Mathematics and Statistics 11(4): 640-645, 2023 http://www.hrpub.org
DOI: 10.13189/ms.2023.110404

# A New Conjugate Gradient Algorithm for Minimization Problems Based on the Modified Conjugacy Condition

**Dlovan Haji Omar***, **Salah Gazi Shareef, Bayda Ghanim Fathi**

Department of Mathematics, Faculty of Science, University of Zakho, Iraq

*Received January 18, 2023; Revised April 14, 2023; Accepted June 9, 2023*

***Cite This Paper in the Following Citation Styles***

***(a):*** *[1] Dlovan Haji Omar, Salah Gazi Shareef, Bayda Ghanim Fathi , "A New Conjugate Gradient Algorithm for Minimization Problems Based on the Modified Conjugacy Condition," Mathematics and Statistics, Vol. 11, No. 4, pp. 640 - 645, 2023. DOI: 10.13189/ms.2023.110404.*

***(b):*** *Dlovan Haji Omar, Salah Gazi Shareef, Bayda Ghanim Fathi (2023). A New Conjugate Gradient Algorithm for Minimization Problems Based on the Modified Conjugacy Condition. Mathematics and Statistics, 11(4), 640 - 645. DOI: 10.13189/ms.2023.110404.*



**Abstract** Optimization refers to the process of finding the best possible solution to a problem within a given set of constraints. It involves maximizing or minimizing a specific objective function while adhering to specific constraints. Optimization is used in various fields, including mathematics, engineering, economics, computer science, and data science, among others. The objective function can be a simple equation, a complex algorithm, or a mathematical model that describes a system or process. There are various optimization techniques available, including linear programming, nonlinear programming, genetic algorithms, simulated annealing, and particle swarm optimization, among others. These techniques use different algorithms to search for the optimal solution to a problem. In this paper, the main goal of unconstrained optimization is to minimize an objective function that uses real variables and has no value restrictions. In this study, based on the modified conjugacy condition, we offer a new conjugate gradient (CG) approach for nonlinear unconstrained problems in optimization. The new method satisfied the descent condition and the sufficient descent condition. We compare the numerical results of the new method with the Hestenes-Stiefel (HS) method. Our novel method is quite effective according to the number of iterations (NOI) and the number of functions (NOF) evaluated, as demonstrated by the numerical results on certain well-known non-linear test functions.

**Keywords** Optimization Problems, Conjugate Gradient Algorithm, Descent Condition, Global Convergence

## 1. Introduction

Suppose the optimization problem as an example

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

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

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

Or $x_{k+1} = x_k + v_k$ where $v_k = \alpha_k d_k = x_{k+1} - x_k$

Here, a line search is used to determine the step size $\alpha_k$, and the search direction is denoted by $d_k$. The steepest descent direction, $d_1 = -g_1$, is often the search direction in the first iteration. The other search directions are follows:

$$d_{k+1} = -g_{k+1} + \beta_k d_k \tag{1.3}$$

where $g_k$ is a gradient of $f$ at $x_k$ and $\beta_k$ is a parameter known as a conjugate parameter. Different conjugate gradient methods have been developed as a result of the literature presenting several choices for $\beta_k$. Some classic conjugate parameters are Fletcher-Reeves (FR) method [1], the Hestenes-Stiefel (HS) method [2], and the Polak-Rivi ère (PR) method [3], Dai and Yuan (DY) [4], which are respectively determined as follows:

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \tag{1.4}$$

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \quad (1.5)$$

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k} \quad (1.6)$$

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T(g_{k+1}-g_k)} \quad (1.7)$$

