The modification conjugate coefficient for conjugate gradient method to solving unconstrained optimization problems

Abed Ahmad Hassan[1*] and *Hamsa Th.* Saeed[1]

[1]Department of Mathematics, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq.

Abstract. In this manuscript, a novel conjugate coefficient is derived through the utilization of equality conjugate gradient direction in conjunction with a modified quasi-Newton (QN) direction. The demonstrated results establish that this innovative approach ensures both sufficient descent and global convergence, meeting the criteria of the Wolf line search condition and inexact line search. Through numerical experiments, we employ the proposed method on a diverse set of established test functions, clearly demonstrating its enhanced efficiency. A comparative analysis is conducted against the FR CG method using the Dolan-More performance profile based on the number of function evaluations (NOF), number of iterations (NOI), and CPU time.

1 Introduction

The Conjugate Gradient (CG) method stands out as a crucial technique for addressing unconstrained optimization problems.

Enable us to consider the objective problem:

$$\underset{x\in R^n}{\text{minimize}}\ f(x) \tag{1}$$

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

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

The step length, represented as α_k, can be determined using a line search procedure.

The direction, denoted as d_k, can be defined as follows:

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

Where β_k define in [1,2,3,4,5,6].

To facilitate the convergence analysis of the Conjugate Gradient (CG) method, the weak Wolfe conditions are commonly utilized:

* Corresponding author : abd.ahmad.alhassan@gmail.com

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

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

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

Utilizing the strong Wolfe conditions involves incorporating condition (5) and

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

Where $0<\delta<\sigma<1$, [7]. "

Ibrahim and Mamat in their work [8] derived a direction, defined as

$$d_{k+1}=-H_{k+1}g_{k+1}+\lambda g_{k+1}\ ,\ 0<\lambda<1 \tag{7}$$

Where H_{k+1} is symmetric positive definite matrix. "

2 The modified conjugate coefficient.

In this manuscript, we will establish a novel conjugate coefficient as follows:
by equality eq. (3) and (7), we get

$$-g_{k+1}+\beta_k^* s_k=-H_{k+1}g_{k+1}+\lambda_k s_k \tag{8}$$

We using formula of H_{k+1} defined by [9]

$$H_{k+1}=I-\frac{s_k y_k^T+y_k s_k^T}{s_k^T y_k}+\left(1+\frac{\|y_k\|}{\|s_k\|}\right)\frac{s_k s_k^T}{s_k^T y_k} \tag{9}$$

Substituting eq. (9) in eq. (8) we get:

$$-g_{k+1}+\beta_k^* s_k=-g_{k+1}+\frac{s_k^T g_{k+1}}{s_k^T y_k}y_k+\frac{y_k^T g_{k+1}}{s_k^T y_k}s_k-\left(1+\frac{\|y_k\|}{\|s_k\|}\right)\frac{s_k^T g_{k+1}}{s_k^T y_k}s_k+\lambda_k s_k \tag{10}$$

Multiplying both side of eq. (10) by y_k, we obtain

$$-g_{k+1}^T y_k+\beta_k^* s_k^T y_k=-g_{k+1}^T y_k+\frac{s_k^T g_{k+1}}{s_k^T y_k}\|y_k\|+\frac{y_k^T g_{k+1}}{s_k^T y_k}s_k^T y_k-\left(1+\frac{\|y_k\|}{\|s_k\|}\right)\frac{s_k^T g_{k+1}}{s_k^T y_k}s_k^T y_k+\lambda_k s_k^T y_k$$

After some algebra operation, we obtain

$$\therefore \beta_k^*=\frac{g_{k+1}^T y_k+\|y_k\|s_k^T g_{k+1}\left(\frac{1}{s_k^T y_k}-\frac{1}{\|s_k\|}\right)-s_k^T g_{k+1}+\lambda_k s_k^T y_k}{s_k^T y_k} \tag{11}$$

