For cited: *Hali, I. H., Abbo, K. K. & Ibrahim, H. H. (2022). A New Conjugate Gradient Algorithm for Unconstrained Optimization. Journal of Current Researches on Educational Studies, 12 (2), 1-12.*

Journal of Current Researches on Educational Studies
(JoCuRES)
www.jocures.com

Year: 2022
Volume: 12
Issue: 2

Crossref doi: 10.26579/jocures.12.2.1

Research Article

A New Conjugate Gradient Algorithm for Unconstrained Optimization

Isam H. HALİ[1], Khalil K. ABBO[2] & Hassan H. IBRAHİM[3]

Keywords
Conjugate gradient algorithm, unconstrained optimization.

Article History
Received
11 Jun, 2022
Accepted
30 Dec, 2022

Abstract

Parallel gradient methods (CG) include a class of unrestricted optimization algorithms with low memory requirements and strong local and global convergence characteristics .The conjugate gradient methods began with the seminal paper of Hestense and Stiefel in 1952 .The algorithm is presented as an approach to solving similar and positive linear systems, but is an alternative to gausi justice that is perfectly suited to solving big problems. In 1964 Fletcher generalized it to solve the unconstrained optimization problems today, the results of unrestricted improvement are applied in different branches of science, as well as generally in practice. Recently different nonlinear conjugate gradient methods are developed. In this paper, we developed a new nonlinear conjugate gradient algorithm. Deriving this technique on the basis of the ratios property and the case of association, the descent property and global convergence with some mild assumption of the algorithm was proved, Numerical comparison of the algorithm with other related (CG) methods are given. The new algorithm is very effective for solving unrestricted optimization problems.

1. Introduction

In order to solve the problem of unrestricted non-linear improvement

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

Where $f: R^n \to R$ is continuously differentiable and, bounded from below. The starting point $x_0 \epsilon R^n$, The Conjugated Gradient (CG) method produces a sequence $\{x_k\} \subset R^n$ s.t :-

$$x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

[1] Corresponding Author. ORCID: 0000-0002-5619-8197. College of Science,University of Kirkuk, Kirkuk, Iraq, isam.h.halil@uokirkuk.edu.iq
[2] ORCID: 0000-0001-5858-625X. College of Basic Education, University of Telafer, Telafer, Iraq, dr.khalil-khuder@uotelafer.edu.iq
[3] ORCID: 0000-0002-0800-7106. College of Computer Science and Mathematics, Tikrit University, Sallahaddin, Tikrit Iraq, Hassan1962pl@tu.edu.iq

Where $\alpha_k > 0$ denotes the phase size as determined by line search and the directions d_k it is presented by [1] [2]

$$d_{k+1} = -g_{k+1} + \beta_k s_k , \quad d_0 = -g_0 \tag{3}$$

β_k is the (CG) parameter in the last reference and $s_k = x_{k+1} - x_k$, $g_k = \nabla f(x_k)$.
Let $\| \cdot \|$ be the Euclidean norm, and write $y_k = g_{k+1} - g_k$ (4)

Hager and Zhang [3] conducted an excellent survey of conjugate gradient (CG) methods. The numerical $\beta_k$values for the various conjugate gradient (CG) methods are different.

The Hybrid conjugative gradient (CG) method is a specific set of different conjugative gradient (CG) methods.

Its aim is to improve the performance of these techniques and prevent jamming.

In order to select the parameter β_k of the method in this research, we list the follows β_k selections [3] :

Fletcher and Reeves: [4] $\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}$; (5)

Dai and Yuan: [5] $\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T(g_k - g_{k-1})}$ (6)

Conjugate descent, proposed by Fletcher: [6] $\beta_k^{CD} = \frac{\|g_{k+1}\|^2}{-g_k^T s_k}$. (7)

Gradient methods associated with the selection of β_k taken in (5), (6) and (7), have a strong convergence goods. They may, however, have mediocre operational performance due to confusion [7] [8] .

On the other side, the ways of Polak and Ribiere [9] , Hestenes and Stiefel [10] [15], also Liu and Storey [11] in generally may not be converging. But the computer usually performs better.

β_k options in these ways are, respectively:

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2} \ , \tag{8}$$

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})} \ , \tag{9}$$

$$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-g_k^T s_k} \ , \tag{10}$$

The difference between FR and CD that is kept by sufficient proportions to search wolfe's strong line (the constraint $\sigma \leq \frac{1}{2}$ that arose with the FR does not require a CD) [12].

