مجلة أبحاث كلية التربية الأساسية ، المجلد ١٨، العدد (٣)، لسنة ٢٠٢٢

College of Basic Education Researchers Journal. ISSN: 7452-1992 *Vol. (18), No.(3), (2022)*

CC BY

A New Spectral Conjugate Gradient method for solving unconstrained Optimization problems

Osama M. T. Wais
University of Mosul / College of Basic Education / Dept. of Mathematic

Khalil A. Abbo
University of Telafer / College of Basic Education / Dept. of Mathematic

Ibrahim A. Saleh
University of Mosul / College of Computer Sciences and Mathematics / Dept. of Software Engineering

(Received in 16/3/2022 Accepted in 17/4/2022)

Abstract:

The spectral conjugate gradient methods are fascinating, and it has been shown that they are useful for strictly convex quadratic reduction when used properly. To handle large-scale unconstrained optimization issues, a novel spectral conjugate gradient approach is suggested in this study. We devise a new methodology for determining the spectral and conjugate parameters, motivated by the benefits of the approximate optimum step size strategy utilized in the gradient method. Additionally, the new search direction meets the spectral property as well as the sufficient descent criterion. The presented method's global convergence is established under a set of appropriate assumptions.

Keywords :

Approximate optimal stepsize , Spectral conjugate gradient method , Global convergence

1- Introduction

Consider the unconstrained optimization problem with the following n variables [1][2]:

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

The conjugate gradient methods are among the most effective optimization strategies for achieving the solution of problem (1), where $f: R^n \to R$ is a continuous differentiable function. The conjugate gradient technique has the following form [3][4]:

$$x_{k+1} = x_k + \alpha_k d_k \quad , \; k = 0,1,2,3,\ldots \tag{2}$$

Where x_0 is the starting point, α_k is a step size , $g_k = \nabla f(x)$ and d_k can be taken as [5][6]:

$$d_k = \begin{cases} -g_k & : \quad k = 0 \\ -g_k + \beta_k d_{k-1} & : \quad k \geq 1 \end{cases} \qquad (3)$$

The exact and inexact line search is commonly used to estimate the step size α_k for nonlinear conjugate gradient methods. As a result, the Wolfe inexact line search was used in this investigation, with the following formula, α_k was needed to meet the normal Wolfe line search [7].

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k \qquad (4)$$

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k \qquad (5)$$

Where $0 < \delta \leq \sigma < 1$. When the value of σ is less, the line search is often more accurate, but the computational time is longer. A stronger criterion based on (4) and (6) was developed to make analysis easier [7].

$$\left|g_{k+1}^T d_k\right| \geq -\sigma g_k^T d_k \qquad (6)$$

In addition to the conditions indicated in (4), the strong Wolfe line search (SWP) has the following characteristics (6).Generally, d_k is required to satisfy [8][9]

$$d_k^T g_k < 0 \qquad (7)$$

Guarantying d_k is a descent direction of (x) at x_k . In order to retain the convergence property, it is often necessary for d_k to meet the descent requirement. [10]

$$g_k^T d_k \leq -c\|g_k\|^2 \qquad (8)$$

Where the constant > 0 .

Bergin et al. proposed a spectral conjugate gradient (SCG) technique, which is a combination of the conjugate gradient method and the spectral gradient method. let $s_k = x_{k+1} - x_k = \alpha_k d_k$ and $y_k = g_{k+1} - g_k$.
The direction d_k is termed as [11][1]

$$d_{k+1} = -\theta_{k+1} g_{+1k} + \beta_{k+1} s_k \qquad (9)$$

Different conjugate gradient algorithms will be determined by different β_k. Previous research in this field has shown that a good selection of this parameter resulted in improved numerical performance. As illustrated below, Hestenes and Stiefel (HS)[12], Fletcher and Reeves (FR)[13], Polak and Ribiere and Polyak (PRP)[14], Fletcher (CD)[15], Liu and Storey (LS)[16], and Dai and Yuan (DY)[8] are some of the most well-known classical $\beta_k$selections :

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \quad \beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k}, \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}$$
$$\beta_k^{CD} = -\frac{\|g_{k+1}\|^2}{d_k^T g_k}, \quad \beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \quad \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k} \qquad (10)$$

where $y_k = g_{k+1} - g_k$ and $\|.\|$ denotes the Euclidean norm.

