African Journal of Advanced Pure and Applied Sciences (AJAPAS)

Online ISSN: 2957-644X

Volume 2, Issue 2, April-June 2023, Page No: 148-154

Website: *https://aaasjournals.com/index.php/ajapas/index*

Development of a New Conjugate Gradient Algorithm for Solving an Unconstrained Nonlinear Optimization Problem

Suaad Madhat Abdullah *

College of Sciences, University of Kirkuk, Kirkuk, Iraq

Corresponding author: suaad_m@uokirkuk.edu.iq

Received: March 13, 2023	Accepted: April 24, 2023	Published: May 06, 2023

Abstract:

This article develops a new conjugate gradient algorithm in a scaled conjugate gradient field. The proposal depends on the following algorithms: Quasi-Newton and classical conjugate gradient. Under certain assumptions, the developed algorithm satisfies the descent direction and global convergence property. Additionally, the hybrid scaled gradient algorithm is involved in the new direction. Compared to the classical algorithm, the numerical outcomes demonstrate the superiority of our algorithm in tackling unconstrained nonlinear optimization problems.

Keywords: Optimization, Conjugate Gradient, Descent, Algorithm.

Cite this article as: S. M. Abdullah, "Development of a New Conjugate Gradient Algorithm for Solving an Unconstrained Nonlinear Optimization Problem," *African Journal of Advanced Pure and Applied Sciences (AJAPAS)*, vol. 2, no. 2, pp. 148–154, April-June 2023.

Introduction

In 1988, Barzilai and Borwein [1] developed the spectral conjugate gradient. The algorithm is often used to solve unimpeded optimization problems. In the late 90s, the algorithm was developed by Raydan [2], [3] to solve the problems of large ruler unimpeded optimization. It is utilized in each line exploration worldwide converge; therefore, the Gradient direction is crucial and held by a non-monotone plan. General unimpeded optimization is formulated under the following:

$$f(x), \ \ x \in R^n \tag{1}$$

The problem is explained iteratively and started from an initial point $x_0 \in R^n$. The conjugate gradient algorithm, according to the relapse formulation, follows as under.

$$x_{k+1} = x_k + a_k d_k, \quad k = 0, 1, 2, ... \tag{2}$$

where x_k denotes the current iteration, the stepsize is $x_k > 0$ where some line exploration process is in the calculation. In the literature, Wolfe [4], [5], Armijo [6] and Goldstein [7] are widely utilized line explorations. In the line explorations, determining stepsize is the change between exact and imprecise. For an exact line search, a_k can be calculated using its rule. On the other hand, for elementary imprecise line exploration, a_k is projected. Additionally, at a minimal cost, it attains an adequate reduction in f. Using Strong Wolfe, we fixated the exploration. The conditions that define Wolfe line search [8] are formulated as follows.

$$f(X_k + a_k d_k) \leq f(X_k) + \mu a_k g_k^T d_k \tag{3}$$

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k \tag{4}$$

Nocedal and Wright [9] are used to build and book σ and μ for this search, and $0 < \mu < \sigma < 1$. Condition (4) is known as the Strong Wolfe line search after replacing it with the following condition.

$$\left|g_{k+1}^T d_k\right| \geq -\sigma g_k^T d_k \tag{5}$$

For d_k, the basic search direction is formulated as follows:

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \qquad d_1 = -g_1 \tag{6}$$

$g_k = f(X_k)$

where different conjugate gradient methods are determined by the coefficient $\beta_k \in R$ with a little common β_{k+1}.

$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}, \beta_k^{HS} = -\frac{g_{k+1}^T y_k}{d_k^T y_k}, \beta_k^{PR} = -\frac{g_{k+1}^T y_k}{g_k^T g_k}, \beta_k^{FR} = -\frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$

where $\|.\|$ and $g_k = \nabla f(x_k)$ indicate the Euclidian norm of vectors.

Fletcher Reeves (FR) [10], Hestenes and Stiefel (HS) [11], Polak-Ribiere-Polyak (PR) [12] and Conjugate Descent (CD) by Fletcher [13].

