Al-Rafidain Journal of Computer Sciences and Mathematics (RJCSM)

www.csmj.uomosul.edu.iq

Investigation a Conjugacy Coefficient for Conjugate Gradient Method to Solving Unconstrained Optimization Problems

Mays Basil Jarjis[1] and Hamsa Th. Chilmerane[2]

[1,2] Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, Iraq
Email: mays.23csp152@student.uomosul.edu.iq[1] and

Article information

Article history:
Received 19 October ,2024
Revised 21 December ,2024
Accepted 21 January ,2025
Published 26 June ,2025

Keywords:
Conjugate,
gradient,
unconstrained,
optimization,
strong Wolfe,
condition,
conjucacy,
coefficient.

Correspondence:
Mays Basil Jarjis
Email:
mays.23csp152@student.uomosul.edu.iq

Abstract

In this research, a new conjugacy coefficient is derived for conjugate gradient method (C.G) and a new direction was obtained. In theory, this direction achieves sufficiently desent condition by using strong wolfe line search and global convergence is proved. When contrasted with the starnder HS (C.G) technique, the numerical performance of this approach is very remarkable. The Dolan-More performance profile was applied in order to carry out this determination. The amount of time that the central processing unit (CPU) spends, the number of iterations (NOI), and the number of function evaluations (NOF) all play a role in determining this profile. It was determined through the utilization of the Dolan-More profile that this was the case.

DOI: 10.33899/csmj.2025.154547.1153, ©Authors, 2025, College of Computer Science and Mathematics, University of Mosul, Iraq.

1. Introduction

The (C.G) algorithm is one of the efficient numerical algorithms that are characterized by simplicity and nice convergence property

$$\min_{x\in R} f(x) \tag{1}$$

Let x be a variable and $f: R^n \to R$. The CG method produces $\{x_k\}$ as follows:

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

Where x_k is the current point , $\alpha_k > 0$ is step size .

$$d_{k+1} = \begin{cases} -g_{k+1} & \text{for } k = 0 \\ -g_{k+1} + \beta_k d_k & \text{for } k \ge 1 \end{cases} \tag{3}$$

Where k is integer , g_k is the gradient of the $f(x)$ and β_k is coefficient of $(C.G)$algorithm [1, 2 ,3 ,4, 5, 6]

The line search of C.G algorithm depend the weak Wolfe conditions are commonly utilized:

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k \nabla f(x_k)^T d_k \tag{4}$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \ge \sigma \nabla f(x_k)^T d_k \tag{5}$$

Additionally, robust Wolfe conditions necessitate the inclusion of condition (4) and

$$\left|g(x_k + \alpha_k d_k)^T d_k\right| \le -\sigma \ g_k^T d_k \tag{6}$$

2. The new conjugacy oefficient

In this section, we will use the modified QN direction in [7] which is defined in eq.(7) satisfy perry condition to generate a new scale (λ_k). Then, we equality stander (C.G) direction which is defined in eq.(3) and the modified QN direction to derive a new conjugacy coefficient (β_k^*) as follows:

$$d_{k+1} = -\nabla^{-1} f_{k+1} g_{k+1} + \lambda_k s_k \tag{7}$$

Multiplying eq (7) by (y)

$$d_{k+1}^{\mathrm{T}} y_k = -y_k^{\mathrm{T}} \nabla^{-1} f_{k+1} g_{k+1} + \lambda_k y_k^{\mathrm{T}} s_k \tag{8}$$

From QN condition ($\nabla^{-1} f_{k+1} y_k = s_k$) and perry condition ($d_{k+1}^{\mathrm{T}} y_k = -t g_{k+1}^{\mathrm{T}} s_k$)

$$-\mathrm{t} g_{k+1}^{\mathrm{T}} s_k = -s_k^{\mathrm{T}} g_{k+1} + \lambda_k y_k^{\mathrm{T}} s_k \tag{9}$$

$$\lambda_k = \frac{(1-t) g_{k+1}^{\mathrm{T}} s_k}{y_k^{\mathrm{T}} s_k} \tag{10}$$

After, we submit eq. (10) in eq. (7), we get

$$d_{k+1} = -\nabla^{-1} f_{k+1} g_{k+1} + \left[\frac{(1-t) g_{k+1}^T s_k}{y_k^T s_k}\right] s_k \tag{11}$$