Furthermore, for a search line that meets generalized conditions of the wolf with $\sigma_1 < 1$ and $\sigma_2 = 0$, it can be shown that the CD style is universally convergent [13].

On the other hand, not much research has been done on β_k^{LS}selecting an update parameter with the exception of paper, But expect the techniques developed to analyze the PRP method to be applied to LS.

2. A New (CG) Method (KHI2)

In this search, we suggest a new beta as follows:

Consider $d_{k+1} = -g_{k+1} + \beta_k d_k$ (11)

Now a new $\beta_{k+1}^{KHI2} = \frac{\|g_{k+1}\|^2}{y_k^T s_k} - \frac{2\|y_k\|^2 (s_k^T g_{k+1})}{\max\{4\|g_{k+1}\|^2, (y_k^T s_k)^2}$ (12)

Then by substitute equation (12) in equation (11) we get :

$$d_{k+1} = -g_{k+1} + \{\frac{\|g_{k+1}\|^2}{y_k^T s_k} - \frac{2\|y_k\|^2 (s_k^T g_{k+1})}{\max\{4\|g_{k+1}\|^2, (y_k^T s_k)^2}\} s_k \quad (13)$$

Algorithm (2.1)

Step 1. The initial point $x_0 \in R^n$, and the accurate $\in > 0$, and we find $d_0 = -g_0$,and the set $k = 0$

Step 2. Testing for the continuation of iterations. If $\|g_{k+1}\| \leq 10^{-6}$ stop. Otherwise, move to the next step.

Step 3. Determine the step length $\alpha_k$by Wolfe line search conditions [14]

$$f_{k+1} \leq f_k + c_1 \alpha_k d_k^T g_k \quad (14)$$

$$d_k^T g_{k+1} \geq c_2 d_k^T g_k \quad (15)$$

And go to step (4).

Step 4. Calculate $x_{k+1} = x_k + \alpha_k d_k$.

Step 5. we compute the search direction by the equation (13).

The set k=k+1 repeat to the step 2.

3. The Descent Property

We will mention the proof of the Descent Property $(d_k^T g_k < 0)$ for our proposed new formula for the conjugated gradient algorithm and that the sufficient descent of the conjugated gradient algorithm is expressed as follows $(d_k^T g_k < -c\|g_k\|^2)$.

Theorem (1)

Let the search direction for each (k≥0) is generated by the formula (13). Suppose that the step size meets the Wolfe standard condition (SDWC) (14) and (15) then achieves sufficient descent property.

proof:-

when k=1 then $d_1 = -g_1 \rightarrow d_1^T g_1 = -\|g_1\|^2$

2- for k=k+1 then

$$d_{k+1} = -g_{k+1} + \{\frac{\|g_{k+1}\|^2}{y_k^T s_k} - \frac{2\|y_k\|^2 s_k^T g_{k+1}}{\max\{4\|g_{k+1}\|^2, (y_k^T s_k)^2}\} s_k$$

1-Let max{ $\max\{4\|g_{k+1}\|^2, (y_k^T s_k)^2 = 4\|g_{k+1}\|^2$ then

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 s_k^T g_{k+1} y_k^T s_k}{(y_k^T s_k)^2} - \frac{2\|y_k\|^2 (s_k^T g_{k+1})^2}{4\|g_{k+1}\|^2}$$

Let $\mathrm{u} = (y_k^T s_k)\mathrm{g}_{k+1}$ $\quad v = (s_k^T g_{k+1}) g_{k+1}$

$$\mathrm{u}^T v \le \tfrac{1}{2}\{\|u\|^2 + \|v\|^2\}$$

$$\frac{\|g_{k+1}\|^2 s_k^T g_{k+1} y_k^T s_k}{(y_k^T s_k)^2} = \frac{\frac{1}{2}\{(y_k^T s_k)^2 \|g_{k+1}\|^2 + (s_k^T g_{k+1})^2 \|g_{k+1}\|^2}{(y_k^T s_k)^2}$$

$$= \frac{1}{2}\|g_{k+1}\|^2 + \frac{(s_k^T g_{k+1})^2 \|g_{k+1}\|^2}{(y_k^T s_k)^2}$$

