alpha_k > 0$ achieves the Wolfe conditions (5), (6).

Step 3: Determine $x_{k+1} = x_k + \alpha_k d_k$. If $\|g_{k+1}\| < \varepsilon$ then stopped.

Step 4: Determine $\beta^F_{k+1}, \theta_k$ from (7) and generate direction from (8).

Step 5: Put $k = k + 1$. Go to step 2.

3. REGRESSION PROPERTY OF THE NEW FORMULA

We will mention the proof of the sufficient descent property for the conjugate gradient method (FR) algorithm formula with improved three-term (8). The sufficient descent property for the conjugated gradient algorithm is expressed as:

$$g_{k+1}^T d_{k+1} \leq -c\|g_{k+1}\|^2$$  \hspace{1cm} for $k \geq 0$ and $c > 0$ \hspace{1cm} (9)

Theorem (1)
The search direction (8) with the conjugation coefficient $\beta^F_k$ and the value of $\theta_k$ given by (7) will achieve (9) for all $k \geq 1$ values.

Proof: by mathematical induction

a) When $k = 0$, then $d_0 = -g_0 \rightarrow g_0^T d_0 = -\|g_0\|^2 < 0$

b) Assume that the relation $g_k^T d_k < 0$ is true for each $k$. 

c) We prove that the relation (9) is correct when \( k = k + 1 \) by multiplying both sides of the (8) by \( g_{k+1} \).

We get:

\[
\begin{align*}
    g_{k+1}^T d_{k+1} &= -g_{k+1}^T g_{k+1} + \beta_k^F g_{k+1}^T d_k - \theta_k^{NEW} g_{k+1}^T g_{k+1} \\
    g_{k+1}^T d_{k+1} &= -g_{k+1}^T g_{k+1} (1 + \theta_k^{NEW}) + \beta_k^F g_{k+1}^T d_k
\end{align*}
\]

If \( \theta_k^{NEW} > 0 \) then

\[
\begin{align*}
    g_{k+1}^T d_{k+1} &< -g_{k+1}^T g_{k+1} (1 + \theta_k^{NEW}) + \beta_k^F g_{k+1}^T d_k \\
    g_{k+1}^T d_{k+1} &< 0
\end{align*}
\]

Thus, the regression property of the new method improved is proved.

4. CONVERGENCE OF THE NEW IMPROVED ALGORITHM

In this section, we will show that the three-term CG method with the coefficient of conjugation \( \beta_k^F \) and the value of \( \theta_k \) given by (7) is absolutely convergent. We need the following assumptions to study the convergence of the new proposed algorithm:

**Assumptions (A1)** [10], [12]-[14]

We will impose the following assumptions on the codomain (target) function:

a) Level set \( S = \{ x \in \mathbb{R}^n : f(x) \leq f(x_0) \} \) is a closed and restricted at the initial point.

b) The codomain (target) function is continuous and derivable in some proximity of \( N \) of level set \( s \), and its grades are continuous (lipschitz continuous). This means that there is a constant \( L > 0 \), as that:

\[
\| g(x) - g(y) \| \leq L \| x - y \| \quad \forall x, y \in N
\]

c) The codomain (target) function \( f \) is uniformly convex function, there is a constant number \( g \) that achieves variance, as that:

\[
\left( \nabla f(x) - \nabla f(y) \right)^T (x - y) \geq \mu \| x - y \|^2 \quad , \text{for any } x, y \in S
\]

On the other hand, using assumptions (A1) there is a positive constant \( B \), as that:

\[
\begin{align*}
    \| x \| &\leq B \quad , \forall x \in S \\
    \gamma &\leq \| g(x) \| \leq \overline{\gamma} \quad , \forall x \in S
\end{align*}
\]

**Lemma** [10], [15], [16]

We present assumptions (A1) and (10) are achieve, and by referring to (8) for the conjugate gradient where \( d_k \) is a sloping search direction, and the length of step \( \alpha_k \) is obtained from the strong search line for Wolfe.