Where $y_k = g_{k+1} - g_k$. The Euclidean norm of vectors is represented by the letter $\|.\|$. Global convergence properties of conjugate gradient methods are those that have received the most attention. Several writers, including Gilbert and Nocedal [5] and Hestenes and Stiefel [2], have investigated the convergence of conjugate gradient algorithms under various line searches. Several studies have also been conducted to enhance the conjugate gradient method's parameters, including ([6] and [7]):

$$\beta_k = \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k} - \mu\left(\frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}\right)^2 \quad (1.8)$$

$$\beta_k = \mu \frac{\|g_{k+1}\|^2}{-d_k^T g_k}\left(1 + \left(k \frac{d_k^T g_{k+1}}{\|y_k\|^2}\right)\left(\frac{\|g_{k+1}\|^2}{d_k^T g_k}\right)\right), \quad (1.9)$$

The structure of this essay is as follows: We'll introduce a novel conjugate gradient method suggestion in Section 2. In section 3, we present proof of the descent and sufficient descent conditions of the new method. Many numerical tests of a new conjugate gradient technique are presented in Section 4. Our last remarks are presented in Section 5.

## 2. Derivation the New Method

In this part, we will derive our new method. Consider the vector below

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

where, $\delta \in (0,1)$, $\mu = 0.1$ and $\gamma = \frac{2\sqrt{\omega}}{\|v_k\|}(1 + \|x_{k+1}\|)$ and $\omega$ is the machine error. Now, suppose that the modified conjugacy condition is

$$d_{k+1}^T y_k = -\bar{g}_{k+1}^T v_k \quad (2.2)$$

and the search direction is given by

$$d_{k+1} = -g_{k+1} + \beta_k^{new} d_k \quad (2.3)$$

multiply both sides by $y_k$, to get

$$d_{k+1} y_k = -g_{k+1} y_k + \beta_k^{new} d_k y_k \quad (2.4)$$

then,

$$-\bar{g}_{k+1}^T v_k = -g_{k+1} y_k + \beta_k^{new} d_k y_k \quad (2.5)$$

here, we get

$$\beta_k^{new} = \frac{g_{k+1}^T y_k - \bar{g}_{k+1}^T v_k}{d_k^T y_k} \quad (2.6)$$

## 3. Algorithm of A New C. G. Method

1. select $x_1$ and $\varepsilon = 10^{-5}$.
2. set $d_1 = -g_1$, $g_k = \nabla f(x_k)$, set $k = 1$
3. compute the step length $\alpha_k > 0$ satisfying the Wolf line search $f(x_k + \alpha_k d_k) - f(x_k) \le c_1 \alpha_k g_k^T d_k$

   $|g_{k+1}^T d_k| \le c_2 |g_k^T d_k|$, where , $0 < c_1 < c_2 < 1$.
4. Calculate $x_{k+1} = x_k + \alpha_k d_k$

   $g_{k+1} = \nabla f(x_{k+1})$, I $f \|g_{k+1}\| \le \varepsilon$,

then stop.

5. Calculate $\beta_k^{new}$ by

$$\beta_k^{new} = \frac{g_{k+1}^T y_k - \bar{g}_{k+1}^T v_k}{d_k^T y_k}$$

6. Compute $d_{k+1} = -g_{k+1} + \beta_k^{new} d_k$
7. If $|g_{k+1}^T g_k| > 0.2\|g_{k+1}\|^2$ then return to step 2.

   Else

   Set $k = k + 1$ and return to step 3.

## 4. The Descent and Sufficient Descent Property of the New Method

**Theorem 1:** The direction in (1.3) with the new CG (2.6) satisfies the descent requirement, i.e. $d_{k+1}^T g_{k+1} \le 0$, with both an exact and an inexact line searches, if the sequence $\{x_k\}$ is created by (1.2).

**Proof:** The equation $d_{k+1} = -g_{k+1} + \left(\frac{g_{k+1}^T y_k - \bar{g}_{k+1}^T v_k}{d_k^T y_k}\right) d_k$
Implies that

$$d_{k+1} = -g_{k+1} + \left(\frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{\bar{g}_{k+1}^T v_k}{d_k^T y_k}\right) d_k \quad (3.1)$$

Multiplying the above equations from the right sides by $g_{k+1}$, we obtain

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T g_{k+1} - \frac{\bar{g}_{k+1}^T v_k}{d_k^T y_k} d_k^T g_{k+1} \quad (3.2)$$

If an exact line search is used to determine the step size which is $d_k^T g_{k+1} = 0$, then we get $d_{k+1}^T g_{k+1} \le 0$.

