Available online at sjuoz.uoz.edu.krd

journals.uoz.edu.krd

Science Journal of University of Zakho

Vol. 11, No. 4, pp.544-547, Octuber–December -2023

p-ISSN: 2410-7549
e-ISSN: 2414-6943

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.2023.11.4.1152

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

$$\text{Min}\, f(x), x \in R^n \tag{1.1}$$

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

$$x_{k+1} = x_k + \alpha_k d_k \tag{1.2}$$

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 as 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:

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \tag{1.3}$$

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \tag{1.4}$$

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k} \tag{1.5}$$

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)} \tag{1.6}$$

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

$$d_{k+1} = -g_{k+1} + \beta_k d_k \tag{1.7}$$

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.

* Corresponding author

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 is a powerful variation of the conjugate gradient method that is especially useful for solving symmetric, positive-definite linear systems of equations. Numerous disciplines, including engineering, physics, computer science, and finance, use the method extensively.

The three-term CG methods 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

$$d_{k+1}^{ZPRP} = -g_{k+1} + \frac{g_{k+1}^T y_k}{g_{k1}^T g_k} d_k - \frac{g_{k+1}^T d_k}{g_k^T g_k} y_k \qquad (1.8)$$

$$d_{k+1}^{ZHS} = -g_{k+1} + \frac{g_{k+1}^T y_k}{d_{k1}^T y_k} d_k - \frac{g_{k+1}^T d_k}{d_{k1}^T y_k} y_k \qquad (1.9)$$

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:

$$d_{k+1}^{NTT-CG} = -g_{k+1} + \beta_k d_k - t_k (\frac{g_{k+1}^T d_k}{d_{k1}^T y_k}) y_k \qquad (1.10)$$

where $t_k = \gamma \frac{\|y_k\|}{\|v_k\|} + (1-\gamma)\frac{v_k^T y_k}{\|v_k\|^2}$, and the parameter β_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 $\bar{g}_{k+1}$ that is defined below

$$\bar{g}_{k+1} = g_{k+1} + (1-\delta)((\frac{g_{k+1}}{\gamma}) - \mu g_{k+1} \qquad (2.1)$$

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

More specifically, the search direction of our method, named as (New1 TT-CG) method, defined by:

$$d_{k+1}^{New} = -g_{k+1} + \frac{g_{k+1}^T y_k}{g_k^T g_k} d_k - \frac{\bar{g}_{k+1}^T d_k}{g_k^T g_k} y_k \qquad (2.2)$$

Where $\bar{g}_{k+1} = g_{k+1} + (1-\delta)((\frac{g_{k+1}}{\gamma}) - \mu g_{k+1}$

Or $\quad d_{k+1}^{New} = -g_{k+1} + \frac{g_{k+1}^T y_k}{g_k^T g_k} d_k - (1 + (1-\delta)((\frac{1}{\gamma}) - \mu)\frac{g_{k+1}^T d_k}{g_k^T g_k} y_k \qquad (2.3)$

Where, $\delta \in (0,1)$, $\mu = 0.1$ and $\gamma = \frac{2\sqrt{\omega}}{\|v_k\|}(1 + \|x_{k+1}\|)$ and ω 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 : Determine α_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 [(2.3)].

Step 7 : If $\|g_{k+1}\|^2 \leq \frac{|g_k^T g_{k+1}|}{0.2}$ is satisfied go to step 2,

Otherwise, set $k = k + 1$ and go to step 3.

3. GLOBAL CONVERGENCE AND (DESCENT AND SUFFICIENT DESCENT PROPERTYS) 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}^{NTT-CG} satisfy the descent condition.