If we use ELS and $\lambda=0, \beta_k^*$ return to $\beta_k^{H/S}$,the new direction is defined as:

$$\therefore d_{k+1}^{new}=-g_{k+1}+\left(\frac{g_{k+1}^T y_k+\|y_k\|s_k^T g_{k+1}\left(\frac{1}{s_k^T y_k}-\frac{1}{\|s_k\|}\right)-s_k^T g_{k+1}+\lambda_k s_k^T y_k}{s_k^T y_k}\right)s_k \tag{12}$$ "

2.1 The new algorithm 2.1:

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

Step2: let $d_0=-g_0$

Step3: Determine the positive step length, denoted as α_k, that satisfies equations (5) and (8), and then set:

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

Step4: if $\|g_k\|\le 10^{-5}$, then stop

Step5: Otherwise, compute the new direction using (12)

Step6: If k= n or Powell restart $\frac{|g_k^T g_{k+1}|}{\|g_{k+1}\|^2}\ge 0.2$ [10], then proceed to step 2; otherwise, set k=k+1 and go to step 3.

"

2.2 Theorem 2.2:

"Let the sequence x_{k+1} and d_{k+1} in equation (12) compute α_k which satisfied strong wolf condition

then sufficient descent condition hold satisfied $d_{k+1}^T g_{k+1} \leq -(1-\delta)\|g_{k+1}\|^2$.

2.2.1 Proof:

After multiplying both side of eq. (12) by $\frac{g_{k+1}}{\|g_{k+1}\|^2}$,we get:

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

Since $s_k^T g_{k+1} \leq s_k^T y_k$ and $g_{k+1}^T y_k \leq \|g_{k+1}\|\|y_k\|$,we get

$$\leq \frac{\|g_{k+1}\|\|y_k\|}{s_k^T y_k} \frac{s_k^T y_k}{\|g_{k+1}\|^2} + \frac{\|y_k\| s_k^T y_k}{s_k^T y_k} \left(\frac{1}{s_k^T y_k} \right) \frac{s_k^T y_k}{\|g_{k+1}\|^2} - \frac{\|y_k\|}{s_k^T y_k \|s_k\|} \frac{s_k^T y_k s_k^T y_k}{\|g_{k+1}\|^2}$$
$$- \frac{s_k^T y_k}{s_k^T y_k} \left(\frac{s_k^T y_k}{\|g_{k+1}\|^2} \right) + \frac{\lambda_k s_k^T y_k}{\|g_{k+1}\|^2}$$
$$\leq \frac{\|y_k\|}{\|g_{k+1}\|} + \frac{\|y_k\|}{\|g_{k+1}\|^2} - \frac{\|y_k\| s_k^T y_k}{\|s_k\| \|g_{k+1}\|^2} - \frac{s_k^T y_k}{\|g_{k+1}\|^2} + \frac{\lambda_k s_k^T y_k}{\|g_{k+1}\|^2}$$
$$\leq \frac{\|y_k\|}{\|g_{k+1}\|} + \frac{\|y_k\|}{\|g_{k+1}\|^2} + \frac{\lambda_k s_k^T y_k}{\|g_{k+1}\|^2} = \delta \qquad , \ 0 < \delta < 1$$
$$\therefore d_{k+1}^T g_{k+1} \leq -(1-\delta)\|g_{k+1}\|^2 \qquad (13)$$ "

3 Assumption

"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 $\|x\| \leq b, \forall x \in s$.

Within a neighborhood N of S, the function f exhibits continuous differentiability, and its gradient adheres to the Lipschitz condition, as expressed by

$$\|g(x) - g(y)\| \leq L\|x - y\|, \ \forall x, y \in N \qquad (14)$$

Given these assumptions on f, we can deduce the existence of a positive constant $\gamma > 0$ such that:

$$\gamma \leq \|\nabla f(x)\| \leq \bar{\gamma} \qquad (15)$$

Furthermore, the inequality

$$\big(g(x) - g(y)\big)(x - y) \geq \hat{\mu}\|x - y\|^2, \forall x, y \in S, \mu > 0 \qquad (15)$$

holds, where $\hat{\mu}$ is a positive constant. [11] "