And by equating eq. (11) and eq.(3), we obtained

$$-\nabla^{-1} f_{k+1} g_{k+1} + \left[\frac{(1-t) g_{k+1}^T s_k}{y_k^T s_k}\right] s_k = -g_{k+1} + \beta_k^* s_k \tag{12}$$

Multiply by both side of eq.(12) $s_k^{\mathrm{T}} \nabla f_{k+1}$ and we get:

$$-s_k^T \nabla f_{k+1} \nabla^{-1} f_{k+1} g_{k+1} + \left[\frac{(1-t) g_{k+1}^T s_k}{y_k^T s_k}\right] s_k^T \nabla f_{k+1} s_k = -s_k^T \nabla f_{k+1} g_{k+1} + \beta^* s_k^T \nabla f_{k+1} s_k \tag{13}$$

$$-s_k^{\mathrm{T}} g_{k+1} + \left[\frac{(1-t) g_{k+1}^{\mathrm{T}} s_k}{y_k^{\mathrm{T}} s_k}\right] s_k^T \nabla f_{k+1} s_k = -s_k^T \nabla f_{k+1} g_{k+1} + \beta^* s_k^{\mathrm{T}} \nabla f_{k+1} s_k \tag{14}$$

$$\beta_k^* = \frac{-s_k^T g_{k+1} + s_k^T \nabla f_{k+1} g_{k+1} + \left[\frac{(1-t) g_{k+1}^T s_k \ s_k^T \nabla f_{k+1} s_k}{y_k^T s_k}\right]}{s_k^T \nabla f_{k+1} s_k} \tag{15}$$

$$\beta_k^* = \frac{-g_{k+1}^T s_k}{s_k^T \nabla f_{k+1} s_k} + \frac{s_k^T \nabla f_{k+1} g_{k+1}}{s_k^T \nabla f_{k+1} s_k} + \left[\frac{(1-t) g_{k+1}^T s_k}{y_k^T s_k}\right] \tag{16}$$

From [8] since

$$s_k^{\mathrm{T}} \nabla f_{k+1} s_k = 2/3 \ s_k^{\mathrm{T}} y_k + 2/3 (f_k - f_{k+1}) y_k$$

, we get :

$$\beta_k^* = \frac{-g_{k+1}^T s_k}{2/3 \ s_k^{\mathrm{T}} y_k + 2/3 \ (f_k - f_{k+1})} + \frac{y_k^T g_{k+1}}{2/3 \ s_k^{\mathrm{T}} y_k + 2/3 \ (f_k - f_{k+1})} + \left[\frac{(1-t) g_{k+1}^T s_k}{y_k^T s_k}\right]$$

Then new direction is defined:

$$d_{k+1} = -g_{k+1} + \left(\frac{-g_{k+1}^T s_k}{\frac{2}{3} s_k^{\mathrm{T}} y_k + \frac{2}{3}(f_k - f_{k+1})} + \frac{y_k^T g_{k+1}}{\frac{2}{3} s_k^{\mathrm{T}} y_k + \frac{2}{3}(f_k - f_{k+1})} + \left[\frac{(1-t) g_{k+1}^T s_k}{y_k^T s_k}\right]\right) s_k \tag{17}$$

The new algorithm :

Step1: given $x_0 \in R^n$, Set k= 0.

Step2: let $d_0 = -g_0$

Step3: Determine the positive step length (α_k) that satisfies equations (4) and (6), and then determine x_{k+1} in eq.(2)

Step4: if $\|g_k\| \le 10^{-5}$, Cease operations; otherwise, calculate the new direction with equation eq (17).

Step5: If k= n or Powell restart $\frac{|g_k^T g_{k+1}|}{\|g_{k+1}\|^2} \ge 0.2$ [9], Then, continue to step 2. Otherwise, set k to k+1 and proceed to step 3.

Theorem [1]:

Let the line search α_k in (2) satisfies the strong Wolfe condition, then the new search direction given by eq (17) is a sufficient descent direction.

