

Computation and Performance Analysis of Existing Three Term Conjugate Gradient Method

Nur Idalisa¹, Nor Farzana Buchek², Nurul Syakirah Norizan³, Nurul Hafawati Fadhilah⁴, Nur Anisah Najwa Mohd Arizi⁵

¹Universiti Teknologi MARA, Malaysia, nuridalisa@uitm.edu.my

²Universiti Teknologi MARA, Malaysia, farzanabuchek97@gmail.com

³Universiti Teknologi MARA, Malaysia, nurulsyakirahnorizan@gmail.com

⁴Universiti Teknologi MARA, Malaysia, nurulhafawatifadhilah@gmail.com

⁵ University of Technology MARA, Malaysia, anisahnajwa483@gmail.com



ABSTRACT

The three-term Conjugate Gradient (CG) method is employed to solve unconstrained optimization problem (UOP) according to its efficiency and numerical performances. The three-term CG methods used were three-term Hestenes-Stiefel (TTHS) and three term new RMIL (3TNRMIL). The efficiency of the method is shown through comparison with all the three-term CG methods. The original three-term is compared with its modified scaling initial direction. The total test run involved in the numerical experiment was 800 that comprised of 5 test functions. The five test functions that were selected were Extended Himmelblau, Extended Rosenbrock, Extended White and Holst, Extended Beale and Shallow Functions with dimensions of 1200, 3000, 6000, and 9000. The performance profile is analysed based on number of iteration (NOI) and CPU time. Result shows that the modified TTHS performed better than the original TTHS and 3TNRMIL.

Key words: Conjugate gradient method, three term, scaling CG, unconstrained optimization.

1. INTRODUCTION

The Unconstrained Optimization Problem (UOP) can be formulated as:

$$\min\{f(x) : x \in \mathcal{R}^n\}. \quad (1)$$

In order to generate the sequence of solutions for the above UOP, the iterative method of the following is used as below:

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

where the step size, $\alpha_k > 0$ is the step-size.

The CG method is well known as the effective iterative method for solving UOP and its search direction d_k is generated by

$$d_{k+1} = \begin{cases} -g_k & \text{if } k < 0 \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 1 \end{cases} \quad (3)$$

2. LITERATURE REVIEW

The CG method is used in order to enhance the behavior of the Steepest Descent (SD) method. There are many types of CG methods that differ in term of CG coefficient. The list below is some of the formula for classical CG coefficient where g is denoted as the gradient. Some of the earlier formula for CG coefficients that have been proposed and most referred to were Hestenes and Stiefel (HS) [1] in 1954, Fletcher and Reeves (FR) [2] in 1964, and Polak, Ribiere and Polyak (PRP) [3] in 1969.

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_k^T (g_k - g_{k-1})} \quad (4)$$

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \quad (5)$$

$$\beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}} \quad (6)$$

It is vital for any CG methods that have been proposed to possess global convergence properties and satisfy a sufficient descent condition. It is also advantageous if the method could fulfil the angle conditions as well as the linear convergence rate. Zoutendijk [4] proved that FR is converge globally under exact line search, but was later refuted by Powell [5]. The shortcomings also resulted from the poor performance which may burden the computation time due to large number of iterations if using FR method. Meanwhile HS and PRP failed to converge globally under exact line search even though their numerical performance are better than FR. Motivated by these facts, numerous researches were conducted to improvise those three existing CG with good performances and global convergence properties. The modified version could be using the alternative step size through inexact line search, the scaled CG method and the hybrid CG. Meanwhile, the CG coefficient proposed by Rivaie, Mustafa, Ismail and Leong in 2012 [6] known as RMIL has been proven to converge globally under exact line search for which its performance is

in between FR and HS/PRP. The RMIL is given as follows [6]:

$$\beta_k^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (d_{k-1} - g_k)} \quad (7)$$

The three-term CG methods are very prevalent among researchers due to their good descent and convergent property [7]. Reference [8] stated that the purpose of choosing three-term CG method is to improve the performance of existing CG methods according to their efficiency and numerical performance.

