# Another Conjugate Gradient Algorithm based on Spectral-scaled Memoryless BFGS Update

I. A. Osinuga
Department of Mathematics,
Federal University of Agriculture, PMB 2240,
Abeokuta, Nigeria

Y. N. Nwodo
Department of Mathematics,
Federal University of Agriculture, PMB 2240,
Abeokuta, Nigeria

## ABSTRACT
In this study, we present another modification of a scaled three-term conjugate gradient (CG) algorithm. The proposed method incorporates the BFGS updating scheme of the inverse Hessian approximation within the frame of a memoryless quasi-Newton approach. In this case, the inverse Hessian approximation is restarted as a multiple of identity matrix with a spectral scaling parameter at every iteration. Under standard Wolfe line search, numerical results from an implementation of the proposed method indicate that the method is promising and competitive when subjected to comparison with other state-of-the art similar algorithms available in literature utilizing performance profiles of Dolan and More.

## Keywords
Unconstrained optimization, conjugate gradient method, spectral-scaled memoryless BFGS, numerical comparisons

## 1. INTRODUCTION
Conjugate gradient algorithm is a class of many tools popularly used for solving large-scale optimization problems due to its relatively single programs and low memory requirements. The quest to construct efficient CG algorithms has prompted researchers to device several techniques for obtaining different CG algorithms. This is because of its applications in astronomy, signal processing, meteorology and energy problems such as minimization of power losses in transmission lines [1, 2, 3]. Hence it is required to develop new method to solve large-scale unconstrained optimization problems. One efficient approach is a generalization of the CG method otherwise referred to as the spectral CG method. The primary objective of this paper is to study the performance of a nonlinear spectral-scaled method for unconstrained optimization satisfying descent and sufficient descent conditions.

Consider the following unconstrained optimization problem

$$\min\{ f(x) : x \in \Re^n \}, \quad \text{----} \quad (1)$$

where $f : \Re^n \to \Re$ is continuously differentiable and its gradient is available. The iterates of the classical CG algorithm can be formulated as

$$x_{k+1} = x_k + \alpha_k d_k \quad \text{----} \quad (2)$$

and

$$d_{k+1} = -g_{k+1} + \beta_k d_k, d_0 = -g_0 \quad \text{----} \quad (3)$$

where $g_k$ is the gradient of $f(x)$ at the point $x_k$, $d_k$ is the search direction, $\beta_k$ is the so-called conjugate parameter, and $\alpha_k$ is the positive scalar step-size which is determined by some line search. For instance, Hestenes and Stiefel (HS) [ 4], Fletcher and Reeves (FR) [5 ], Polak-Ribiere- Polyak (PRP) [6, 7], Liu and Storey (LS) [8 ], Dai and Yuan (DY) [9 ], and Fletcher (CD) [10] used conjugate parameter, respectively, given by

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2},$$

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}, \beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T y_k},$$

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}, \beta_k^{CD} = \frac{\|g_{k+1}\|^2}{-d_k^T y_k} \quad \ldots (4)$$

where $y_{k+1} = g_k - g_{k+1}$ and $\| \ \|$ denotes Euclidean norm of vectors. As it is well known that the choice of $\beta_k$ affect the numerical performance of the method, hence many researchers studied choices of $\beta_k$. The CG algorithms, based on $\beta_k$ computation, can be classified as classical, hybrid, scaled and parametric [11, 28]. The classical algorithms are defined by (2) and (3), where the CG parameter is computed as in (4). Modified classical algorithms are abound in literature, consult [12-16], to mention a few.

However, one of the earliest developed three term classical CG method may be found in Beale [17] as another important innovation to CG methods. In recent time, Babaie - Kafaki and Ghanbari [15] gave an extension of the three - term CG method proposed by Zhang et al. [33]. Taqi [12], developed a three - term CG algorithm for training feed - forward neural networks which was a vector based training algorithm derived from DFP quasi - Newton and has 0(n) memory. Application of the three - term CG method to regression analysis was reported by Moyi et al. [3]. Hybrids have been derived to exploit the exciting features of the classical algorithms using projective manner [20 -22], consideration of linear and convex combination of classical schemes [31] as well the use of notion involving the classical CG and quasi-Newton methods which started with Buckley [19 ]. Several others in this category can be found in

[18-19, 30]. Another innovation to the so-called three-term CG method is the case in which the search direction is determined as a linear combination of $g_k, s_k$ and $y_k$ as

$$d_k = -g_k - t_1 s_k + t_2 y_k$$

where $t_1$ and $t_2$ are scalars. CG methods in this category are termed scaled three-term and can be found in [25-26, 29] and references therein.