If

\[
\sum_{k=1}^{\infty} \frac{1}{\| d_{k+1} \|^2} = \infty
\]

we get

\[
\lim_{k \to \infty} (\inf \| g_k \|) = 0
\]

**Theorem:**

We propose assumptions (A1) and (10) are accomplished by regression condition. The conjugate gradient method with the coefficient of conjugation \( \beta_k^F \) and the value of \( \theta_k \) is given by (7), as if \( \alpha_k \) is fulfilled with two strong wolf conditions (5) and (6). Since the codomain (target) function is uniformly convex at the plane of the set \( S \), then the equation \( \lim_{k \to \infty} \inf \| g_k \| = 0 \) is achieved.
Proof:
\[
\|d_{k+1}\| = \|-g_{k+1} + \beta_k^{FR}d_k - \theta_k^{NEW}g_{k+1}\| \\
\|d_{k+1}\| \leq \|g_{k+1}\| + \beta_k^{FR}\|d_k\| + \theta_k^{NEW}\|g_{k+1}\| \\
\|d_{k+1}\| \leq \|g_{k+1}\|(1 + \theta_k^{NEW}) + \frac{\|g_{k+1}\|^2}{\|g_k\|^2}\|d_k\| \\
\|d_{k+1}\| \leq (1 + \theta_k^{NEW}) + \frac{\|g_{k+1}\|^2}{\|g_k\|^2}\|d_k\| \\
\sum_{k=1}^1 \frac{1}{\|d_{k+1}\|} \geq \left( \frac{1}{(1 + \theta_k^{NEW}) + \frac{\|g_{k+1}\|^2}{\|g_k\|^2}\|d_k\|} \right)^\frac{1}{\gamma^2} \sum_{1} = \infty
\]

using the lemma above

\[
limit_{k \to \infty} \|g_k\| = 0
\]

5. NUMERICAL RESULTS

In this section, we discuss the numerical results of the new improved algorithm that we obtained from using the new formula in (8) for a set of test functions in unrestricted non-linear optimization [17]. To evaluate the performance of this proposed algorithm, the results of (75) functions [18] that were included in this study were chosen to compare with the other classical conjugate gradient method (FR, TTFR), shown in the source [17]. All codes were written using Fortran 77 and MATLAB R2009b. Using a comparison of Dolan and More’ we notice through Figures 1-3 a clear superiority of the new improved algorithm with respect to the number of iterations in Figure 1 and the number of times the function is calculated in Figure 2 and also in terms of the cpu time taken to implement the program in Figure 2 in Dimensions 100, 200, …, 1000 [19]. We also wrote a Table 1 for (22) unrestricted non-linear optimization functions to show the efficiency of the new improved method for numbers of iterations (Iter), and the number of function evaluations (FE) in Dimensions 100, with stop test \(\|g_{k+1}\| < 10^{-6}\). There are other research in the same field but with different test functions. For more see [5], [20]-[29].

| Problems                      | Dim | Iter NEW | Iter FR | Iter TTFR | FE NEW | FE FR | TTFR |
|-------------------------------|-----|----------|---------|-----------|--------|-------|-------|
| Freudenstein & Roth          | 100 | 84       | 1529    | 328       | 1979   | 44092 | 9131  |
| Extended Rosenbrock          | 100 | 35       | 43      | 42        | 73     | 87    | 82    |
| Extended White & Holst       | 100 | 32       | 37      | 35        | 68     | 77    | 70    |
| Extended Beale BEALE         | 100 | 14       | 15      | 15        | 27     | 29    | 28    |
| Perturbed Quadratic          | 100 | 96       | 101     | 100       | 144    | 155   | 153   |
| Raydan 1                     | 100 | 78       | 90      | 86        | 118    | 138   | 133   |
| Diagonal 2                   | 100 | 63       | 64      | 71        | 105    | 105   | 121   |
| Diagonal 3                   | 100 | 190      | 203     | 242       | 2772   | 3075  | 4869  |
| Hager                        | 100 | 27       | 47      | 31        | 44     | 565   | 49    |
| Generalized Tridiagonal 1    | 100 | 23       | 23      | 24        | 45     | 45    | 20    |
| Extended Powell              | 100 | 53       | 71      | 79        | 101    | 136   | 151   |
| Extended Cliff               | 100 | 12       | fail    | 19        | fail   | fail  | 46    |
| Quadratic Diagonal           | 100 | 53       | 54      | 53        | 95     | 95    | 96    |
| Extended Hiebert             | 100 | 78       | 87      | 85        | 170    | 188   | 187   |
| Extended Quadratic Penalty   | 100 | 24       | 29      | 25        | 55     | 61    | 53    |
| BDQRTIC                      | 100 | 310      | 532     | 587       | 6684   | 11664 | 11457 |
| TRIDIA                       | 100 | 348      | 364     | 392       | 542    | 566   | 624   |
| Broyden Tridiagonal          | 100 | 29       | 31      | 31        | 53     | 49    | 49    |
| Tridiagonal Perturbed Quadratic | 100 | 99       | 109     | 105       | 157    | 173   | 167   |
| Extended DENSCHNC            | 100 | 14       | 15      | 18        | 26     | 27    | 31    |
| BIGGSB1                      | 100 | 477      | 617     | 533       | 750    | 985   | 837   |
| Extended Block-Diagonal      | 100 | 14       | 15      | 15        | 24     | 26    | 26    |
6. CONCLUSIONS