If we have an inexact line search which is $d_k^T g_{k+1} \ne 0$.

Using mathematical induction, from the first search direction, we get $d_1^T g_1 = -\|g_1\|^2 \le 0$. We assume that it is true for case $k$, that is mean $d_k^T g_k \le 0$. To prove case $k+1$.

Because the HS parameter satisfies the descent requirement, the first two terms of the above equation are less than or equal to zero. To complete the proof, we only need proof that the third term is

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

Since, $\delta \in (0,1)$, $\mu = 0.1$ and $\gamma = \frac{2\sqrt{\omega}}{\|v_k\|}(1 + \|x_{k+1}\|)$

$$0 < \|v_k\| = \|x_{k+1} - x_k\| \leq \|x_{k+1}\| \Rightarrow 0 \leq \frac{\|v_k\|}{2\sqrt{\omega}1 + \|x_{k+1}\|)} \leq 1$$

So, we get

$$-\frac{\bar{g}_{k+1}^T v_k}{d_k^T y_k} d_k^T g_{k+1} = -\left(1 + \left(\frac{(1-\delta)}{\gamma}\right) - (1-\delta)\mu\right) \frac{\alpha_k (d_k^T g_{k+1})^2}{d_k^T y_k} < 0$$

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

**Theorem 2:** suppose that $\{x_k\}$ is generated by (1.2), then the direction in (1.3) with the new CG (2.2) satisfies the sufficient descent requirement, i.e. $d_{k+1}^T g_{k+1} \leq -C\|g_{k+1}\|^2$.

**Proof:** Consider the equation

$$d_{k+1} = -g_{k+1} + (\frac{g_{k+1}^T y_k - \bar{g}_{k+1}^T v_k}{d_k^T y_k}) d_k \tag{3.3}$$

Implies that

$$d_{k+1} = -g_{k+1} + (\frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{\bar{g}_{k+1}^T v_k}{d_k^T y_k}) d_k \tag{3.4}$$

By multiplying both sides of above equation by $g_{k+1}$ from right, we obtain

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T g_{k+1} - \frac{\bar{g}_{k+1}^T v_k}{d_k^T y_k} d_k^T g_{k+1} \tag{3.5}$$

Since the parameter of HS is to satisfy the descent property, then we get the following inequality

$$\begin{aligned} d_{k+1}^T g_{k+1} &\leq -\frac{\bar{g}_{k+1}^T v_k}{d_k^T y_k} d_k^T g_{k+1} \\ &= -(\frac{v_k^T (g_{k+1} + (1-\delta)((\frac{g_{k+1}}{\gamma}) - \mu g_{k+1})}{d_k^T y_k} d_k^T g_{k+1} \\ &\leq -\left(1 + \left(\frac{(1-\delta)}{\gamma}\right) - (1-\delta)\mu\right) \frac{\alpha_k (d_k^T g_{k+1})^2}{d_k^T y_k} * \frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2} \end{aligned} \tag{3.6}$$

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

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

Now, to show the global convergence of our new conjugate gradient method, the following presumptions are necessary [8].