The proposed spectral-scaled three-term method studied in this paper incorporates the BFGS updating scheme of the inverse Hessian approximation within the frame of a memoryless quasi-Newton approach. In this case, the inverse Hessian approximation is restarted as a multiple of the identity matrix with a spectral scaling parameter in every iteration.

The rest of the paper is organized as follows. In Section 2, we present another spectral scaled three-term CG and prove its descent property without any line search. Section3 presented preliminary numerical results under standard Wolfe line search using the performance profiles of Dolan and More. Finally, Section 4 presents our concluding remarks.

## 2. DERIVATION OF THE SPECTRAL-SCALED THREE-TERM CG ALGORITHM

This section deals with derivation of the spectral-scaled (ON) method. Motivated by the methods proposed in [25, 26, 34], we state the ideas to propose a spectral-scaled CG method as follows.

According to Broyden, Fletcher, Goldfarb and Shanno (1970) and reported in Andrei [ 25], one of the most efficient quasi-Newton methods for solving small and medium sized unconstrained problem is the BFGS method. Likewise the theory behind the method and its global convergence are very well established [25]. The proposed method is derived by utilizing the BFGS update for the inverse Hessian approximation which preserves the sufficient descent conditions. Unlike Arzuka et al. [26] the quasi-Newton update is restarted with a multiple of the identity matrix with a spectral scaling parameter. As usual the search direction in the quasi-Newton method is given by

$$B_k d_k = -g_k \quad \text{---} \quad (5)$$

and $g_k$ is as defined above. The matrix $B_k$ is the BFGS approximation to the Hessian $\nabla^2 f(x_k)$ of $f$ at $x_k$ , being updated by the formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k s_k^T} \quad \text{---} \quad (6)$$

where $s_k = x_{k+1} - x_k, y_k = g_{k+1} - g_k$ and $B_0$ being symmetric and positive definite. However, the search direction is computed in practical implementation as

$$d_k = -H_K g_k \quad \text{---} \quad (7)$$

If $H_k$ is updated by the BFGS then

$$H_{k+1} = H_k - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} + \frac{y_k y_k^T}{y_k s_k^T} \quad \text{--- (8)}$$

such that the secant equation

$$H_k y_k = s_k \quad \text{---} \quad (9)$$

is satisfied. The resulting BFGS update scheme takes the form adaptively from [26]

$$Q_{k+1} = \theta_k I - \theta_k \frac{s_k s_k^T}{s_k^T s_k} + \frac{y_k y_k^T}{y_k s_k^T} \quad \text{---} \quad (10)$$

and thus, the search direction is defined by

$$d_{k+1} = -Q_{k+1} g_{k+1} = -\theta_k g_{k+1} - \theta_k \frac{g_{k+1} s_k s_k^T}{s_k^T s_k} + \frac{y_k g_{k+1} y_k^T}{y_k s_k^T} \quad (11$$

The proposed method new search direction is defined by

$$d_{k+1} = -\theta_k g_{k+1} - t_1 s_k + t_2 y_k \quad (12)$$

where

$$\theta_k = \frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2} [32,34], t_1 = \theta_k \frac{g_{k+1} s_k^T}{s_k^T s_k}, t_2 = \frac{g_{k+1} y_k}{s_k^T y_k} \quad (13)$$

In this paper, we solve (1) using a new iterative scheme in which the iterative point is generated by (12), where $\theta_k$ is the spectral gradient parameter. Obviously, if $\theta_k = \mu_k$, it reduces to Arzuka et al. [26] provided the updating scheme is DFP and $\mu_k$ is due to Wolkowicz (1994) as reported in [26]. In this method, the parameter $\theta_k$ is selected in such a way that sufficient descent condition is guaranteed.