2- The new spectral conjugate gradient algorithm

Spectral conjugate gradient methods of the kind (9) are proposed in this part, and the following is the name of our novel spectral conjugate gradient method:

$$d_k^{OKI2} = -(1+\omega_3)g_{k+1} + \beta^{FR} s_k \qquad (11)$$

Where $\omega_3 = \frac{\omega_1}{\omega_2}$,and

$$\omega_1 = (g_{k+1}^T g_{k+1})(y_k^T s_k) + (s_k^T g_{k+1})(y_k^T g_{k+1}) - (g_{k+1}^T g_{k+1})(y_k^T g_{k+1})$$
$$\omega_2 = (g_k^T g_k)(y_k^T g_{k+1})$$

The algorithm is denoted by the following:

<u>Algorithm 1</u>

Stage 1: Initialization. Given $x_0 \in R^n$, set $k = 0$.
Stage 2 : Calculate β_k based on (10)
Stage 3 : Calculate d_k based on (3) . If $g_k = 0$,then stop.
Stage 4 : Calculate α_k based on (4) and (6) .
Stage 5 : Update d_k based on (11)
Stage 6 : Update new point based on (2).
Stage 7 : Convergent test and stopping criteria.
If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| \le \epsilon$ then stop. Otherwise go to Step 1 with $k = k + 1$.

3- Descent property

Here, we demonstrate, using the following theorem, that our method specified in (3) and (11) yields descent direction for all iterations.

<u>Theorem 1:</u>

Consider the procedure stated in equation (2), where d_k is derived from (3) and (11). When a given increment in the step size α_k meets the Wolfe conditions (4) and (6) and $y_k^T g_{k+1} > 0$, it follows that the directions d_k are descent for every k.

Proof :

Since $d_{k+1}^{OKI2} = -(1+\omega_3) g_{k+1} - \beta^{FR} s_k$

We can prove the descent property $d_{k+1}^T g_{k+1} < 0$ by mathematical induction :

For the initial point $d_0 = -g_0$, then the new direction (9) multiplied by g_{k+1} is sufficient descent .

Let the direction in (9) hold for all k.

For k+1 we have :

$$\omega_3 = \frac{\omega_1}{\omega_2} = \frac{(g_{k+1}^T g_{k+1})(y_k^T s_k) + (s_k^T g_{k+1})(y_k^T g_{k+1}) - (g_{k+1}^T g_{k+1})(y_k^T g_{k+1})}{(g_k^T g_k)(y_k^T g_{k+1})} \tag{12}$$

$$\omega_3 = \frac{\|g_{k+1}\|^2 (y_k^T s_k)}{\|g_k\|^2 (y_k^T g_{k+1})} + \frac{(s_k^T g_{k+1})(y_k^T g_{k+1})}{(g_k^T g_k)(y_k^T g_{k+1})} - \frac{\|g_{k+1}\|^2 (y_k^T g_{k+1})}{\|g_k\|^2 (y_k^T g_{k+1})} \tag{13}$$

$$\omega_3 = \beta^{FR} \frac{(y_k^T s_k)}{(y_k^T g_{k+1})} + \frac{(s_k^T g_{k+1})}{(g_k^T g_k)} - \beta^{FR} \tag{14}$$

$$\omega_3 = \beta^{FR} \left(\frac{(y_k^T s_k)}{(y_k^T g_{k+1})} - 1 \right) + \frac{(s_k^T g_{k+1})}{(g_k^T g_k)} \tag{15}$$