$$\because \quad \frac{2(s_k^T g_{k+1})^2 \|y_k\|^2}{4\|g_{k+1}\|^2} \le \frac{(s_k^T g_{k+1})^2 \|g_{k+1}\|^2}{(y_k^T s_k)^2}$$

$$g_{k+1}^T d_{k+1} \le -\frac{1}{2}\|g_{k+1}\|^2 + \frac{2(s_k^T g_{k+1})\|y_k\|^2}{4\|g_{k+1}\|^2} - \frac{2(s_k^T g_{k+1})\|y_k\|^2}{4\|g_{k+1}\|^2}$$

$$g_{k+1}^T d_{k+1} < -\frac{1}{2}\|g_{k+1}\|^2$$

2-If $\max\{4\|g_{k+1}\|^2, (y_k^T s_k)^2\} = (y_k^T s_k)^2$

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \{\frac{\|g_{k+1}\|^2 s_k^T g_{k+1}}{y_k^T s_k} - \frac{2(s_k^T g_{k+1})^2 \|y_k\|^2}{(y_k^T s_k)^2}$$

$$\leq -\|g_{k+1}\|^2 + \frac{2\|y_k\|^2 (s_k^T g_{k+1})^2}{(y_k^T s_k)^2} - \frac{2(s_k^T g_{k+1})^2 \|y_k\|^2}{(y_k^T s_k)^2}$$

$$\therefore g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2$$

We see that the sufficient property holds for any line search

4. The Global Convergent

In this we prove the global convergence the proposed method for strongly convex function. First we need the following assumption.

Assumption: (1) [4]

1-The level group $\Omega = \{x \in R^n \mid f(x) \leq f(x_0)\}$ is bounded which there are exits $B > 0$ so that $\|x\| \leq B, \ \forall x \in \Omega$

2-In some neighborhood N, of level set , f bounded and constantly different and gradient is Lipshitz continuous, there exists $L > 0$ such that:

$\|g(x) - g(y)\| \leq L\|x - y\| \quad \forall x, y \in N.$

Note that these assumptions mean that there is a constant Γ so that $\|g(x)\| \leq \Gamma$, $\forall x \in \Omega$

Lemma(1) [18]

Suppose that d_k is the direction of descent and Δf satisfies the Lipschitz condition for all x on the Line segment connecting x_k and x_{k+1} if the line search satisfies the Wolfe conditions, then

$$\alpha_k \geq \frac{1-\sigma}{L} \frac{|g_k^T d_k|}{\|d_k\|^2}$$

Theorem (2)

Assume that f is Lipschitz continuous and strongly convex on the level set Ω ,i.e., Constants L and μ>0 exist ,so that

$$\|\Delta f(x) - \Delta f(y)\| \leq L\|x - y\| \tag{16}$$

$$y_k^T s_k \geq \mu\|x - y\|^2 \tag{17}$$

For all x,y ϵ Ω. If the combined gradient (13) method is performed with the Wolfe line search terms at each step, either $g_k = 0$ for some k, or $\lim_{k\to\infty} g_k = 0$

Proof:-

Assume that $g_k \neq 0$ for all k, otherwise the theorem is true

Since f is strongly convex it follows that

$y_k^T s_k \geq \mu\alpha_k \|d_k\|^2$

By the sufficient descent and assumption $g_k \neq 0$ ensure that $d_k \neq 0$, by Lemma(1) $\alpha_k > 0$ there for

$$y_k^T s_k > 0$$

Because f is strongly convex over Ω, and bounded below

In the Wolfe condition, after summing over k, the upper bound, it is found that

$$\sum_{k=1}^{\infty} \alpha_k g_k^T d_k > -\infty$$

When this is combined with the lower bound for α_k given by Lemma(1) and sufficient descent
gives

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty \tag{18}$$

From Lipschitz continuity

$$\|y_k\| = \|g_{k+1} - g_k\| \leq L\alpha_k \|d_k\|$$

Now using $y_k^T s_k > \mu\alpha_k \|d_k\|$ and $\beta^{KHI} = \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} - \frac{2\|y_k\|^2 s_k^T g_{k+1}}{\max\{4\|g_{k+1}\|^2, (y_k^T s_k)^2\}}$ the following estimation