This research is structured as follows. Section 2 presents our proposal, A new spectral conjugate gradient algorithm. For each repetition, Section 3 shows the algorithm's ancestry conditions. Satisfaction of the global convergence condition is presented in Section 4. Evaluation of the algorithm via some numerical experiments is conducted in Section 5.

A New Spectral Conjugate Gradient Algorithm

for Unconstrained Optimization, the algorithm is formulated as follows.

$$-H_{k+1}^{DFP} g_{k+1} = -g_{k+1} + \beta_k^{HS} d_k \tag{7}$$

$$-H_{k+1}^{DFP} g_{k+1} + g_{k+1} = \beta_k^{HS} d_k \tag{8}$$

$$d_k = \frac{-H_{k+1}^{DFP} g_{k+1} + g_{k+1}}{\beta_{k+1}^{HS}} \tag{9}$$

$$d_{k+1} = -g_{k+1} + \beta_{k+1} \left(\frac{-H_{k+1}^{DFP} g_{k+1} + g_{k+1}}{\beta_{k+1}^{HS}} \right) \tag{10}$$

Suppose that $-H_{k+1}^{DFP} g_{k+1} = d_{k+1}$

$$d_{k+1} = -g_{k+1} + \beta_{k+1} \left(\frac{-d_k + g_{k+1}}{\beta_{k+1}^{HS}} \right) \tag{11}$$

$$d_{k+1} = \left(\beta_{k+1} \frac{d_k^T y_k}{g_{k+1}^T y_k} - 1 \right) g_{k+1} - \beta_{k+1} \frac{d_k^T y_k}{g_{k+1}^T y_k} \tag{12}$$

The steps for implementing the proposed algorithm are as follows:

Firstly (Initialization), choose $x_1 \in R^n$ and calculate $f(X_1)$ and g_1. Let $d_1 = -g_1$, and set the initial to:

$$guess\ a_1 = \frac{1}{\|g_1\|}$$

Secondly (test for iterations' continuation), if $\|g_{k+1}\| \le 10^{-6}$, then stop.

Thirdly (line search), calculate $a_{k+1} > 0$ satisfying the Wole line search conditions (three and four), then the variables $X_{k+1} = X_k + a_k d_k$ will be updated.

Lastly (direction new computation), calculate $d_{k+1} = \left(\beta_{k+1}^{DY} \frac{d_k^T y_k}{g_{k+1}^T y_k} - 1 \right) g_{k+1} - \beta_k^{DY} \frac{d_k^T y_k}{g_{k+1}^T y_k} d_k$. $d_{k+1} = -g_{k+1}$ if Powell's restart criterion $\left| g_{k+1}^T g_k \right| \ge 0.2 \|g_{k+1}\|^2$. Otherwise $d_{k+1} = d$ is defined. Calculate the initial guess $a_k = a_{k-1} \frac{\|d_{k-1}\|}{\|d_k\|}$, k and continue with the second step.

Algorithm's Descent Property

The descent property of the proposed algorithm can be denoted via $d_{k+1} = \left(\beta_{k+1} \frac{d_k^T y_k}{g_{k+1}^T y_k} - 1 \right) g_{k+1} - \beta_{k+1} \frac{d_k^T y_k}{g_{k+1}^T y_k} d_k$ in the following theorem.

Theorem (1)

The following equation formulated the search direction β_{k+1} and d_{k+1}

$$d_{k+1} = \left(\beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k} - 1\right) g_{k+1} - \beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k} d_k \quad (**)$$

Will hold for all $k \geq 1$

Proof:

Induction is used for the proof.

1. $g_1 d_1 < 0, d_1 = -g_1 \rightarrow< 0$ when k equals one. Thus, using Wolfe's conditions $d_k^T y_k > 0$.
2. For all k, suppose $d_k^T y_k < 0$.
3. When $k = k + 1$ and in g_{k+1}^T, to prove the above relation is correct by multiplying the equation (**), we achieved the following.

$$g_{k+1}^T d_{k+1} = \left(\beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k} - 1\right) g_{k+1}^T g_{k+1} - \beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k} g_{k+1}^T \quad (13)$$