3.1 Lemma 3.1:

"Assuming the fulfillment of assumption (3.1) and considering any conjugate gradient method, where d_{k+1} is a descent direction and α_k satisfies conditions (5) and (6), if:"

$$\sum_{k\geq 1}\frac{1}{\left\|d_{k+1}\right\|^2}=\infty \qquad (16)$$

Then

$$\lim_{k\to\infty}(\inf\left\|g_k\right\|)=0\ [12] \qquad (17)\ \text{“}$$

3.2 Theorem 3.2:

“If we assume that assumption (3.1) is fulfilled and the direction d_{k+1} defined by equation (12) qualifies as a descent direction, with α_k being computed using equations (7) and (9), then:

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

. “

3.2.1 Proof

By considering the absolute value of β_k^*, we obtain:

$$|\beta_k^*|=\left|\frac{g_{k+1}^T y_k+\|y_k\|s_k^T g_{k+1}\left(\frac{1}{s_k^T y_k}-\frac{1}{\|s_k\|}\right)-s_k^T g_{k+1}+\lambda_k s_k^T y_k}{s_k^T y_k}\right|$$

$$|\beta_k^*|\leq\left|\frac{g_{k+1}^T y_k}{s_k^T y_k}\right|+\left|\frac{\|y_k\|s_k^T g_{k+1}}{(s_k^T y_k)^2}\right|+\left|\frac{\|y_k\|s_k^T g_{k+1}}{s_k^T y_k\|s_k\|}\right|+\left|\frac{s_k^T g_{k+1}}{s_k^T y_k}\right|+|\lambda_k|$$

from Wolfe condition, and $g_{k+1}^T y_k\leq\|g_{k+1}\|\|y_k\|$ we have

$$|\beta_k^*|\leq\frac{\|g_{k+1}\|\|y_k\|}{s_k^T g_{k+1}-s_k^T g_k}+\frac{\|y_k\|\,s_k^T g_{k+1}}{(s_k^T g_{k+1}-s_k^T g_k)^2}+\frac{\|y_k\|s_k^T g_{k+1}}{(s_k^T g_{k+1}-s_k^T g_k)\|s_k\|}+\frac{s_k^T g_{k+1}}{(s_k^T g_{k+1}-s_k^T g_k)}+\lambda_k$$

From eq. (6) and eq. (7) and since $s_k=-g_k$

$$|\beta_k^*|\leq\frac{\|g_{k+1}\|\|y_k\|}{-\sigma\|g_k\|^2+\|g_k\|^2}+\frac{\|y_k\|\sigma\|g_k\|^2}{(-\sigma\|g_k\|^2+\|g_k\|^2)^2}+\frac{\|y_k\|\sigma\|g_k\|^2}{(-\sigma\|g_k\|^2+\|g_k\|^2)\|s_k\|}+\frac{\sigma\|g_k\|^2}{-\sigma\|g_k\|^2+\|g_k\|^2}+\lambda_k$$

$$|\beta_k^*|\leq\frac{\|g_{k+1}\|\|y_k\|}{(1-\sigma)\|g_k\|^2}+\frac{\|y_k\|\sigma\|g_k\|^2}{((1-\sigma)\|g_k\|^2)^2}+\frac{\|y_k\|\sigma\|g_k\|^2}{((1-\sigma)\|g_k\|^2)\|s_k\|}+\frac{\sigma\|g_k\|^2}{(1-\sigma)\|g_k\|^2}+\lambda_k$$

$$|\beta_k^*|\leq\frac{\|g_{k+1}\|\|y_k\|}{(1-\sigma)\|g_k\|^2}+\frac{\sigma\|y_k\|}{(1-\sigma)^2\|g_k\|^2}+\frac{\sigma\|y_k\|}{(1-\sigma)\|s_k\|}+\frac{\sigma}{(1-\sigma)}+\lambda_k=\tau$$

By taking norm value of eq. (12), we get

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

$$\sum_{k\geq 1}\frac{1}{\|d_{k+1}\|}\geq\frac{1}{\sum D}\sum_{k\geq 1}1=\infty$$

$\therefore\ \lim_{k\to\infty}\|g_{k+1}\|=0$“

4 Numerical result and comparisons

“In this section, we present the numerical results of both the new method and the FR CG method using test problems from [13], with the step length determined by the Wolf condition

(5), and considering $\|g_k\| \leq 10^{-6}$ as the stopping criterion. Both methods are implemented with cubic lines. The Dolan-More tool [14] is employed to showcase the performance of the techniques.