We presented in this research a new type of TTCG algorithm to solve the problems of unconstrained optimization, and the proposed algorithm has shown a high efficiency in solving these problems with the least number of iterations and with higher accuracy in reaching the approximate solution of the function.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to my supervisors Prof. Dr. Khalil K. Abbo for this valuable suggestion, encouragement and invaluable remark during writing this paper.
REFERENCES

[1] R. Fletcher and C. M. Reeves, “Function minimization by conjugate gradients,” *Comput. J.*, vol. 7, no. 2, pp. 149-154, 1964, doi: 10.1093/comjnl/7.2.149.

[2] Y. H. Dai and Y. Yuan, “A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property,” *SIAM J. Optim.*, vol. 10, no. 1, pp. 177-182, 1999, doi: 10.1137/S1052623497318992.

[3] H. M. Khudhur and K. K. Abbo, “A New Conjugate Gradient Method for Learning Fuzzy Neural Networks,” *J. Multidiscip. Model. Optim.*, vol. 3, no. 2, pp. 57-69, 2021.

[4] A. Ahmed, “Optimization Methods For Learning Artificial Neural Networks,” University of Mosul, 2018.

[5] Hisham M. Khudhur, Khalil K. Abbo, “A New Type of Conjugate Gradient Technique for Nonlinear Fuzzy Algebraic Equations,” 2020.

[6] M. R. Hestenes and E. Stiefel, “Methods of conjugate gradients for solving linear systems,” vol. 49, no. 1. NBS Washington, DC, 1952.

[7] L. C. W. Dixon, “Conjugate gradient algorithms: quadratic termination without linear searches,” *IMA J. Appl. Math.*, vol. 15, no. 1, pp. 9-18, 1975.

[8] E. Polak and G. Ribiere, “Note sur la convergence de méthodes de directions conjuguées,” *ESAIM Math. Model. Numer. Anal. Mathématique Analy. Numérique*, vol. 3, no. R1, pp. 35-43, 1969.

[9] L. Zhang and W. Zhou, “Two descent hybrid conjugate gradient methods for optimization,” *J. Comput. Appl. Math.*, vol. 216, no. 1, pp. 251-264, 2008.

[10] K. K. ABBO, Y. A. Laylani, and H. M. Khudhur, “A New Spectral Conjugate Gradient Algorithm For Unconstrained Optimization,” *Int. J. Math. Comput. Appl. Res.*, vol. 8, pp. 1-9, 2018, [Online]. Available: www.tjprc.org.

[11] H. N. Jabbar, K. K. Abbo, and H. M. Khudhur, “Four--Term Conjugate Gradient (CG) Method Based on Pure Conjugacy Condition for Unconstrained Optimization,” *kirkuk Univ. J. Sci.*, vol. 13, no. 2, pp. 101-113, 2018.

[12] H. M. Khudhur and K. K. Abbo, “New hybrid of Conjugate Gradient Technique for Solving Fuzzy Nonlinear Equations,” *J. Soft Comput. Artif. Intell.*, vol. 2, no. 1, pp. 1-8, 2021.