Let us consider

$a = \left(\beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k} - 1\right)\beta_{k+1}$, and $b = \frac{d_k^T y_k}{g_{k+1}^T y_k}$

$$g_{k+1}^T d_{k+1} = a g_{k+1}^T g_{k+1} - b g_{k+1}^T d_k \quad (14)$$

Additionally, $b > 0$ and $a > 0$

$$g_{k+1}^T d_{k+1} = a g_{k+1}^T g_{k+1} - b g_{k+1}^T d_k < 0 \quad (15)$$

$$g_{k+1}^T d_{k+1} < 0 \quad (16)$$

Global Convergence Analysis

with β_{k+1} converges, next, we investigate the conjugate gradient algorithm globally. For the convergence of the suggested algorithm, the assumptions are required.

Assumption (1) [14], [15]

1. In some Initial points and the level set $S = \{X \in R^n : f(x) \leq f(X_\circ)\}$, we assume f is bound.
2. f is differentiable continuously, and its gradient is Lipshitz continuous, $L > 0$ exists such as:

$$\|g(x) - g(y)\| \leq L\|X - Y\| \quad (17)$$

3. f is a uniformly convex function, then there exists $\mu > 0$ such that,

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \mu\|x - y\|^2, for\ any\ x, y, \in S) \quad (18)$$

or equivalently

$$y_k^T S_k \geq \mu\|S_k\|^2 and\ \mu\|S_k\|^2 \leq y_k^T S_k \leq L\|S_k\|^2 \quad (19)$$

By contrast, it is clear that positive constant β exists under assumption 1, such as:

$$\|x\| \leq \beta, \forall x \in S \quad (20)$$

$$\|\nabla f(x)\| \leq \underline{y}, \forall x \in S \quad (21)$$

Lemma (1) [16], [17]

Let eq. (20) and assumption (1) are held. Consider any conjugate gradient method in the forms (2) and (6). The descent direction is defined by d_k and the strong Wolfe line search is used to obtain a_k. If

$$\sum_{k>1} \frac{1}{\|d_{k+1}\|^2} = \infty \quad (22)$$

Accordingly, we get

$inf\|g_k\| = 0$

We refer the reader to [18]–[20] for further information.

Theorem (3)

Suppose the descent condition, the eq. (20) and assumption (1) are held.

$$d_{k+1} = \left(\beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k} - 1\right) g_{k+1} - \beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k} \quad (23)$$

We use the Wolfe line search conditions (3) and (4) to compute a_k. Note that $inf\|g_k\| = 0$ if the objective function is uniformly on set S.

Proof:

In the beginning, we substitute β_{k+1} in the direction $d_{k+1}$to determine the following.

$$d_{k+1} = \left(\beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k} - 1\right) g_{k+1} - \beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k} \quad (24)$$

$$\|d_{k+1}\|^2 = \| \left(\beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k} - 1\right) g_{k+1} - \beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k} \|^2 \quad (25)$$

Assume that

a $=\left(\beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k} - 1\right)\beta_{k+1}$, and $b = \frac{d_k^T y_k}{g_{k+1}^T y_k}$

$$\|d_{k+1}\|^2 = \|a g_{k+1} - b d_k\|^2 \quad (26)$$

$$\|d_{k+1}\|^2 \le a\|g_{k+1}\|^2 + b\|d_k\|^2 \quad (27)$$

$$\|d_{k+1}\|^2 \le a\underline{Y}^2 + b\|d_k\|^2 \quad (28)$$

$$\|d_{k+1}\|^2 \le \frac{1}{\underline{Y}^2} a(\underline{Y}^2)^2 + \underline{Y}^2 b\|d_k\|^2 \quad (29)$$