Proof:

After we multiplying both sides of Eq. (17) by $\left(\frac{g_{k+1}}{\|g_{k+1}\|^2}\right)$then we get:

$$\frac{d_{k+1}^{\mathrm{T}} g_{k+1}}{\left\|g_{k+1}\right\|^2} + 1 = \left(\frac{-g_{k+1}^T s_k}{\frac{2}{3} s_k^{\mathrm{T}} y_k + \frac{2}{3}(f_k - f_{k+1})} + \frac{y_k^T g_{k+1}}{\frac{2}{3} s_k^{\mathrm{T}} y_k + \frac{2}{3}(f_k - f_{k+1})} + \left[\frac{(1-t) g_{k+1}^T s_k}{y_k^T s_k}\right]\right) \frac{s_k^{\mathrm{T}} g_{k+1}}{\left\|g_{k+1}\right\|^2} \tag{18}$$

By using eq.(6),we obtain

$$\le \left(\frac{\sigma g_k^T s_k}{\frac{2}{3} s_k^{\mathrm{T}} y_k + \frac{2}{3}(f_k - f_{k+1})} + \frac{y_k^T g_{k+1}}{\frac{2}{3} s_k^{\mathrm{T}} y_k + \frac{2}{3}(f_k - f_{k+1})} + \left[\frac{(1-t) - \sigma g_k^T s_k}{y_k^T s_k}\right]\right) \frac{-\sigma s_k^{\mathrm{T}} g_k}{\left\|g_{k+1}\right\|^2}$$

since $y_k^{\mathrm{T}} g_{k+1} \le \left\|y_k\right\| \left\|g_{k+1}\right\|$ and from (4), we obtain

$$\leq \Big(\frac{-\sigma^2 (g_k^T s_k)^2}{[2/3\ s_k^T y_k - 2/3\ \delta\alpha_k g_k^T d_k]\|g_{k+1}\|^2} + \frac{-\sigma s_k^T g_k \|y_k\|\|g_{k+1}\|}{[2/3\ d_k^T y_k - 2/3\ \delta\alpha_k g_k^T d_k]\|g_{k+1}\|^2} + \left[\frac{(1-t)\sigma^2 (g_k^T d_k)^2}{y_k^T s_k \|g_{k+1}\|^2}\right]\Big)$$

Since $s_k^T g_k \leq \frac{-s_k^T y_k}{(\sigma+1)}$

$$\leq \frac{-\sigma^2 \left(\frac{s_k^T y_k}{(\sigma+1)}\right)^2}{\left[2/3\ s_k^T y_k [1 - 2/3\ \delta\alpha_k \frac{(\sigma+1)}{s_k^T y_k}\right]\|g_{k+1}\|^2} + \frac{\frac{\sigma s_k^T y_k}{(\sigma+1)}\|y_k\|\|g_{k+1}\|}{\left[2/3\ s_k^T y_k - 2/3\ \delta\alpha_k \frac{(\sigma+1)}{s_k^T y_k}\right]\|g_{k+1}\|^2} + \left[\frac{(1-t)\sigma^2 \left(\frac{s_k^T y_k}{(\sigma+1)}\right)^2}{y_k^T s_k \|g_{k+1}\|^2}\right]\Big)$$

$$\leq \frac{\frac{\sigma s_k^T y_k}{(\sigma+1)}\|y_k\|\|g_{k+1}\|}{\left[2/3\ s_k^T y_k [1 - \delta\alpha_k \left(\frac{\sigma+1}{(s_k^T y_k)}\right)\ \right]\|g_{k+1}\|^2} + \left[\frac{(1-t)\sigma^2 \left(\frac{1}{(\sigma+1)}\right)^2 s_k^T y_k}{\|g_{k+1}\|^2}\right] = \hat{u}$$

where û is small constant

$$d_{k+1}^T g_{k+1} \leq -(1-\gamma)\|g_{k+1}\|^2$$
(19)

The proof is complet.

ASSUMPTION(A) [10]

(i) The set S, defined as $S = \{x: f(x) \leq f(x_0)\}$, is bounded, indicating the existence of a positive scalar $b > 0$ such that $\|x\| \leq b, \forall x \in s$.