Is obtained

$$\left|\beta_k^{KHI2}\right| = \left| \frac{g_{k+1}^T g_{k+1}}{\mu\alpha_k \|d_k\|^2} - \frac{2\|y_k\|^2 \|d_k\| \|g_{k+1}\|}{\max\{4\|g_{k+1}\|^2, (y_k^T s_k)^2} \right|$$

$$\leq \frac{\|g_{k+1}\|^2}{\mu\alpha_k \|d_k\|^2} + 2\frac{\|y_k\| \|d_k\| \|g_{k+1}\|}{\max\{4\|g_{k+1}\|^2, (y_k^T s_k)^2\}}$$

$$\leq \frac{\|y_k\| \|g_{k+1}\|}{\mu\alpha_k \|d_k\|^2} + 2\frac{L^2\alpha^2 \|d_k\|^3 \|g_{k+1}\|}{\max\{4\|g_{k+1}\|^2, (y_k^T s_k)^2\}}$$

If $\max\{4\|g_{k+1}\|^2, (y_k^T s_k)^2\} = 4\|g_{k+1}\|^2$. Since $\|g_{k+1}\|^2 \leq \|y_k\|^2$

$$\left|\beta_k^{KHI2}\right| \leq \frac{L\alpha_k \|d_k\| \|g_{k+1}\|}{\mu\alpha_k \|d_k\|^2} + \frac{2L^2\alpha^2 \|d_k\|^3 \|g_{k+1}\|}{2\mu^2\alpha_k^2 \|d_k\|^4}$$

$$\leq \left(\frac{L}{\mu} + \frac{L^2}{\mu^2}\right)\frac{\|g_{k+1}\|}{\|d_k\|}.$$

Hence $\|d_{k+1}\| \leq \|g_{k+1}\| + \left|\beta_k^{KHI}\right| \|d_k\| \leq \left(1 + \frac{L}{\mu} + \frac{L^2}{\mu^2}\right)\|g_{k+1}\|$

Using this upper for d_k in (18), It follows that $\sum_{k=1}^{\infty} \|g_k\|^2 < \infty$

If $\max\{4\|g_{k+1}\|^2, (y_k^T s)\} = (y_k^T s_k)$ then the prove is similar as in [18] and is omitted.

5. Numerical Experiment

In this search, we compare the new official KHI 2 performance developed with the new CG method. We've chosen (75) large-scale unconstrained optimization problems to investigate, taken from [1] For any of the test problems. We also considered numerical tests with the number of variables n=100, 1000 for every test element. These new versions are compared to the HS, FR algorithms, known conjugate gradient algorithms. With search terms, wolf solid line search terms, all these algorithms are implemented. In each of these examples the stop criterion is $\|g_k\| = 10^{-6}$. All of the codes are written in F77 default compiler settings in double-precision FORTRAN Language. Usually the test functions are standard initially, and the summary numerical results are recorded in fig. 1, 2 and 3. The performance profile is used by [16] to demonstrate the performance of the new two-period CG algorithm developed with KHI 2. Determine P = 75 as the complete set of $n_p$test problems and S = 3, the set of interested solutions. Suppose $l_{p,s}$ is the number of objective function assessments required by solver S of problem P. Determine the performance ratio as follows:

$$r_{p,s} = \frac{l_{p,s}}{l_p^*} \tag{19}$$

Where $l_p^* = min\{l_{p,s}: s \in S\}$. It is clear that $r_{p,s} \geq 1 \; for\, all\, p, s$. When a solution fails to solve a problem, the $r_{p,s}$ ratio is set to a large number M. The following output cumulative distribution function $r_{p,s}$ is used to describe the performance profile for each analyzer.

$$p_s(t) = \frac{size\,\{p \in P : r_{p,s} \leq t\}}{n_p}$$

$p_s(1)$ of course, denotes the percentage of problems for which solver S is the highest.