Let $c = (a(\underline{Y}^2)^2 + \underline{Y}^2 b\|d_k\|^2)$

$$\|d_{k+1}\|^2 \le c\frac{1}{\underline{Y}^2} \quad (30)$$

$$\sum_{k=1}^{\infty} \frac{1}{\|d_{k+1}\|^2} \le \frac{1}{c}\underline{Y}^2 \sum_{k\ge 1} 1 = \infty \quad (31)$$

$$\|g_k\| = 0 \quad (32)$$

Performance Evaluation and Comparisons

Here, we present some preliminary numerical results of comparison between our algorithm and Classical conjugate gradient direction one. Specifically, for unconstrained Optimization, we used $\beta_k^{DY}$to evaluate the performance of the new formal d_{k+1} in both algorithms. For each test problem taken from [21], (70) large-scale unconstrained optimization problem is selected. For each test function, the number of variables is taken as $(n = 1000, \ldots, 10000)$ and is considered in the numerical examples. The comparisons are conducted of the new versions with the classical direction. The algorithms are deployed using the standard Wolfe line search conditions (3) and (4). We assigned stopping criteria to be $\|g_k\| = 10^{-6}$ in all the cases. Via F77 default compiler settings, the utilized software is FORTRAN Language. Usually, the test functions begin point standard initially. As shown in Figures (1, 2 and 3), the findings are then drawn using Matlab. The performance profile utilized by Dolan and More$'$ in [22] has been used to evaluate and show our algorithm's performance. Furthermore, we used β_k^{FR} to compare our algorithm with the classical direction algorithm. We considered the interested solvers set $S = 2$ and $p = 700$ as the whole set of n_p test problems. For the problem p, suppose that $I_{p,s}$ is the number of objective function evaluations needed by the solver S. Accordingly, the performance ratio can be formulated as follows.

$$r_{p,s}\frac{I_{p,s}}{I_p^*} = \infty \quad (33)$$

where$I_p^* = min\,\{I_{p,s}: s \in S\}$. For all $p, and\ s$, it is clear that $r_{p,s} \ge 1$. The ration $r_{p,s}$ is considered a large number M if the mathematician couldn't solve the problem. For performance ratio $r_{p,s}$, cumulative distribution function (in below) is used to define the profile for each solver s.

$$p_s(\tau) = \frac{size\{p\in P: r_{p,s} \le \tau\}}{n_p} \quad (34)$$

Clearly, for each solver s, the percentage of problems $(p_s(1))$ is the best. We refer the reader to [22] for further information on the performance profile. We used the performance profile to analyze CPU time, the number of gradient evaluations and the number of iterations. Furthermore, in the following figures, we considered the horizontal coordinate a log scale for clear observation.

Note the following points:

- Choose a_k [23] using Wolfe conditions (11) and (12).

- Let ξ=0.01.

$d_{k+1} = \left(\beta_{k+1}^{DY} \frac{d_k^T y_k}{g_{k+1}^T y_k} - 1\right) g_{k+1} - \beta_{k+1}^{DY} \frac{d_k^T y_k}{g_{k+1}^T y_k} d_k$

- DY is

DYC is $d_{k+1} = -g_{k+1} + \beta_{k+1}^{DY} d_k$.

The following figures indicate a comparison between the new DYN algorithm and the classical DYC algorithm, where the blue curve indicates the new DYN algorithm and the red curve indicates the classical DYC algorithm. The closer the curve is to one, the better the result will be according to Donald and Moore comparisons; Hence, it became clear to us that the new DYN algorithm is clearly close to 1 in the figures, which means that it is better than the classic DYC algorithm.