Pre-multiplying by $g_{k+1}^T$, this gives

$$g_{k+1}^T d_{k+1} = -\theta_k \|g_{k+1}\|^2 - t_1 s_k g_{k+1}^T + t_2 y_k g_{k+1}^T \quad (14)$$

$$= -\|g_{k+1}\|^2 \frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2} - \theta_k \frac{g_{k+1} s_k^T}{s_k^T s_k} s_k g_{k+1}^T + \frac{g_{k+1} y_k^T}{s_k^T y_k} y_k g_{k+1}^T \quad (15)$$

$$= \|g_{k+1}\|^2 \frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2} \left(\frac{-\|g_{k-1}\|^2}{d_{k-1}^T y_{k-1}} \theta_k - \frac{\|g_{k-1}\|^2}{d_{k-1}^T y_{k-1}} \cdot \frac{\|g_{k+1}\|^2 s_k^T}{s_k^T s_k} \theta_k s_k + \frac{\|g_{k-1}\|^2}{d_{k-1}^T y_{k-1}} \cdot \frac{\|g_{k+1}\|^2 y_k^T}{s_k^T y_k} y_k\right) \quad (16)$$

then we have

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} = \frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2}\ \eta_k \qquad \text{---} \qquad (17)$$

where

$$\eta_k = \frac{-\|g_{k-1}\|^2}{d_{k-1}^T y_{k-1}}\theta_k - \frac{\|g_{k-1}\|^2}{d_{k-1}^T y_{k-1}}.\frac{\|g_{k+1}\|^2 s_k^T}{s_k^T s_k}\theta_k s_k$$

$$+ \frac{\|g_{k-1}\|^2}{d_{k-1}^T y_{k-1}}.\frac{\|g_{k+1}\|^2 y_k^T}{s_k^T y_k} y_k$$

$$g_{k+1}^T d_{k+1} = -\frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2}\|g_{k+1}\|^2 \eta_k$$

Assuming that $\eta_k \equiv 1$ holds for any $k \geq 1,$ it follows from (3) and (17) that

$$= -\|g_{k+1}\|^2, \forall k \geq 0$$

i.e $g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2, \forall k \geq 0$

Next, we present specific algorithm for ON method

**Algorithm 2.1 (ON algorithm)**. Step 0: Give the initial point $x_0 \in \Re^n$ and set $\sigma, \varepsilon, \delta > 0.$ Set $k = 0$.

Step 1: If $\|g_k\| \leq \varepsilon,$ stop.

Step 2: Determine $\alpha > 0$ using the standard Wolfe line search:

$$f(x_k + \alpha_k d_d) \leq f(x_x) + \delta\alpha_k g_k^T d_k, \text{---} \qquad (18)$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \qquad \text{--} \qquad (19)$$

where $0 < \delta < \sigma < 1.$

Step 3: Let the next iterate be $x_{k+1} = x_k + \alpha_k \nu_k d_k,$ where $\nu_k$ are computed according to ([26], page 4) else $x_{k+1} = x_k + \alpha_k d_k$

Step 4: Generate the next direction $d_{k+1}$ by (12 ) where $\theta_k, t_1$ and $t_2$ are computed by (13 )

Step 5: Let $k := k + 1,$ go to step1

## 3. NUMERICAL EXPERIMENTS

In this section, by numerical experiments, we are going to study the effectiveness and robustness of Algorithm 2.1 for solving unconstrained optimization problems.

### 3.1 Benchmark test problems

All the test functions used in our computational study were drawn from the CUTE library [23] and Andrei [24]. We selected 15 large-scaled problems in extended or generalized form. Each problem is tested 9 times for a gradually increasing number of variables: $n =$ 2, 10, 100, 500, 800, 1000, … , 5000, 10000. These problems are listed below in Table 1:

**Table 1: Test Problems Functions**

| No | Name | Dimensions |
|---|---|---|
| 1. | Rayden 2 | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 2. | Quadratic QF 1 | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 3. | Extended Rosenbrock | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 4. | Rayden 1 | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 5. | Fletcher Function | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 6. | Extended PSC 1 | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 7. | Extended Himmelblau | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 8. | Extended Tridiagonal 1 | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 9. | Diagonal 5 | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 10. | Nonquadratic Function | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 11. | Hager Function | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 12. | Extended Bohr 3 | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 13. | Extended Quadratic Penalty QP 1 | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 14. | EG 2 | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |
| 15 | Diagonal 8 | 2, 50, 70, 180, 500, 863, 1000, 6500, 11400 |