(ii) The function f demonstrates continuous differentiability inside a neighborhood N of S, and its gradient satisfies the Lipschitz condition, as given by the equation:

$$\|g(x) - g(y)\| L \|x - y\|, \forall x, y \in N\ \mathrm{N} \qquad (20)$$

On the basis of these assumptions concerning f, we are able to arrive at the conclusion that there is a positive constant represented by γ>0 that occurs in such a way that:

$$\gamma \leq \|\nabla f(x)\| \leq \bar{\gamma} \qquad (21)$$

Below the assumptions (i) and (ii) on *f*, we are able to deduce that there exists $\gamma > 0$ such as

$$\gamma \leq \|\nabla f(x)\| \leq \bar{\gamma} \qquad (22)$$

For strictly convex function we have :

(iii)

$$(g(x) - g(y))(x-y) \geq \mu\|x-y\|^2, \quad \forall x, y \in S, \mu > 0 \qquad (23)$$

Theorm[2] : **Global convergence**

Given Assumption [A]and theorem [1] hold then

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

Proof

Since this is the case, taking into consideration the absolute value of β_k^*, we obtain:

$$|\beta^*| = \left| \frac{-g_{k+1}^T s_k}{\frac{2}{3}s_k^T y_k + \frac{2}{3}(f_k - f_{k+1})} + \frac{y_k^T g_{k+1}}{\frac{2}{3}s_k^T y_k + \frac{2}{3}(f_k - f_{k+1})} + \left[\frac{(1-t)g_{k+1}^T s_k}{y_k^T s_k}\right] \right|$$

$$|\beta^*| \leq \left| \frac{-g_{k+1}^T s_k}{-\frac{2}{3}[s_k^T y_k + \frac{2}{3}(f_{k+1} - f_k)]} \right| + \left| \frac{y_k^T g_{k+1}}{\frac{2}{3}s_k^T y_k + \frac{2}{3}(f_{k+1} - f_k)} \right| + \left| \left[\frac{(1-t)g_{k+1}^T s_k}{y_k^T s_k}\right] \right|$$

From eq, (4),eq.5 and since $y_k^T g_{k+1} \leq \|y_k\|\|g_{k+1}\|$ we have:

$$|\beta^*| \leq \left| \frac{\sigma g_k^T s_k}{[2/3\ s_k^T y_k - 2/3\ \delta\alpha_k g_k^T d_k]} \right| + \left| \frac{\|y_k\|\|g_{k+1}\|}{[2/3\ s_k^T y_k - 2/3\ \delta\alpha_k g_k^T d_k]} \right| + \left| \frac{(1-t)\sigma g_k^T s_k}{y_k^T s_k} \right|$$

Since $s_k = -g_k$ [11] we obtain

$$|\beta^*| \leq \left| \frac{\sigma\|g_k\|^2}{\left[2/3\ \mu_k\|s_k\| + 2/3\ \delta\alpha_k\|g_k\|^2\right]} \right| + \left| \frac{\|y_k\|\|g_{k+1}\|}{\left[2/3\ \mu_k\|s_k\| + 2/3\ \delta\alpha_k\|g_k\|^2\right]} \right| + \left| \frac{(1-t)\sigma\|g_k\|^2}{\mu_k\|s_k\|} \right| = \varpi$$

$$\|d_{k+1}\| \leq \|g_{k+1}\| + \varpi\|s_k\| = D$$

$\sum_{k\geq 1} \frac{1}{\|d_{k+1}\|^2} \geq \frac{1}{D^2} \sum_{k\geq 1} 1 = \infty.$

i.e. $\lim_{k\to\infty} \|g_k\| = 0.$

4. Numerical result and comparisons

The numerical results of the new technique and the HS C.G method are reported in this section. These results are based on test problems chosen from [12], and they are presented in this section. Taking into consideration the fact that the halting criterion is$\|g_k\| \leq 10^{-5}$, we are having this in mind. The employment of cubic lines is what allows for the unique method to be accomplished. It is necessary to make use of the Dolan-More tool [13] in order to guarantee that the performance of the new approach is satisfactory.