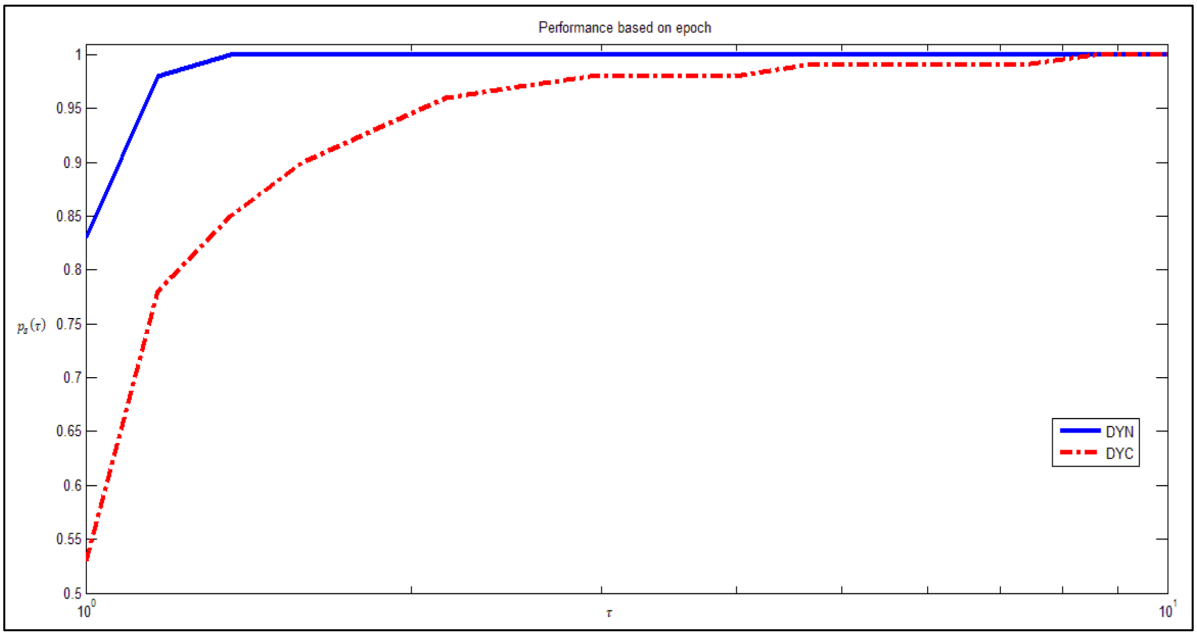

Figure 1. Performance-based on iteration

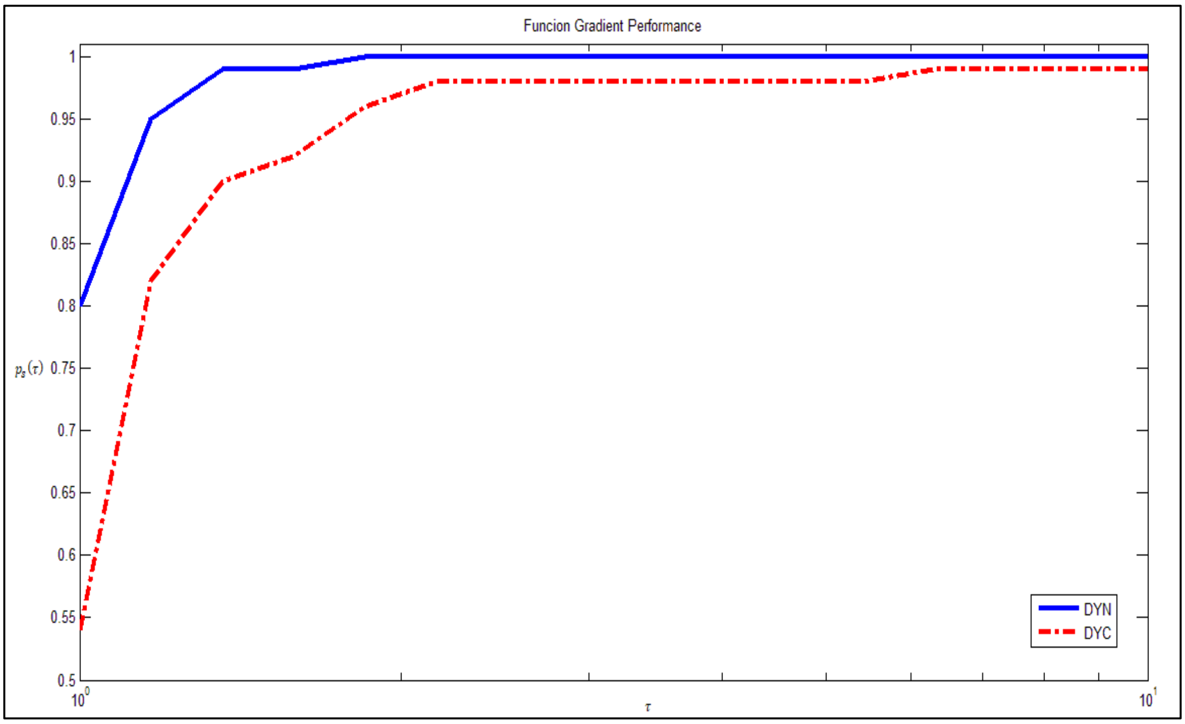

Figure 2. Performance-based on Function

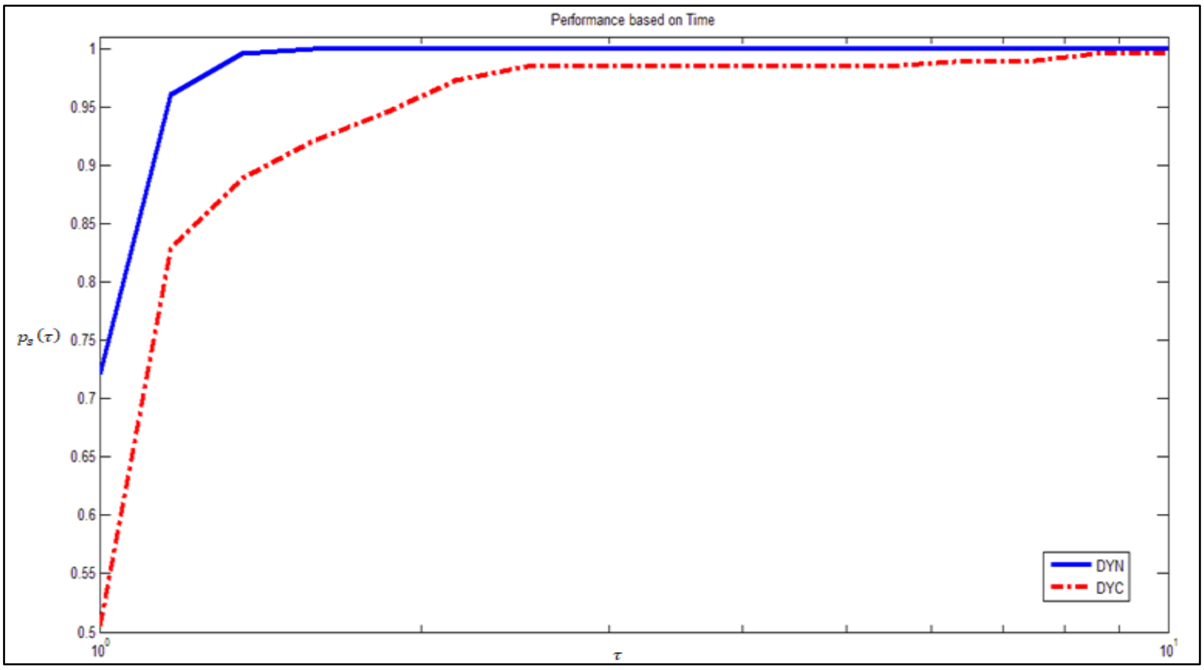

Figure 3. Performance-based on Time

Conclusion

In this research, we developed a new conjugate gradient depended on Quasi-Newton and classical conjugate gradient. The proposed algorithm satisfied the descent direction and global convergence property under certain assumptions. The numerical example demonstrated our algorithm's superiority over the classical conjugate gradient direction in solving unconstrained nonlinear optimization problems.

References

[1] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, 1988, doi: 10.1093/imanum/8.1.141.

[2] Y. Laylani, B. A. Hassan, and H. M. Khudhur, "A New Class of Optimization Methods Based on Coefficient Conjugate Gradient," *Eur. J. Pure Appl. Math.*, vol. 15, no. 4, pp. 1908–1916, 2022.

[3] H. M. Khudhur, "Modified Barzilai-Borwein Method for Steepest Descent Method to Solving Fuzzy Optimization Problems(FOP)," *Albahir J.*, vol. 12, no. 23–24, pp. 63–72, 2020.

[4] G. Zoutendijk, "Nonlinear programming, computational methods," *Integer nonlinear Program.*, pp. 37–86, 1970.

[5] Y. A. Laylani, B. A. Hassan, and H. M. Khudhur, "Enhanced spectral conjugate gradient methods for unconstrained optimization," *Comput. Sci.*, vol. 18, no. 2, pp. 163–172, 2023.