### 3.2 Parameter Settings

The parameters such as number of iterations and CPU time in seconds were indicators considered to evaluate the computational capability of ON as compared with the conventional FR [5], AZ [26], AN [25], ZH [29] methods. For each test problem, the stopping criteria used is $\| g_k \| \leq 10^{-6}$. All problems implement the standard Wolfe line search with $\sigma = 0.9$ and $\delta = 0.0001$ using MATLAB R2013 with CPU 1.30 GHz and 3.00GB RAM, on SAMSUNG PC notebook.

### 3.3 Discussion of Results

The performance profiles of Dolan and More [27] was used to compare the numerical strength of the proposed method (ON) against some known CG methods such as AZ, AN, ZH and FR methods based on number of iterations and CPU time. We plot fraction ($\rho$) of the test problems for which the method is within a factor $\tau$of the best time for each method. The left

hand side of the figures gives the % of how fast is a particular method in solving the test problems. The right hand side of the figures gives the % of test problems that are successfully solved by each method. The solver with large probability is regarded as the best solver for the test problems. According to an anonymous referee's suggestion, we detail the experimental analysis as follows:

Figures 1-2, show that ON is the fastest solver on approximately 44% of the test problems (together with AZ (43%), AN (38%), ZH (18%) and FR (14%)) for iterations while AZ and ZH on approximately 25% each are the fastest solver based on CPU time follow by ON (20%), FR (8%) and AN (7%).

However, ON competes well with AZ, ZH and FR by solving 82% of the test problems with AN solving only 38% based on iterations and 83% of CPU time alongside AZ and ZH with FR solving 82% while AN exhibiting poor performance solving only 38% of the problems. Since all CG methods have been implemented with the same line search, we conclude that the efficiency and robustness of the proposed method is encouraging.

## 4. CONCLUSIONS

This paper, based on previous ideas of quasi-Newton approach and recently proposed scaled three-term CG methods, has presented another CG algorithm which satisfies descent condition and promising. Therefore, as part of future work, new hybrid methods including the one introduced in this paper and those exhibiting good global convergence properties would be developed, tested numerically and compared with highly efficient methods.