Figures 1,2 in this study illustrate the performance of our technique through the Dolan-More graph, focusing on the number of function evaluations (NOF), particularly for problem dimensions around (1000, 10000). In Figures 3, 4 , a similar pattern is observed, emphasizing the performance of our method in comparison to the baseline methods, with a focus on the number of iterations (NOI) for problem dimensions around (1000, 10000).

Figure 5,6 depicts the graphical representation for our new method, based on CPU time. Therefore, from the previous three figures, we can deduce that the new method is well-suited for addressing multi-dimensional issues within the dimensions of (1000, 10000). "

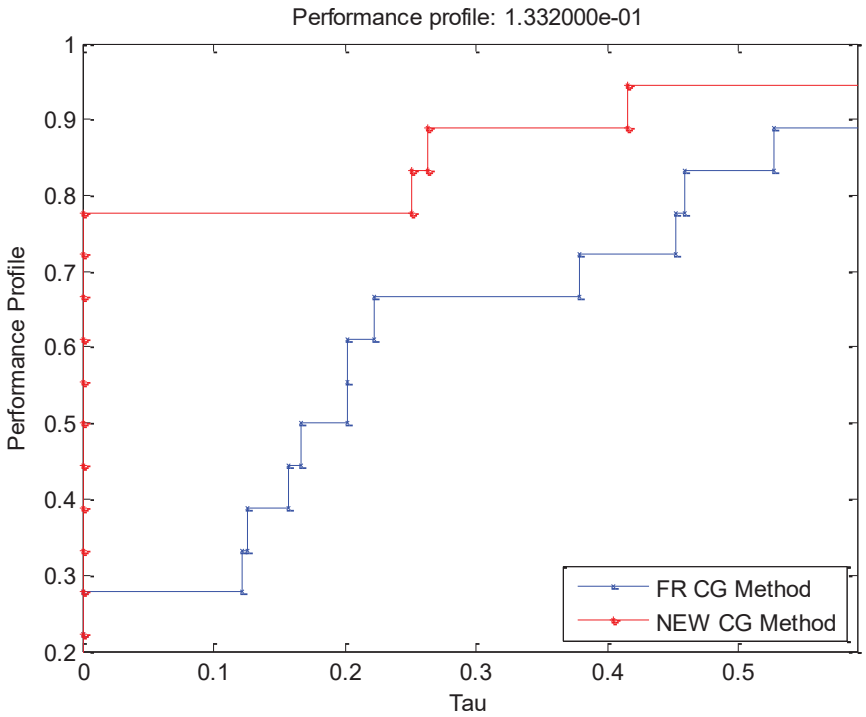

Fig.1. Performance profiles of NOF with (n=1000)