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

$$d_{k+1}^{New^T} g_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{g_k^T g_k} d_k^T g_{k+1} - \frac{\bar{g}_{k+1}^T d_k}{g_k^T g_k} g_{k+1}^T y_k$$

(3.1)

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

By mathematical induction, from the first search direction, we get $d_1^T g_1 = -\|g_1\|^2 \leq 0$, and we assume that it is true for case k that is mean

$d_k^T g_k \leq 0$. 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.

$$-\frac{\bar{g}_{k+1}^T d_k}{g_k^T g_k} g_{k+1}^T y_k = -(\frac{d_k^T(g_{k+1} + (1-\delta)((\frac{g_{k+1}}{\gamma}) - \mu g_{k+1})}{g_k^T g_k} g_{k+1}^T y_k$$

from Lipschitz Condition $\|y_k\| \leq L\|v_k\|$ and

$g_{k+1}^T y_k \leq L g_{k+1}^T d_k$ where $L > 0$ (3.2)

Then,

$$d_{k+1}^{New^T} g_{k+1} \leq -\frac{\bar{g}_{k+1}^T d_k}{g_k^T g_k} d_k^T g_{k+1}$$

$$= -\left(1 + \left(\frac{(1-\delta)}{\gamma}\right) - (1-\delta)\mu\right) L \frac{(d_k^T g_{k+1})^2}{g_k^T g_k} < 0 \qquad (3.3)$$

Finally, we have $d_{k+1}^{New1^T} g_{k+1} \leq 0$. ■

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 (31) and (3.3), we have

$$d_{k+1}^{New^T} g_{k+1} \leq= -\left(1 + \left(\frac{(1-\delta)}{\gamma}\right) - (1-\delta)\mu\right) L \frac{(d_k^T g_{k+1})^2}{g_k^T g_k} * \frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2}$$

Let $C = \left(1 + \left(\frac{(1-\delta)}{\gamma}\right) - (1-\delta)\mu\right) L \frac{\left(d_k^T g_{k+1}\right)^2}{g_k^T g_k \|g_{k+1}\|^2}$

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

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

$$\|x\| \leq B, \forall\, x \in S \tag{3.4}$$

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

$$\|g(x) - g(x_k)\| \leq L\|x - x_k\|, \forall\; x, x_k \in \Omega \tag{3.5}$$

From Assumptions (3.1) and (3.2), $\exists\, M > 0$ such that

$$\|g(x)\| \leq M, \quad \forall\; x \in S. \tag{3.6}$$

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 α_k is determined by strong Wolfe conditions and by assuming that Assumptions (3.5) and (3.5) are true. If

$$\sum_{k\geq 1} \frac{1}{\|d_k\|^2} = \infty \tag{3.7}$$

The $\lim_{k\to\infty} \inf \|g_k\| = 0.$ (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 α_k satisfies the strong Wolfe line search conditions, then

$$\lim_{k\to\infty} \inf \|g_k\| = 0$$

Proof: From equation (3.2), we have

$$\left\|d_{k+1}^{New}\right\| \leq \|g_{k+1}\| + \left|\frac{g_{k+1}^T y_k}{g_k^T g_k}\right| \| d_k\| + \left|\frac{\bar{g}_{k+1}^T d_k}{g_k^T g_k}\right| \| y_k\|, \tag{3.9}$$

by using (3.6) and (3.2)

$$\left\|d_{k+1}^{New}\right\| \leq M + \left|\frac{L g_{k+1}^T d_k}{g_k^T g_k}\right| \| d_k\| + (1 + (1-\delta)(\left(\frac{1}{\gamma}\right) - \mu 1) \left|\frac{g_{k+1}^T d_k}{g_k^T g_k}\right| L\|v_k\| \tag{3.10}$$

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

$$\left\|d_{k+1}^{New}\right\| \leq M + LM + (1 + (1-\delta)\left(\left(\frac{1}{\gamma}\right) - \mu 1\right) LM\alpha_k = \beta$$

$$\Rightarrow \sum_{k\geq 1} \frac{1}{\left\|d_{k+1}^{New}\right\|^2} \geq \sum_{k\geq 1} \frac{1}{\beta^2} = \infty$$