Substituted (15) in (11) we have

$$d_{k+1}^{new} = -(1 + \left(\beta^{FR} \left(\frac{(y_k^T s_k)}{(y_k^T g_{k+1})} - 1 \right) + \frac{(s_k^T g_{k+1})}{(g_k^T g_k)} \right) g_{k+1} + \beta^{FR} s_k \tag{16}$$

By multiplying (16) by g_{k+1}^T we have

$$d_{k+1}^T g_{k+1} = -(1 + \left(\beta^{FR} \left(\frac{(y_k^T s_k)}{(y_k^T g_{k+1})} - 1 \right) + \frac{(s_k^T g_{k+1})}{(g_k^T g_k)} \right) g_{k+1}^T g_{k+1} + \beta^{FR} s_k^T g_{k+1} \tag{17}$$

$$= (-1 - \left(\beta^{FR} \left(\frac{(y_k^T s_k)}{(y_k^T g_{k+1})} - 1 \right) - \frac{(s_k^T g_{k+1})}{(g_k^T g_k)} \right) \|g_{k+1}\|^2 + \beta^{FR} s_k^T g_{k+1} \tag{18}$$

$$= -\|g_{k+1}\|^2 - \beta^{FR} \left(\frac{(y_k^T s_k)}{(y_k^T g_{k+1})} - 1 \right) \|g_{k+1}\|^2 - (s_k^T g_{k+1}) \beta^{FR} + \beta^{FR} s_k^T g_{k+1} \tag{19}$$

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 - \beta^{FR} \left(\frac{(y_k^T s_k)}{(y_k^T g_{k+1})} - 1 \right) \|g_{k+1}\|^2 \tag{20}$$

We have $y_k^T s_k > 0$ by Wolfe condition and $y_k^T g_{k+1} > 0$ by assumption , then $\tau = \frac{y_k^T s_k}{y_k^T g_{k+1}} > 0$

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 - \beta^{FR}(\tau - 1)\|g_{k+1}\|^2 \quad (21)$$

$$= -\|g_{k+1}\|^2 - \tau\beta^{FR}\|g_{k+1}\|^2 + \beta^{FR}\|g_{k+1}\|^2 \quad (22)$$

$$= (-1 + \beta^{FR})\|g_{k+1}\|^2 - \tau\beta^{FR}\|g_{k+1}\|^2 \quad (23)$$

$$= -c\|g_{k+1}\|^2 - \tau\beta^{FR}\|g_{k+1}\|^2 \quad :c = 1 - \beta^{FR} \quad (24)$$

$0 \le \beta^{FR} \le 1$ then we have

$$d_{k+1}^T g_{k+1} \le -c\|g_{k+1}\|^2 - \tau\beta^{FR}\|g_{k+1}\|^2 = (-c - \tau\beta^{FR})\|g_{k+1}\|^2 < 0 \quad (25)$$

$d_{k+1}^T g_{k+1} \le 0$; then the new direction is descent .

4- Global convergence

Following that, we shall demonstrate that d_{k+1}^{OKI2} converges globally. For the suggested new method to be successful in its convergence, we must make the following assumptions.

Assumption (1)[4][6][7][17]

1- Suppose that f is bounded below in the level set $S = \{x \in R^n : f(x) \le f(x_0)\}$

2- In the case of f that is continuously differentiable and whose gradient is Lipchitz continuous, there exists $L > 0$ such that :

$$\|g(x) - g(y)\| \le L\|x - y\| \quad \forall x, y \in N \quad (26)$$

3- If f is a uniformly convex function, then a constant $\rho > 0$ exists such that:

$$\left(\nabla f(x) - \nabla f(y)\right)^T (x - y) \ge \rho\|x - y\|^2 \text{ , for any } x, y \in S \quad (27)$$

or equivalently

$$y_k^T s_k \ge \rho\|s_k\|^2 \quad \text{and} \quad \rho\|s_k\|^2 \le y_k^T s_k < L\|s_k\|^2 \quad (28)$$