Figures (1,2) Using the Dolan-More graph, this research presents an illustration of the performance of the novel method. This illustration is offered in this study. Particularly for problem dimensions that are somewhat close to (1000, 10000), the focus is on the number of function evaluations (NOF), which will be discussed further below. In **Figures (3, 4)**, It is possible to accomplish the goals that have been set by placing an emphasis on the performance of the new technique, which is dependent on the (NOI) with dimensions that are typically between 1000 and 10000. Because of this, we are able to accomplish the specified results.

Figure (5,6) displays the graphical description of the new technique, which is based on the amount of time spent by the central processing unit (CPU) and has dimensions of 1000 and 10000. These parameters, which are used to characterize the approach, are displayed presently.

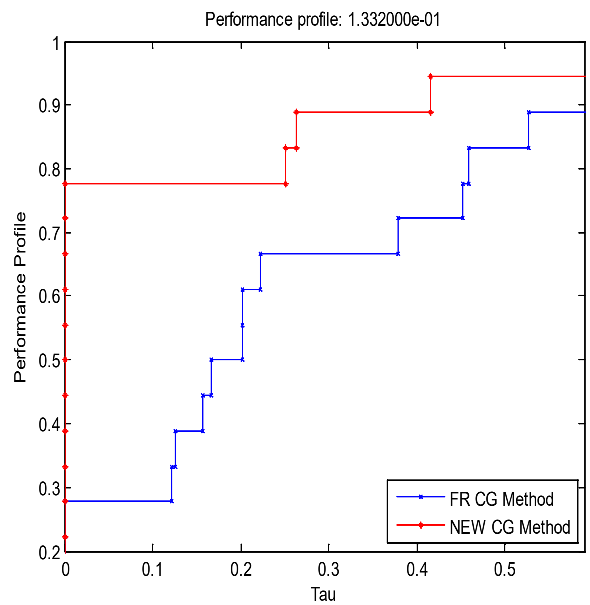

Figure 1. performance profiles of NOF with (n=1000)

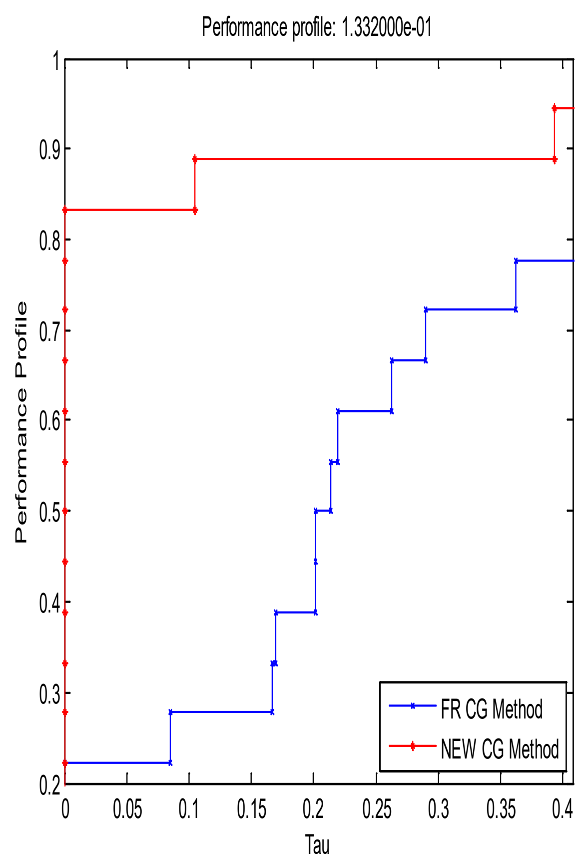

Figure 2. performance profiles of NOF with (n=10000)