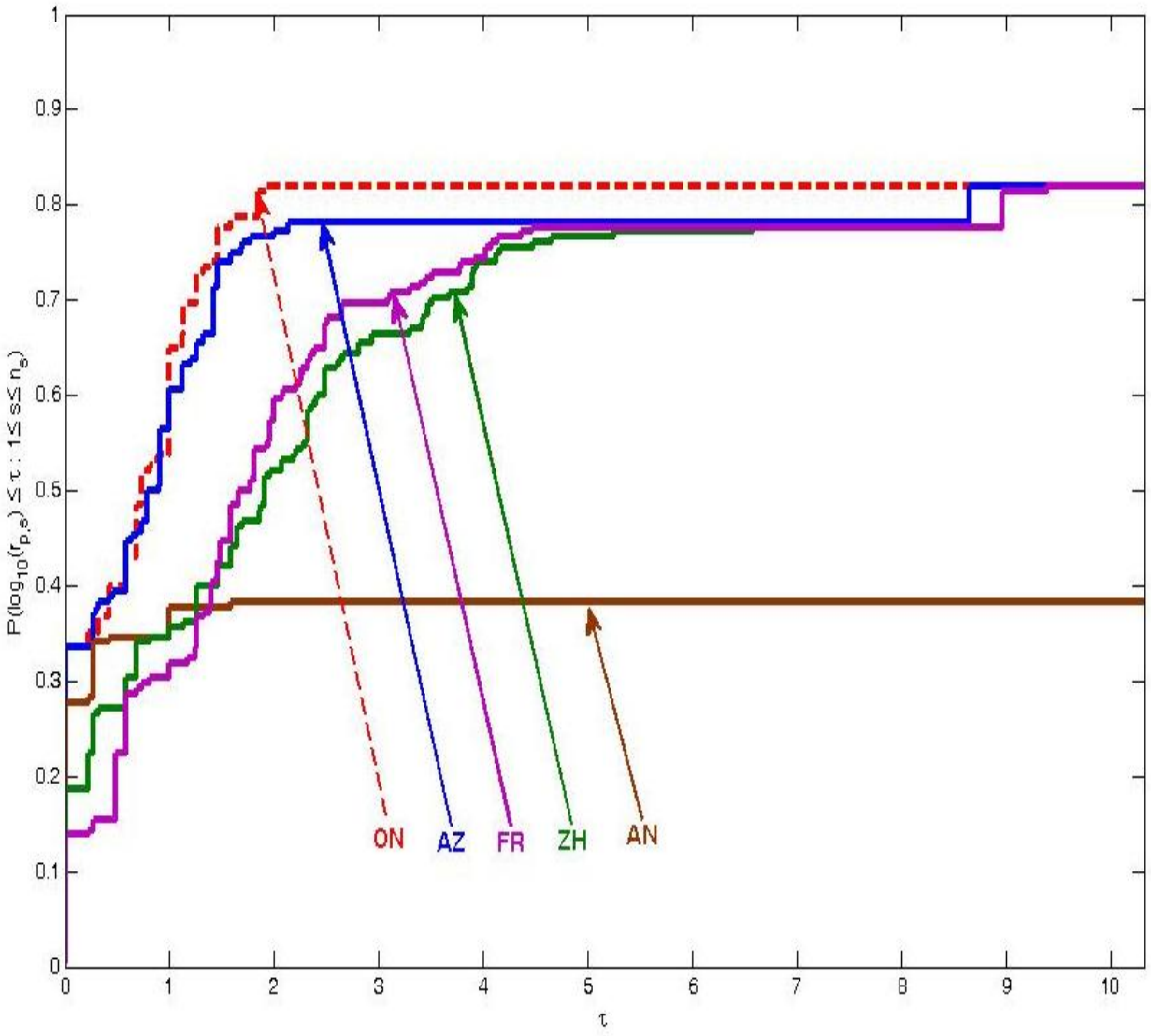


**Fig 1: Performance profiles of ON versus AZ, AN, ZH, and FR based on iterations**

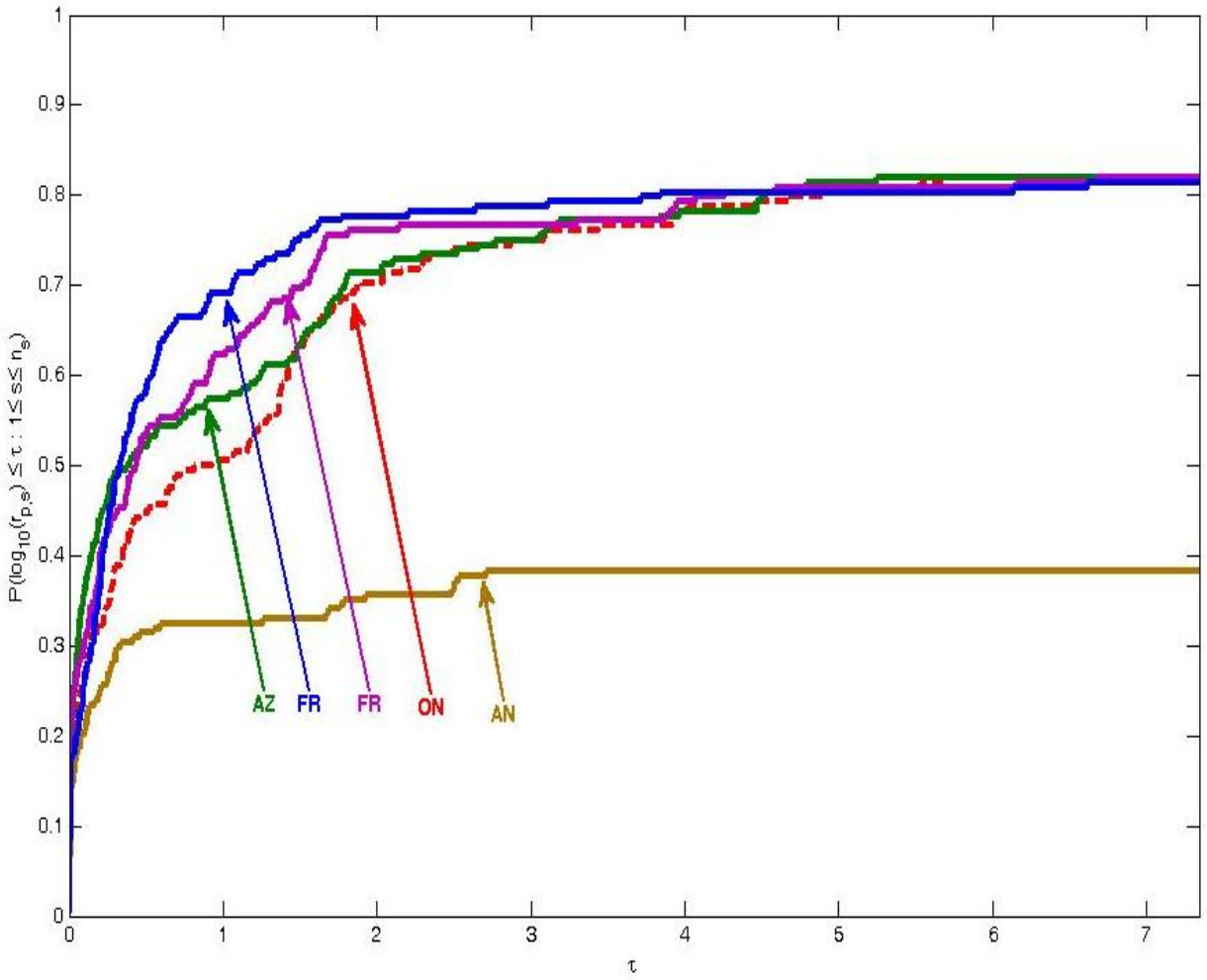


**Fig 1: Performance profiles of ON versus AZ, AN, ZH, and FR based on CPU time**

## 5. ACKNOWLEDGEMENTS

The authors would like to thank the anonymous referees for their constructive comments which helped to bring this paper in the current form.

## 6. REFERENCES

[1] Navon, M. and Legler, D. N. 1987 Conjugate gradient Methods for large-scale minimization in meteorology, descent property, Miskolc Mathematical Notes, 16 (1), 45 – 55.

[2] Bamigbola, O. M., Ali, M. M. and Oke, M. O. 2014Mathematical modeling of electric power flow and the minimization of power losses on transmission lines, Applied Mathematics and Computation, 241, 214 – 221.

[3] Moyi, A. U., Leong, W. J., Ibrahim, S. 2014 On the Application of Three - Term Conjugate Gradient Method in Regression Analysis, International Journal of Computer Applications, 102 (8), 1-4.

[4] Hestenes, M. R. and Stiefel, E. L. 1952 Method of conjugate gradients for solving linear systems,Journal of Research National Bureau Standards, 49, 409-436.