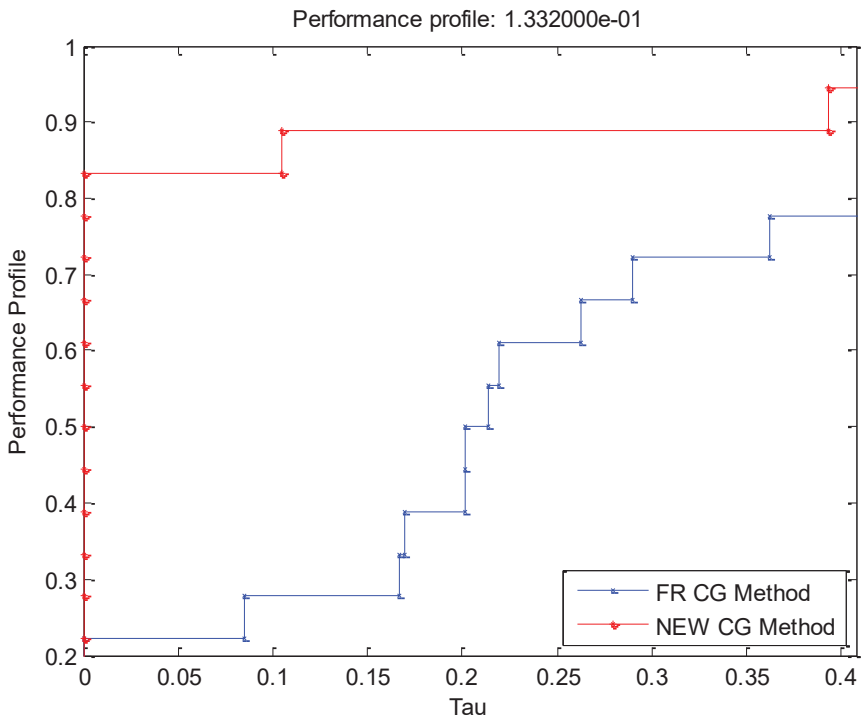

Fig. 2. Performance profiles of NOF with (n=10000)

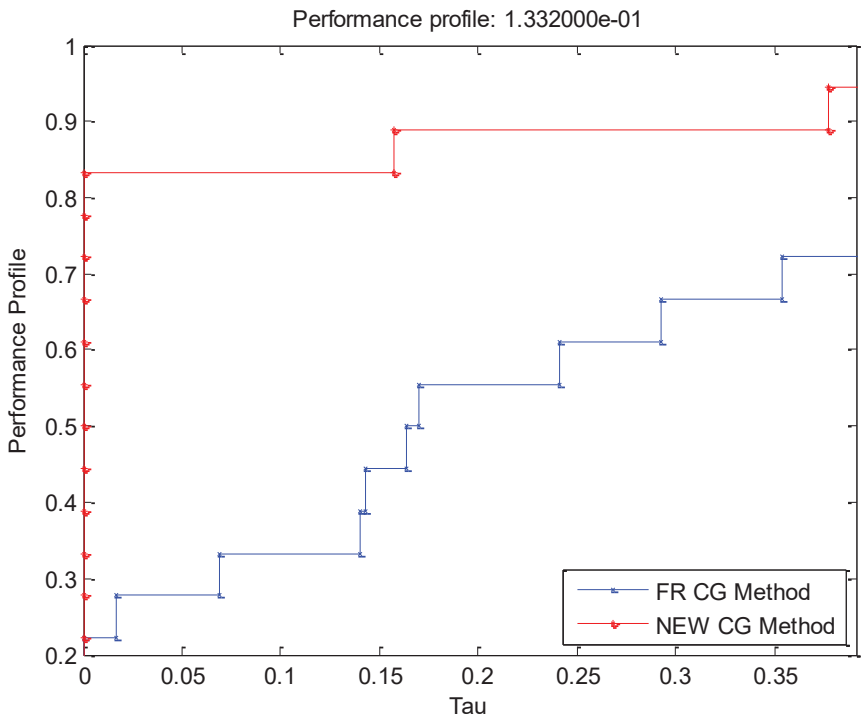

Fig. 3. Performance profiles of NOI with (n=1000)