$$\Rightarrow \sum_{k\geq 1} \frac{1}{\left\|d_{k+1}^{New}\right\|^2} = \infty$$

Now, by using lemma (3.1), we get $\lim_{k\to\infty} \inf \|g_k\| = 0.$ ■

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

Test function.	DIM.	PR		New TT-CG	
		NOI	NOF	NOI	NOF
G-Cantrel	4	22	159	21	149
	10	22	159	21	149
	100	22	159	21	149
	500	23	171	22	161
	1000	23	171	22	161
	5000	30	270	23	175
G-Wolfe	4	11	24	16	33
	10	32	65	34	76
	100	49	99	46	98
	500	58	117	45	97
	1000	64	129	55	120
	5000	99	214	58	127
Cubic	4	15	45	14	39
	10	16	47	15	43
	100	16	47	15	43
	500	16	47	15	43
	1000	16	47	15	43
	5000	16	47	16	45
Miele	4	37	116	39	122
	10	37	116	40	124
	100	44	148	40	124
	500	44	148	43	143
	1000	50	180	46	160
	5000	50	180	46	160
G-Wood	4	29	67	26	61
	10	29	67	26	61
	100	30	69	26	61
	500	30	69	26	61
	1000	30	69	26	61
	5000	30	69	27	63
Rosen	4	30	85	28	66
	10	30	85	28	66
	100	30	85	28	66
	500	30	85	28	66
	1000	30	85	28	66
	5000	30	85	28	66
Total		1170	3825	1053	3348

Table 3.2: -The Percentage Of Improving Of New TT-CG Algorithm

	PR	New TT-CG
NOI	100%	90 %
NOF	100%	89.524 %

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

Fletcher, R. and Reeves, C.M., Function minimization by conjugate gradients. The Computer Journal. 7, 149-154, 1964.

Hestenes, M. R. and Stiefel, E., Methods of conjugate gradients for solving linear systems. Journal of Research of the National Bureau of Standards. 49, 409-436, 1952.

Polak, E. and Ribiere, G., Note surla convergence des méthodes de directions conjuguées., 3(16), 35-43, 1969.

Dai, Y. H. and Yuan, Y., A nonlinear conjugate gradient method with a strong global convergence property, SIAM Journal on Optimization, 10, 177-182,1999.

A. L. Ibrahim and S. G. Shareef, Modified Conjugate Gradient Method For Training Neural Networks Based On Logistic Mapping, Journal of Duhok University, vol. 22, no. 1, 45–51, 2019. https://doi.org/10.26682/sjuod.2019.22.1.7

H. Dlovan, A new Suggested Conjugate Gradient Algorithm with Logistic Mapping, Journal of University of Zakho, 4(2), 2016, 244-247, https://doi.org/10.25271/2016.4.2.113. 12), 733–757.

Jie Guo, Zhong Wan. A new three-term conjugate gradient algorithm with modified gradient-differences for solving unconstrained optimization problems[J]. AIMS Mathematics, 2023, 8(2): 2473-2488. doi: 10.3934/math.2023128

Sulaiman, I.M., Malik, M., Awwal, A.M. et al. On three-term conjugate gradient method for optimization problems with applications on COVID-19 model and robotic motion control. Adv Cont Discr Mod 2022, 1 (2022). https://doi.org/10.1186/s13662-021-03638-9

A. L. Ibrahim and M. G. Mohammed, "A new three-term conjugate gradient method for training neural networks with global convergence" Indonesian Journal of Electrical Engineering and Computer Science, Vol. 28, No. 1, October 2022, pp. 547~554. DOI: http://doi.org/10.11591/ijeecs.v2 8.i1.pp551-558

L. Zhang, W. Zhou & D-H. Li, A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence, IMA J. Numer. Anal, 26(4), 629-640,2006, https://doi.org/10.1093/imanum/drl016

M. R. Hestenes and E. Stiefel, Methods for conjugate gradients for solving linear systems, Journal of Research of the National Bureau of Standards, vol. 49, no. 6, pp. 409–436, 1952.

A. L., Ibrahim, & S. G., Shareef, A new class of three-term conjugate gradient methods for solving unconstrained minimization problems. General Letters in Mathematics Vol 7 (2) 79-86, 2019.
https://doi.org/10.31559/glm2019.7.2.4

Jusoh, M. Mamat and M. Rivaie, A new edition of conjugate gradient methods for large-scale unconstrained optimization, International Journal of Mathematical Analysis, Vol. 8, No. 46, (2014), 2277 – 2291.

K. Sugiki, Y. Narushima, and H. Yabe, Globally convergent three–term conjugate gradient methods that use secant conditions and generate descent search directions for unconstrained optimization, J. Optim. Theory Appl. 153, (20