On the other hand, given Assumption(1), it is evident that there are positive constants δ such that :

$$\|x\| \le \delta \ , \ \forall x \in S \quad (29)$$

$$\|\nabla f(x)\| \le \gamma \ , \forall x \in S \quad (30)$$

Lemma (1)[18][19]

Suppose that assumption(1) and equation (30) hold. Consider any conjugate gradient method in from (2) and (3), where d_{k+1} is a descent direction and α_k is obtained by the strong Wolfe line search. If

$$\sum_{k>1} \frac{1}{\|d_{k+1}\|^2} = \infty \tag{31}$$

then we have

$$\lim_{k\to\infty} \inf\|g_k\| = 0 \tag{32}$$

Theorem 2:

Assuming that Assumption (1) is true and that any iterative technique of the type (11) where d_{k+1} satisfies descent condition, where α_k is determined from Wolfe conditions (4) and (6), if the objective function is uniformly convex on set S, then $\lim_{k\to\infty} \inf\|g_k\| = 0$.

Proof :

We have the direction $d_{k+1}^{OKI2} = -(1+\omega_3)g_{k+1} - \beta^{FR} s_k$

$$d_{k+1}^{OKI2} = -(1 + \frac{(g_{k+1}^T g_{k+1})(y_k^T s_k) + (s_k^T g_{k+1})(y_k^T g_{k+1}) - (g_{k+1}^T g_{k+1})(y_k^T g_{k+1})}{(g_k^T g_k)(y_k^T g_{k+1})}) g_{k+1} - \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} s_k \tag{3١}$$

By taking $\|.\|$ to the both sides , we have

$$\|d_{k+1}^{OKI2}\| \le \|-1\| - \frac{(\|g_{k+1}^T g_{k+1}\|)(\|y_k^T s_k\|) - (\|s_k^T g_{k+1}\|)(\|y_k^T g_{k+1}\|) + (\|g_{k+1}^T g_{k+1}\|)(\|y_k^T g_{k+1}\|)}{(\|g_k^T g_k\|)(\|y_k^T g_{k+1}\|)}) \|g_{k+1}\| - \frac{\|g_{k+1}^T g_{k+1}\|}{\|g_k^T g_k\|} \|s_k\| \tag{32}$$

$$\|d_{k+1}^{OKI2}\| \le \left(\|-1\| - \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \frac{\|y_k^T s_k\|}{\|y_k^T g_{k+1}\|} - \frac{\|s_k^T g_{k+1}\|}{\|g_k\|^2} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2}\right) \|g_{k+1}\| + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \|s_k\| \tag{33}$$

$$\le \left(\|-1\| - |\beta^{FR}| \frac{\|s_k\|}{\|g_{k+1}\|} - \frac{\|s_k\|}{\|g_k\|} + |\beta^{FR}|\right) \|g_{k+1}\| + |\beta^{FR}| \|s_k\| \tag{34}$$

$$\le (\|g_{k+1}\| - |\beta^{FR}| \|s_k\| - \|s_k\| |\beta^{FR}| + |\beta^{FR}| \|g_{k+1}\|) + \beta^{FR} \|s_k\| \tag{35}$$

$$\le (\|g_{k+1}\| - |\beta^{FR}| \|g_{k+1}\|) - \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \|s_{k+1}\| \tag{36}$$

$$\le \|g_{k+1}\| \left(1 + |\beta^{FR}| + \frac{\|s_{k+1}\|}{\|g_k\|^2}\right) \tag{37}$$