**Assumption 1.** The level set $S = \{x : x \in R^n, f(x) \leq f(x_0)\}$ is bounded. i.e. $\exists\ B > 0$ such that

$$\|x\| \leq B, \ \forall \ x \in S \tag{3.7}$$

**Assumption 2.** In a neighborhood $\Omega \in S$, $E$ is differentiable and its gradient $g$ is Lipschitz continuous, i.e. $\exists\ L > 0$ such that

$$\|g(w) - g(w_k)\| \leq L\|w - w_k\|, \forall \quad w, w_k \in \Omega \tag{3.8}$$

From Assumptions 1 and 2, $\exists\ M > 0$ such that

$$\|g(w)\| \leq M, \qquad \forall \quad w \in S. \tag{3.9}$$

**Lemma 1 [9].** Suppose that the above Assumptions true and the $\{x_k\}$ is a sequence constructed by equations (1.2) and (1.3), where $d_k$ satisfies the descent property and the step size is computed by strong Wolfe conditions. If

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} = \infty \tag{3.10}$$

Then

$$\lim_{k\to\infty} inf\|g_k\| = 0. \tag{3.11}$$

**Theorem 3.** Assume that the above Assumptions hold. If the iteration is computed from equations (1.2) and (1.3) where $\beta_k^{new}$ is defined in (2.6), and $\alpha_k$ satisfies the strong Wolfe line search conditions, then

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

**Proof:** From equations (1.3) and (2.6), we have

$$\|d_{k+1}\| \le \|g_{k+1}\| + \left|\frac{g_{k+1}^T y_k - \bar{g}_{k+1}^T v_k}{d_k^T y_k}\right| \| \ d_k\|, \tag{3.12}$$

$$\|d_{k+1}\| \le \|g_{k+1}\| + \left(\left|\frac{g_{k+1}^T y_k}{d_k^T y_k}\right| + (1 + \frac{(1-\delta)}{\gamma} - \mu)\left|\frac{g_{k+1}^T v_k}{d_k^T y_k}\right|\right)\| \ d_k\|$$

$$\|d_{k+1}\| \le \|g_{k+1}\| + \left(\left|\frac{g_{k+1}^T y_k}{d_k^T y_k}\right| + (1 + \frac{(1-\delta)}{\gamma} - \mu)\left|\frac{g_{k+1}^T v_k}{d_k^T y_k}\right|\right)\| \ d_k\|$$

Since $g_{k+1}^T v_k \le \alpha_k d_k^T y_k$ and by using (3.9), also from Lipschitz Condition $\|y_k\| \le L\|v_k\|$ and $g_{k+1}^T y_k \le L g_{k+1}^T d_k$ where $L > 0,$ we have

$$\|d_{k+1}\| \le M + \left(L + ((1 + \frac{(1-\delta)}{\gamma} - \mu)\alpha_k \right)\| \ d_k\| \tag{3.13}$$

since, $\|v_k\| = \|x - x_k\|$, and $D = max\{\|x - x_k\|, \forall \ \ x, x_k \in R\}$.

It follows that

$$\|d_{k+1}\| \le M + \left(L + ((1 + \frac{(1-\delta)}{\gamma} - \mu)\alpha_k \right) D\alpha_k = \beta$$

this implies that $\sum_{k\ge1} \frac{1}{\|d_{k+1}\|^2} \ge \sum_{k\ge1} \frac{1}{\beta^2} = \infty$

and so, $\sum_{k\ge1} \frac{1}{\|d_{k+1}\|^2} = \infty$

By using lemma (1), we get $\lim_{k\to\infty} inf\|g_k\| = 0$, which completes the proof.

## 5. Numerical Results

This section is designed to evaluate how effectively the new CG method has been used. We conduct a number of computational experiments on a concatenation of test problems for unconstrained nonlinear optimization derived from [10]. To demonstrate the application and efficacy of the suggested approach with various dimensions of $4 \le n \le 5000$, all programs are written in the FORTRAN95 language. The performance of the new CG method and the original CG method using NOI and NOF for 7 different non-linear functions is compared in Table 1. Table 2 shows that the rate of improvement for the modified CG technique illustrates how it outperforms the standard method in terms of NOI and NOF.