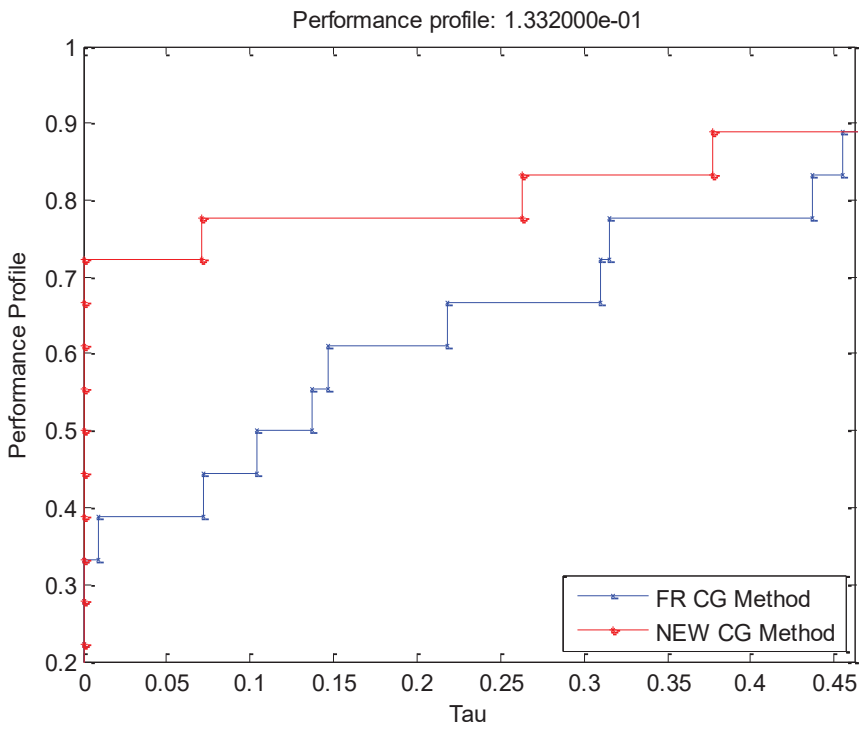

Fig. 4. Performance profiles of NOI with (n=10000)

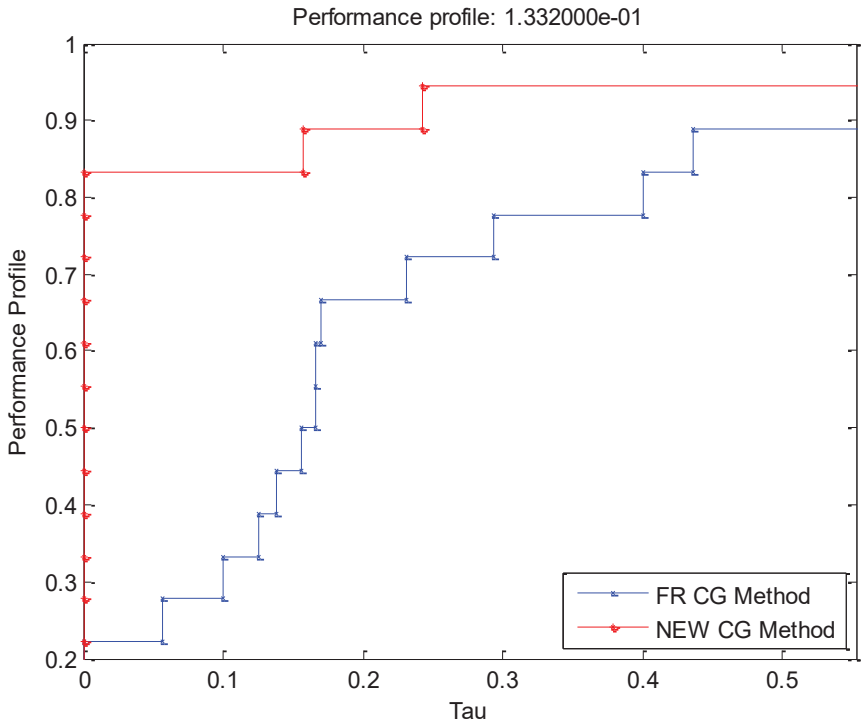

Fig. 5. Performance profiles of CPU time with (n=1000)