$$\|d_{k+1}^{OKI2}\| \le \mu \|g_{k+1}\| \tag{38}$$

$$\frac{1}{\|d_{k+1}^{OKI2}\|} \ge \frac{1}{\mu \|g_{k+1}\|} \tag{39}$$

$$\sum_{k=0}^{\infty} \frac{1}{\|d_{k+1}^{OKI2}\|} \ge \frac{1}{\mu \|g_{k+1}\|} \sum_{k\ge 1} 1 = \infty \tag{40}$$

$$\lim_{k\to\infty} \|g_k\| = 0 \tag{41}$$

5- Numerical results

The numerical results of the proposed new approach generated from the application of the new formula for the beta conjugation coefficients as well as the Wolfe (4) and (5) conditional set of test functions in the unconstrained optimization (Andrei, 2008)[20] will be explained in this section. We always require the practical side while calculating unconstrained optimization methods since it complements the theoretical side. To fully appreciate the algorithm's strength, we must use it in practice and test it on a variety of non-linear unconstrained situations. Many test functions have been chosen to evaluate the performance of the proposed method, which are provided in this work and detailed in the Appendix. The functions are chosen for dimensions n=100,...,1000, and the performance of these new suggested algorithms is compared to that of the FR method. The $\|g_k\| = 10^{-6}$ stopping condition has been used. All of the code is written in double precision FORTRAN using the F77 compiler settings. The test routines normally begin with a point standard and then record numerical findings in Matlab's figures (1), (2), and (3).