**Table 1.** Comparison Between Two Algorithms (HS. and New CG)

| Test function. | DIM. | HS | | New. | |
|---|---|---|---|---|---|
| | | NOI | NOF | NOI | NOF |
| G-Cantrel | 4 | 22 | 159 | 18 | 113 |
| | 10 | 22 | 159 | 18 | 113 |
| | 100 | 22 | 159 | 19 | 129 |
| | 500 | 23 | 171 | 21 | 144 |
| | 1000 | 23 | 171 | 21 | 144 |
| | 5000 | 28 | 248 | 32 | 255 |
| G-Wolfe | 4 | 11 | 24 | 11 | 25 |
| | 10 | 32 | 65 | 31 | 63 |
| | 100 | 49 | 99 | 48 | 101 |
| | 500 | 52 | 105 | 57 | 127 |
| | 1000 | 70 | 141 | 57 | 119 |
| | 5000 | 165 | 348 | 62 | 139 |
| Powell | 4 | 37 | 102 | 34 | 95 |
| | 10 | 37 | 102 | 34 | 95 |
| | 100 | 40 | 117 | 34 | 95 |
| | 500 | 44 | 136 | 34 | 95 |
| | 1000 | 44 | 136 | 34 | 95 |
| | 5000 | 44 | 136 | 37 | 112 |
| Rosen | 4 | 30 | 83 | 30 | 83 |
| | 10 | 30 | 83 | 30 | 83 |
| | 100 | 30 | 83 | 30 | 83 |
| | 500 | 30 | 83 | 30 | 83 |
| | 1000 | 30 | 83 | 30 | 83 |
| | 5000 | 30 | 83 | 30 | 83 |
| Miele | 4 | 28 | 85 | 28 | 84 |
| | 10 | 31 | 102 | 28 | 84 |
| | 100 | 33 | 114 | 34 | 118 |
| | 500 | 40 | 146 | 35 | 120 |
| | 1000 | 46 | 176 | 41 | 152 |
| | 5000 | 54 | 211 | 48 | 188 |
| G-Wood | 4 | 30 | 68 | 26 | 61 |
| | 10 | 30 | 68 | 26 | 61 |
| | 100 | 30 | 68 | 26 | 61 |
| | 500 | 30 | 68 | 26 | 61 |
| | 1000 | 30 | 68 | 27 | 63 |
| | 5000 | 30 | 68 | 27 | 63 |
| OSP. | 4 | 8 | 45 | 8 | 49 |
| | 10 | 13 | 58 | 12 | 46 |
| | 100 | 49 | 185 | 51 | 194 |
| | 500 | 112 | 353 | 107 | 327 |
| | 1000 | 156 | 475 | 156 | 473 |
| | 5000 | 256 | 774 | 256 | 774 |
| Total | | 1951 | 6208 | 1744 | 5536 |

**Table 2.** The Percentage of Improving of New Algorithm

| Tools | HS | New CG |
|---|---|---|
| NOI | 100% | 89.3906 % |
| NOF | 100% | 89.1753 % |

The suggested strategy raises NOI and NOF by 10.6094% and 10.8247%, respectively. The modified New CG method has generally outperformed the standard HS approach by a factor of 10.71705%.

# 6. Conclusions

For unconstrained optimization issues, we proposed a new conjugate gradient algorithm in this paper. The decent condition and sufficient decent condition, which are essential steps of the proofs with an exact line search and an inexact line search. Based on the numerical results, it can be clearly observed that the new algorithm is more efficient in terms of the number of iteration and the number of function evaluation.

# REFERENCES

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

[2] 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.

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

[4] 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.

[5] J. C. Gilbert, J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM J. Optim. 2, 21-42, 1992.

[6] 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

[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.

[8] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning internal representations by error propagation,” in Parallel Distributed Processing: Explorations in the Microstructure of Cognition, D. Rumelhart and J.McClelland, Eds., 318–362, Cambridge, Mass, USA, 1986.

[9] R. Fletcher and C. M. Reeves, “Function minimization by conjugate gradients,” Computer Journal, vol. 7, 149–154, 1964.

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