Figure 1. Performance Based on Iteration

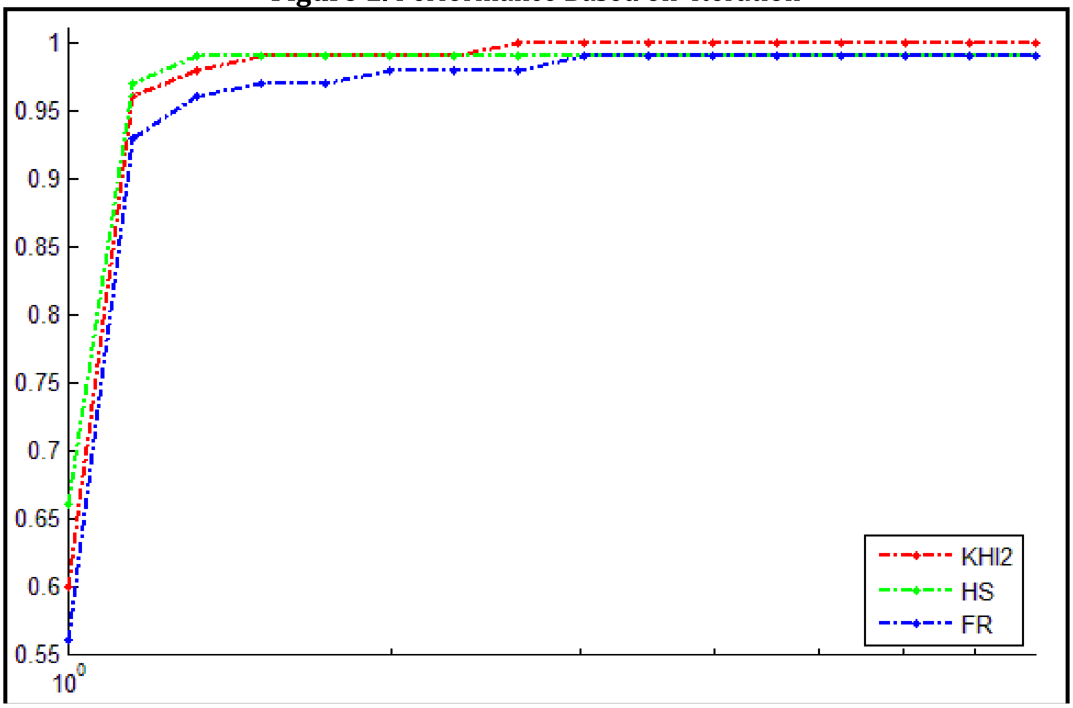


Figure 2. Performance-Based on Function evalution

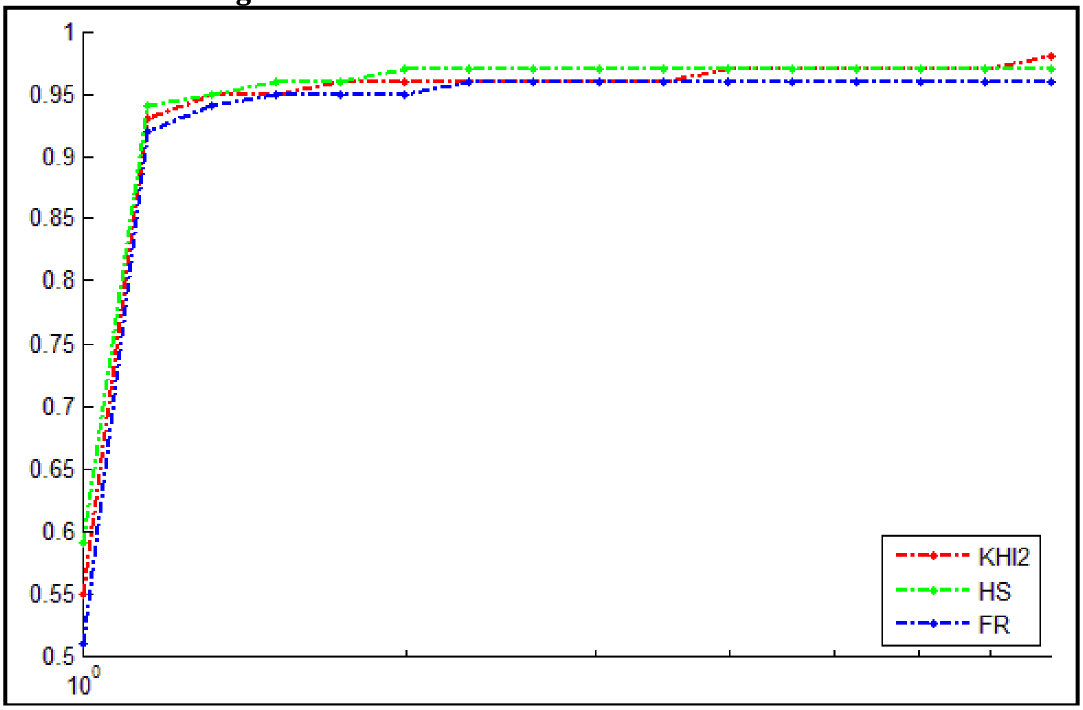
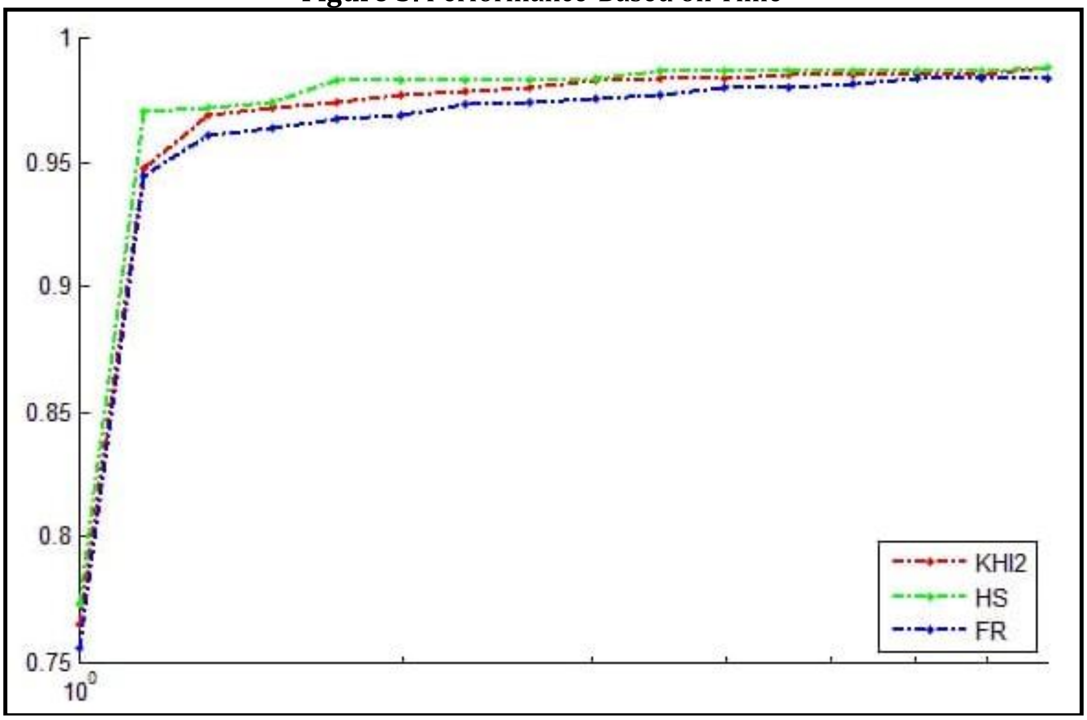


Figure 3. Performance-Based on Time



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.
2. J. E. Dennis, Jr. and J. J. Moré, “Quasi-Newton Methods, Motivation and Theory,” *SIAM Rev.*, vol. 19, no. 1, 1977, doi: 10.1137/1019005.
3. N. Andrei,'Nonlinear Conjugate Gradient Methods For Uncostrained
4. Optimization'. Springer Nature Switzerland AG, 2020
5. N. Andrei, “A simple three-term conjugate gradient algorithm for unconstrained optimization,” *J. Comput. Appl. Math.*, vol. 241, no. 1, 2013, doi: 10.1016/j.cam.2012.10.002.
6. 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.
7. 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.
8. R. Fletcher, *Practical methods of optimization*. John Wiley & Sons, 1987.
9. H. M. Khudhur, “Numerical and analytical study of some descent algorithms to solve unconstrained Optimization problems,” University of Mosul, 2015.
10. Hisham M. Khudhur; Khalil K. Abbo, “A New Type of Conjugate Gradient Technique for Solving Fuzzy Nonlinear Algebraic Equations,” 2020.
11. 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.
12. M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, vol. 49, no. 1. NBS Washington, DC, 1952.
13. 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.
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, 2015.
15. 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, 2015.
16. M. A. Wolfe, *Numerical methods for unconstrained optimization: An introduction*. Van Nostrand Reinhold New York, 1978.
17. 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.

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

19. W. W. Hager 'A new conjugate gradient method with guranteed descent and an sufficient line search'. SIAM J. OPTIM. Vol 16, No.1.