[5] Fletcher, R. and Reeves, C. 1964 Function minimization by conjugate gradients, Computer Journal, 7,149-154.

[6] Polak, E. and Ribieré, G. 1969 Note sur la convergence de directions conjugeés, Rev. Francaise Informat Recherche Operationelle, 3e Année, 16: 35-43.

[7] Polyak, B. T. 1969 The conjugate gradient method in extreme problems, USSR Computer Mathematics and Mathematical Physics, 9, 94-112

[8] Liu, Y. and Storey, C. 1991 Efficient generalized conjugate gradient algorithms, Part 1: Theory Journal of Optimization Theory Application, 69, 129-137.

[9] Dai Y. H. and Yuan Y. 2001 An efficient hybrid conjugate gradient method for unconstrained optimization, Annals of Operations Research, 103, 33-47.

[10] Fletcher, R., 1987. Practical Methods of Optimization vol. 1: Unconstrained Optimization, John Wiley & Sons, New York.

[11] Andrei N. 2007 Numerical comparison of conjugate algorithms in unconstrained optimization, Studies in Informatics and Control, 16, 333-352.

[12] Taqi, A. H. 2015 Improved three - term conjugate Algorithm for training neural network, Journal of Kufa for Mathematics and Computer, 2 (3), 93 – 100.