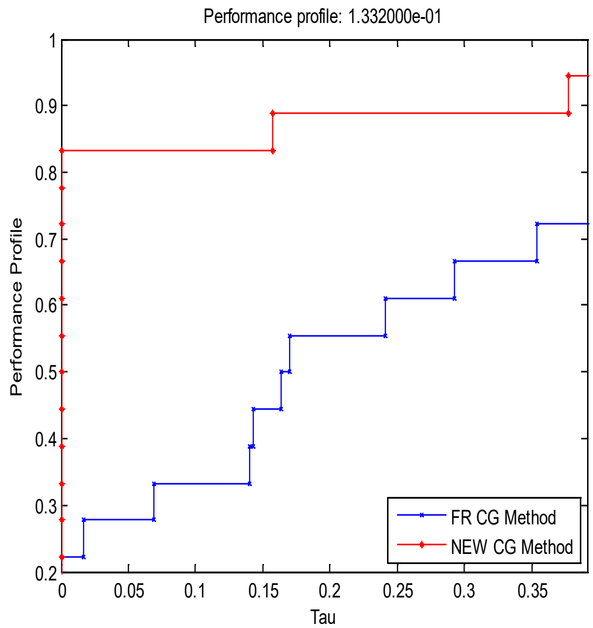

Figure 3. performance profiles of NOI with (n=1000)

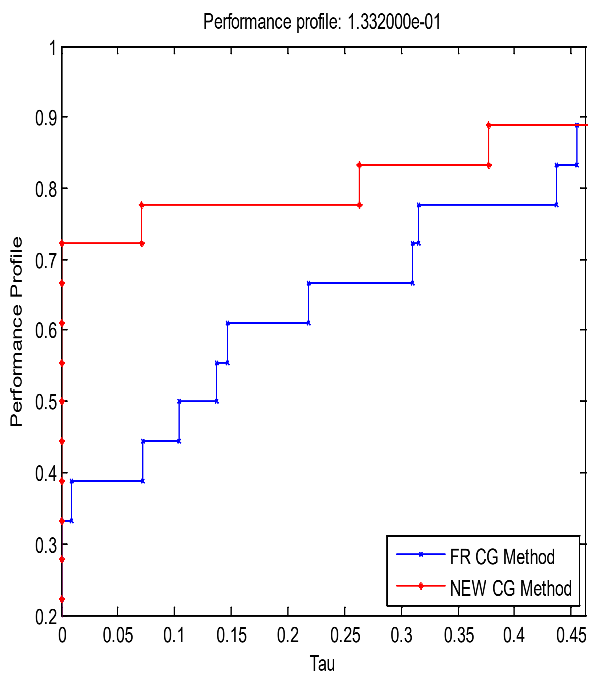

Figure 4. performance profiles of NOI with (n=10000)

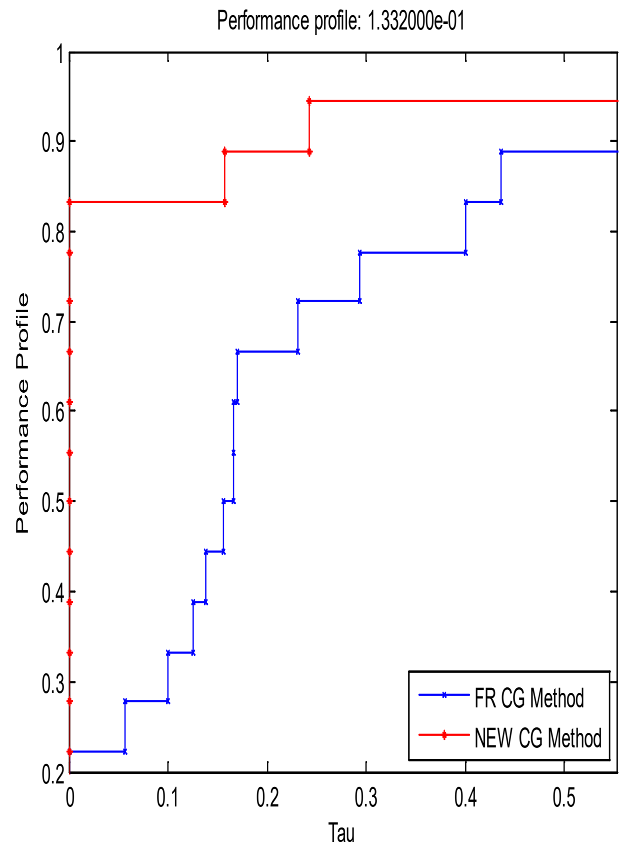

Figure 5. performance profiles of CPU time with (n=1000)