[13] K. K. Abbo and H. M. Khudhur, “New A hybrid Hestenes-Stiefel and Dai-Yuan conjugate gradient algorithms for unconstrained optimization,” *Tikrit J. Pure Sci.*, vol. 21, no. 1, pp. 118-123, 2016.

[14] K. K. Abbo and H. M. Khudhur, “New A hybrid conjugate gradient Fletcher-Reeves and Polak-Ribiere algorithm for unconstrained optimization,” *Tikrit J. Pure Sci.*, vol. 21, no. 1, pp. 124-129, 2016.

[15] K. K. Abbo, Y. A. Laylani, and H. M. Khudhur, “Proposed new Scaled conjugate gradient algorithm for Unconstrained Optimization,” *Int. J. Enhanc. Res. Sci. Technol. Eng.*, vol. 5, no. 7, 2016.

[16] Z. M. Abdullah, M. Hameed, M. K. Hisham, and M. A. Khaleel, “Modified new conjugate gradient method for Unconstrained Optimization,” *Tikrit J. Pure Sci.*, vol. 24, no. 5, pp. 86-90, 2019.

[17] N. Andrei, “An unconstrained optimization test functions collection,” *Adv. Model. Optim.*, vol. 10, no. 1, pp. 147–161, 2008.

[18] H. M. Khudhur, “Numerical and analytical study of some descent algorithms to solve unconstrained Optimization problems,” University of Mosul, 2015.

[19] E. D. Dolan and J. J. Moré, “Benchmarking optimization software with performance profiles,” *Math. Program.*, vol. 91, no. 2, pp. 201-213, 2002.

[20] B. A. Hassan, H. O. Dahawi, and A. S. Younus, “A new kind of parameter conjugate gradient for unconstrained optimization,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 17, no. 1, 2019, doi: 10.11591/ijeeecs.v17.i1.pp404-411.

[21] B. A. Hassan and H. K. Khalo, “A new class of BFGS updating formula based on the new quasi-newton equation,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 13, no. 3, 2019, doi: 10.11591/ijeeecs.v13.i3.pp945-953.

[22] B. A. Hassan, “A new type of quasi-newton updating formulas based on the new quasi-newton equation,” *Numer. Algebr. Control Optim.*, vol. 10, no. 2, 2020, doi: 10.3934/naco.2019049.

[23] B. A. Hassan, Z. M. Abdullah, and H. N. Jabbar, “A descent extension of the Dai - Yuan conjugate gradient technique,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 16, no. 2, 2019, doi: 10.11591/ijeeecs.v16.i2.pp661-668.

[24] B. A. Hassan, A. O. Owaid, and Z. T. Yasen, “A variant of hybrid conjugate gradient methods based on the convex combination for optimization,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 20, no. 2, 2020, doi: 10.11591/ijeeecs.v20.i2.pp1007-1015.

[25] B. A. Hassan and M. W. Taha, “A modified quasi-Newton equation in the quasi-Newton methods for optimization,” *Appl. Math. Sci.*, vol. 13, no. 10, 2019, doi: 10.12988/ams.2019.9351.

[26] B. A. Hassan, “A new formula for conjugate parameter computation based on the quadratic model,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 13, no. 3, 2019, doi: 10.11591/ijeeecs.v13.i3.pp954-961.

[27] B. A. Hassan and M. W. Taha, “A new variants of quasi-newton equation based on the quadratic function for unconstrained optimization,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 2, 2020, doi: 10.11591/ijeeecs.v19.i2.pp701-708.

[28] H. M. Khudhur, “Modified Barzilai-Borwein Method for Steepest Descent Method to Solving Fuzzy Optimization Problems (FOP),” *Albahir J.*, vol. 12, pp. 23-24, 2020.

[29] K. K Abbo and F. H Mohammed, “Spectral Fletcher-Reeves Algorithm for Solving Non-Linear Unconstrained Optimization Problems,” *Iraqi J. Stat. Sci.*, vol. 11, no. 19, pp. 21-38, 2011.