[13] Adeleke, O. J. and Osinuga, I. A. 2018 A five-term Hybrid conjugate gradient method with global convergence and descent properties for unconstrained optimization problems, Asian Journal of Scientific Research, 11, 185-194. DOI: 10.3923/ajsr.2018.

[14] Andrei N. 2010 Accelerated hybrid conjugate gradient algorithm with modified secant condition for unconstrained optimization. Numerical Algorithms, 54 (1), 23 – 46.

[15] Babaie – Kafaki S. and Ghanbari R. 2015 An extendedthree - term conjugate gradient method with sufficient descent property, Miskolc Mathematical Notes, 16 (1), 45 – 55.

[16] Gilbert E. G. and Nocedal J. 1992 Global convergence properties of conjugate gradient methods for Optimization, SIAM Journal of Optimization, 2, 21- 42.

[17] Beale, E. M. I. 1972 A derivative of conjugate gradients in numerical methods for nonlinear optimization, F.A. Lootsma, Ed., 39-43, Academic Press, London, UK.

[18] Ibrahim M. A. H., Mustafa M., Leong W. J. 2014 The hybrid BFGS - CG method in solving unconstrained optimization problems,. Abstract and Applied Analysis, Vol. 2014 Article ID 507102, 6 pp

[19] Buckley, A. G., 1978 A combined conjugate-gradient quasi-Newton minimization algorithm, Mathematical Programming, 15, 200 – 210.

[20] Ibrahim M. A. H., Mustafa M., Leong W. J., Azfi Z. S. 2014 The Algorithms of Broyden – CG for unconstrained optimization problems, International Journal of Mathematical Analysis, 8 (52), 2591 – 2600, http://dx.doi.org/10.12988/ijma.2014.49272

[21] i J. K.. and Li S. J. 2014 New hybrid conjugate Gradient Method for Unconstrained Optimization, Applied Mathematics Computation, 245, 36 – 43.

[22] Touati - Ahmed D., Storey C. 1990 Efficient hybrid conjugate gradient techniques, J. Optim. Theory Appl. 64 (2), 379 – 397.

[23] Andrei, N. 2008 An unconstrained optimization test functions collection. Advanced Modelling and Optimization, 10 (1), 147- 161.

[24] Bongart, I, Conn, A. R., Gould, N. I. M. and Toint, P. L. 1995, CUTE: constrained and unconstrained testing environments, ACM TOMS, 21, 123-160.

[25] Andrei, N. 2013 On three-term conjugate gradient algorithm for unconstrained optimization, Applied Mathematics and Computation, 219 (11), 6316 – 6327.

[26] Arzuka, I., Abubakar, M. R. and Leong, W. J. 2016 A scaled three-term conjugate gradient method for unconstrained optimization, Journal of inequalities and applications, 325, DOI.10-1186/s13660-016-1239-1

[27] Dolan, E., More, J. J. 2002 Benchmarking optimization software with performance profile, Mathematical Programming", 91, 201 – 213.

[28] Hager, W. W. and Zhang, H. 2000 Asurvey of nonlinear conjugate gradient methods, Pacific Journal of Optimization, 2, 35 – 58.

[29] Zhang L., Zhou W. and Li D. 2007 Some descent conjugate gradient method and their global convergence, Optim. Methods Softw., 22 (4), 697 – 711.

[30] Wan Osman W. F. H., Ibrahim, M. A. H., Mamat M. 2017 Hybrid DFP-CG method for solving Unconstrained optimization problems", Journal of Physics: Conference Series, 890, DOI:10.1088/1742-6596/890/1/012033.

[31] Xu X. and Kong F. 2016 New hybrid conjugate gradient methods with the generalized Wolfe line search". Springerplus, 5:881.

[32] Zhang L., Zhou W. and Li D. 2006 Global convergence of a modified FR conjugate gradient method with Armidjo-type line search, Numer. Math., 2006, 104, 561 – 572.

[33] Zhang L., Zhou W. and Li D. 2006 A descent modified PRP conjugate gradient method and its global convergence, IMA J. Numer. Anal., 26, 629 – 649.

[34] Liu, J. K., Feng, Y. M. and Zou, L. M. 2019 A spectral conjugate gradient method for solving large-scale unconstrained optimization, Computers and Mathematics with Applications, 77, 731-739.