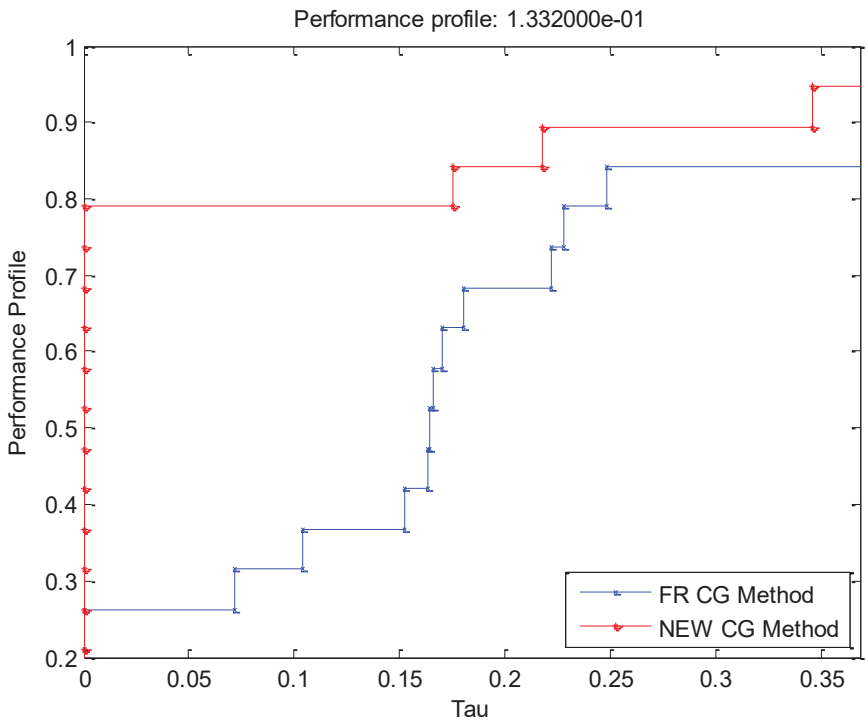

Fig. 6. Performance profiles of CPU time with (n=10000)

5 Conclusion

"Concerning the theoretical aspects of our novel algorithm, we have established both sufficient descent and global convergence under specific assumptions. The numerical results depicted in the aforementioned figures showcase the effectiveness of our algorithm through a direct comparison with the standard HS CG method. "

References

1. 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, 1952.
2. R. Fletcher, and C. M. Reeves, "Function minimization by conjugate gradients," The computer journal, vol. 7, no. 2, pp. 149-154, 1964.
3. H. I. Ahmed, E. T. Hamed, and H. T. Saeed . chilmeran, " A Modified Bat Algorithm with Conjugate Gradient Method for global Optimization," International Journal of Mathematicsan Mathematical Scinces, vol. 2020,2020
4. Basim A. Hassan M. Sadiq, (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 (Bulletin SUSUMMCS), 2022, vol. 15, no. 4, pp. 123-130 .
5. H.I. Ahmed, R.Z. Al-Kawaz and A.Y. Al- Bayati, "Spectral three-term constrained conjugate gradient algorithm for function minimizations," Hindawi J. Appl. Math., vol 2019,.https://doi.org/10.1155/2019/6378368, 2019.
6. E.T. Hamed, H.I. Ahmed, H. Y. Najm, "Global Convergence of Conjugate Gradient Method inUnconstrained Optimization Problems, "International Conference of Mathematical Sciences (ICMS 2018) AIP.

7. P. Wolfe, , " Convergence conditions for a scent methods II: some corrections". SIAM Review, Vol.13, No. 2,1971, pp.185-188.
8. Ibrahim, M. A. H. Bin, Mamat, M. and June, L. W. (2014) " BCGF method: Anew search direction", Sains Malayasiana, 43(10), pp. 1591-1597.
9. Najm, H. Y., & Ahmed, H. I. (2022, December). Improving Jellyfish Swarm Algorithm for Solving Unconstrained Optimization Problems. In 2022 3rd Information Technology To Enhance e-learning and Other Application (IT-ELA) (pp. 226-231). IEEE.
10. Powell, M. J. D., (1976a), "Rstart Procedure For The conjugate Gradient Methods, Mathematical programming", 12, 241-254 .
11. H. Yabe, and M. Takano, "Global convergence properties of nonlinear conjugate gradient methods with modified secant condition," Computational optimization and Applications, vol. 28, no.2, pp. 203-225, 2004.
12. Andrei, N. (2020) Nonlinear conjugate gradient methods for unconstrained optimization, Springer Optimization and It's a'pplications.
13. Andrei, N. (2008) 'An Unconstrained Optimization Test Functions Collection', Advanced Modelling and Optimization, 10(1), pp. 147–161
14. E. D. Dolan, and J. J. Moré," Benchmarking optimization software with performance profiles," Mathematical programming, vol. 91, no. 2, pp. 201-213, 2002