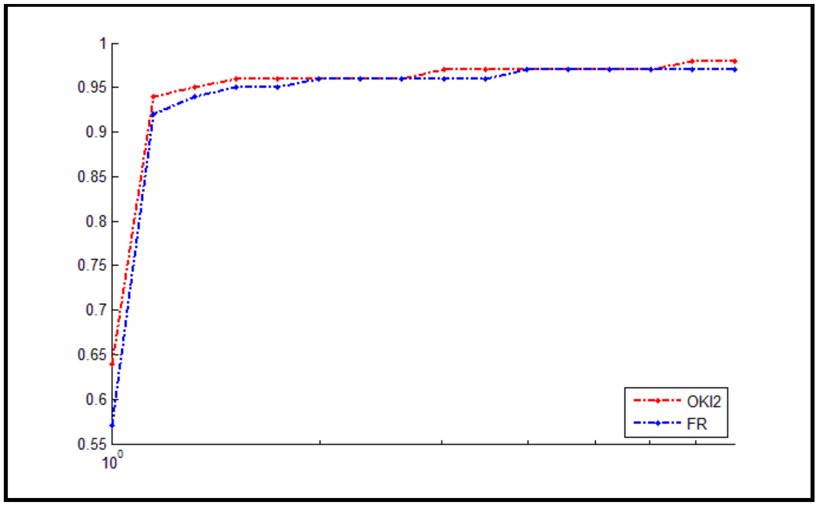

Figure (1): Performance according to Function

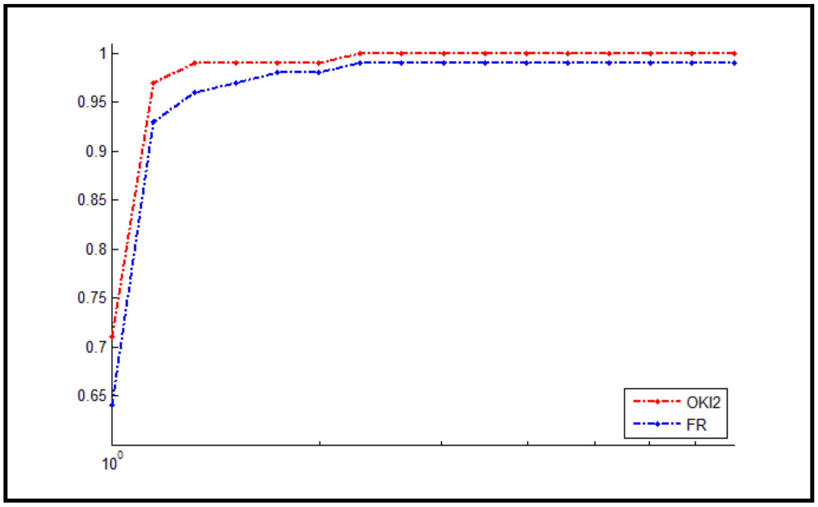

Figure (2): Performance according to iteration

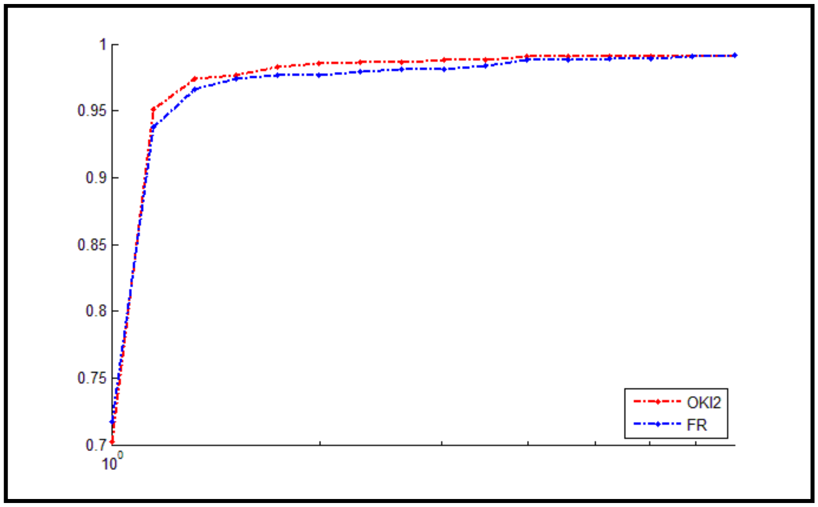

Figure (3): Performance according to Time

6- Conclusions

Based on the (FR) method, we provide a novel Spectral Conjugate Gradient Methods based on the (FR) approach in this study. We investigate the qualities of the formula from a scientific standpoint and demonstrate the properties of descent and convergence by using a number of assumptions. We also looked into the characteristics of the matrices and compared their performance to that of the FR approach, which produced favorable results.

Reference

[1] O. M. T. Wais, "Two Modified Spectral Conjugate Gradient methods for Optimization," *Coll. Basic Educ. Res. J.*, vol. 14, no. 4, pp. 533–553, 2018.

[2] 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.

[3] 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.

[4] K. K. Abbo and O. M. T. Waiss, "Investigation on Extended Conjugate Gradient Non-linear methods for solving Unconstrained Optimization."

[5] 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.

[6] 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. Stud.*, vol. 13, no. 2, pp. 101–113, 2018.

[7] M. A. Wolfe, *Numerical methods for unconstrained optimization: An introduction*. Van Nostrand Reinhold New York, 1978.

[8] 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.

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

[10] Y. Dai and L. Liao, "New conjugacy conditions and related non-linear CG methods' Appl," *Math optim.,(43). Spring-verlag, New York*, 2001.

[11] B. A. Hassan, S. S. Ismael, and O. M. Taher, "A new Spectral Gradient Coefficient is based on the Convex Combination of the two Different Conjugate Coefficients for Optimization."

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

[13] 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.

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

[15] R. Fletcher, *Practical methods of optimization*. John Wiley & Sons, 1987.

[16] Y. Liu and C. Storey, “Efficient generalized conjugate gradient algorithms, part 1: theory,” *J. Optim. Theory Appl.*, vol. 69, no. 1, pp. 129–137, 1991.

[17] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.

[18] B. A. Hassan, O. M. T. Wais, and A. A. Mahmood, “A Class of Descent Conjugate Gradient Methods for Solving Optimization Problems,” *Appl. Math. Sci.*, vol. 13, no. 12, pp. 559–567, 2019.

[19] 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.

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