[6] L. Armijo, "Minimization of functions having lipschitz continuous first partial derivatives," *Pacific J. Math.*, vol. 16, no. 1, 1966, doi: 10.2140/pjm.1966.16.1.

[7] A. A. Goldstein, "On steepest descent," *J. Soc. Ind. Appl. Math. Ser. A Control*, vol. 3, no. 1, pp. 147–151, 1965.

[8] P. Wolfe, "Convergence Conditions for Ascent Methods. II: Some Corrections," *SIAM Rev.*, vol. 13, no. 2, pp. 185–188, Apr. 1971, doi: 10.1137/1013035.

[9] J. Nocedal and S. J. Wright, "Numerical optimization," in *Springer Series in Operations Research and Financial Engineering*, Springer Science & Business Media, 2006, pp. 1–664.

[10] R. Fletcher, "Function minimization by conjugate gradients," *Comput. J.*, vol. 7, no. 2, pp. 149–154, Feb. 1964, doi: 10.1093/comjnl/7.2.149.

[11] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Res. Natl. Bur. Stand. (1934).*, vol. 49, no. 6, p. 409, Dec. 1952, doi: 10.6028/jres.049.044.

[12] E. Polak and G. Ribiere, "Note sur la convergence de méthodes de directions conjuguées," *Rev. française d'informatique Rech. opérationnelle. Série rouge*, vol. 3, no. 16, pp. 35–43, 1969, doi: 10.1051/m2an/196903r100351.

[13] C. Witzgall and R. Fletcher, "Practical Methods of Optimization," *Math. Comput.*, vol. 53, no. 188, p. 768, Oct. 1989, doi: 10.2307/2008742.

[14] M. M. Abed, U. Öztürk, and H. M. Khudhur, "Spectral CG Algorithm for Solving Fuzzy Nonlinear Equations," *Iraqi J. Comput. Sci. Math.*, vol. 3, no. 1, pp. 1–10, Jan. 2022, doi: 10.52866/ijcsm.2022.01.01.001.

[15] Z. M. Abdullah, H. M. Khudhur, and A. Khairulla Ahmed, "Modification of the new conjugate gradient algorithm to solve nonlinear fuzzy equations," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 27, no. 3, p. 1525, Sep. 2022, doi: 10.11591/ijeecs.v27.i3.pp1525-1532.

[16] H. M. Khudhur and K. K. Abbo, "A New Type of Conjugate Gradient Technique for Solving Fuzzy Nonlinear Algebraic Equations," *J. Phys. Conf. Ser.*, vol. 1879, no. 2, p. 022111, May 2021, doi: 10.1088/1742-6596/1879/2/022111.

[17] A. S. Ahmed, H. M. Khudhur, and M. S. Najmuldeen, "A new parameter in three-term conjugate gradient algorithms for unconstrained optimization," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 23, no. 1, p. 338, Jul. 2021, doi: 10.11591/ijeecs.v23.i1.pp338-344.

[18] Y. Ismail Ibrahim and H. Mohammed Khudhur, "Modified three-term conjugate gradient algorithm and its applications in image restoration," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 3, p. 1510, Dec. 2022, doi: 10.11591/ijeecs.v28.i3.pp1510-1517.

[19] G. Li, C. Tang, and Z. Wei, "New conjugacy condition and related new conjugate gradient methods for unconstrained optimization," *J. Comput. Appl. Math.*, vol. 202, no. 2, pp. 523–539, 2007, doi: 10.1016/j.cam.2006.03.005.

[20] 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, 2020.

[21] N. Andrei, "An Unconstrained Optimization Test Functions Collection," *Adv. Model. Optim.*, vol. 10, no. 1, pp. 147–161, 2008.

[22] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Math. Program. Ser. B*, vol. 91, no. 2, pp. 201–213, 2002, doi: 10.1007/s101070100263.

[23] POWELL MJD, "SURVEY OF NUMERICAL METHODS FOR UNCONSTRAINED OPTIMIZATION," *SIAM Rev.*, vol. 12, no. 1, 1970, doi: 10.1137/1012004.