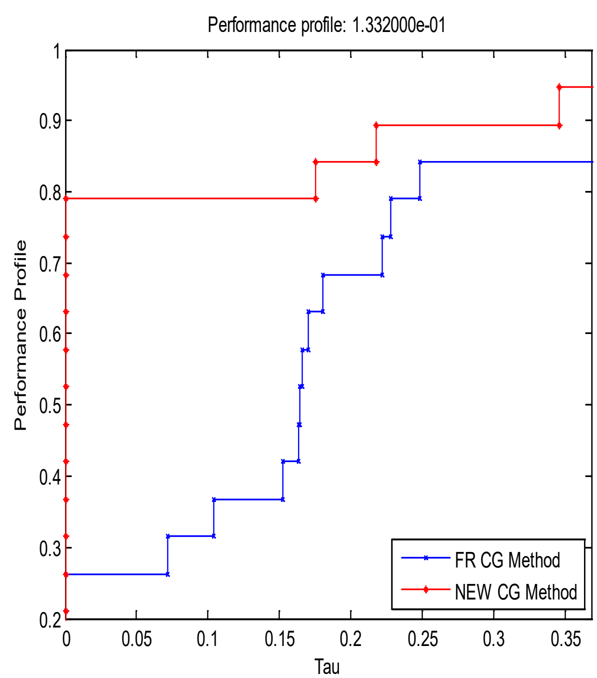

Figure 6. performance profiles of CPU time with (n=10000)

Conclusion

On the other hand, this was only the case when those aspects were taken into consideration. Despite the fact that the one-of-a-kind method was successful in establishing both sufficient descent and global convergence in certain instances, this was only the case when those criteria were taken into consideration. This success of our algorithm is proved by the numerical results that are depicted in the graphics that were explained earlier in this paragraph. Making a direct contrast with the traditional method to high-speed computing is the means by which this is accomplished. “

Conflict of interest

None.

References

[1] Hestenes, M. R. & Stiefel, E., (1952). Methods of Conjugate Gradientsfor Solving Linear Systems. Journal of Research of the National Bureau of Standards, 5(49), 409-436.

[2] Fletcher, R. & Reeves, C. M., (1964). Function Minimization byConjugate Gradients. Computer Journal, 7, 149-154.

[3] Polak, E. & Ribiére, G., (1969). Note Sur la Convergence de Directions Conjuguée. Revue Francaise Information, Recherche. Operationnelle,(16), pp.35-43.

[4] Hassan, B. A., & Sadiq, H. M. (2022). A new formula on the conjugate gradient method for removing impulse noise images, Bulletin of the south Ural State University. Ser. Mathematical Modelling, Programming &Computer Software,15(4),123-130.

[5] Chilmeran, H. S., Ahmed, H. I., Hamed, E. T., & Al-Bayati, A. Y. (2021). A new modification of the quasi-newton method for unconstrained optimization. Indonesian Journal of Electrical Engineering and Computer Science, 21(3), 1683-1691.

[6] Najm, H. Y., & Ahmed, H. I. (2022). Improving Jellyfish Swarm Algorithm for Solving Unconstrained Optimization Problems.Information Technology To Enhance e-learning and Other Application (226-231).

[7] Ibrahim, M. A. H. Bin, Mamat, M. and June, L. W. (2014) " BCGF method: Anew search direction", Sains Malayasiana, 43(10), pp. 1591-1597.Malaysiana, 43(10), 1591-1597.

[8] Hassan, B. A., & Abdullah, A. A. A. (2024). Influence to new formulas gradient for removing impulse noise images. Bulletin of the South Ural State University series Mathematics Mechanics Physics, 17(1),64–74

[9] Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. Mathematical programming, 12, 241-254.

[10] Yabe, H., & Takano, M. (2004). Global convergence properties of nonlinear conjugate gradient methods with modified secant condition.Computational optimization and Applications, 28, 203-225.

[11] Andrei, N. (2010) Amodified Polak-Ribiere-polyak conjugate gradient algorithm for unconstrained optimization, Research Institute for InformaticsCenter for Advanced Modeling and Optimization 60(12) 1457-1471.

[12] Andrei, N. (2008) An Unconstrained Optimization Test Functions Collection, Advanced Modelling and Optimization, 10(1), 147–161.

[13] Dolan, E. D., & More, J. J. (2002). Benchmarking optimization software with performance profiles. Mathematical programming, 91(2), 201-213.