Reference [9] proposed a new three-term CG with exact line search using CG parameter of FR, HS and RMIL. The RMIL method which has been applied into the new proposed three term search direction is denoted as 3TNRMIL and performed better than FR and HS.

Reference [10] were doing a research on a new three-term CG method with exact line search where the initial search direction is the scaling of the original SD method. The CG coefficient proposed by [11] is adopted in proposing the new CG method in [10] because it has been proven to satisfy the sufficient descent condition. The following is the modified version of CG search direction formula [11] with parameter $\gamma \in (0,1)$:

$$d_{k+1} = \begin{cases} -g_k + \gamma g_k & \text{if } k < 0 \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 1 \end{cases} \quad (8)$$

It is agreed that exact line search is expensive and impractical [12-14]. Therefore, another option is to employ inexact line search to compute a step-length that can reduce the value of f at each iteration. Hence the strong Wolfe line search is a suitable option to control the inexactness of the line search given by the following :

$$\begin{cases} |g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k \\ \text{where } 0 < \delta < \sigma < 1 \end{cases} \quad (9)$$

Generally, a good CG method is independent with any line searches. In 2007, Zhang et.al [8] proposed a three-term HS (TTHS) CG method with the evidence that it possesses global convergence property under the standard Wolfe line search. This paper is intended to further analyse the performance of the already proven three-term CG methods which are TTHS and 3TNRMIL. The comparison involves the performance of TTHS, 3TNRMIL and their modified version by adding the scaling initial direction inspired from [11] using Strong Wolfe line search. The modified three-term CG of TTHS and 3TNRMIL are named as Method 1 and Method 2 respectively. The general algorithm for the method is presented in Section 3. The following are the methods used in this paper:

a) The search direction of TTHS [8] :

$$d_k = \begin{cases} -g_k & \text{if } k = 0 \\ -g_k + \beta_k^{HS} d_{k-1} + \theta_k y_{k-1} & \text{if } k \geq 1 \end{cases} \quad (10)$$

$$\text{where } \theta_k = \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} \quad \text{and } y_{k-1} = g_k - g_{k-1}.$$

b) The search direction of 3TNRMIL [9]:

$$d_k = \begin{cases} -g_k & \text{if } k = 0 \\ -g_k + \beta_k^{RMIL} d_{k-1} - \beta_k^{RMIL} \theta_k d_{k-1} & \text{if } k \geq 1 \end{cases} \quad (11)$$

$$\text{where } \theta_k = \frac{g_k^T g_{k-1}}{\|g_{k-1}\|^2}.$$

c) The modified TTHS named as Method 1:

$$d_k = \begin{cases} -g_k + \gamma g_k & \text{if } k = 0 \\ -g_k + \beta_k^{HS} d_{k-1} + \theta_k y_{k-1} & \text{if } k \geq 1 \end{cases} \quad (12)$$

$$\text{where } \theta_k = \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} \quad \text{and } y_{k-1} = g_k - g_{k-1}.$$

d) The modified 3TNRMIL named as Method 2:

$$d_k = \begin{cases} -g_k + \gamma g_k & \text{if } k = 0 \\ -g_k + \beta_k^{RMIL} d_{k-1} - \beta_k^{RMIL} \theta_k d_{k-1} & \text{if } k \geq 1 \end{cases} \quad (13)$$

$$\text{where } \theta_k = \frac{g_k^T g_{k-1}}{\|g_{k-1}\|^2}.$$

3. ALGORITHM

- Step 1: Given x_0 , set $k = 0$.
- Step 2: Compute the CG coefficient based on (4) and (7).
- Step 3: Compute the search direction based on (10) to (13).
If $g_k = 0$, then stop.
- Step 4: Compute the step size using (9).
Solve the step size such as the Strong Wolfe line search conditions.
- Step 5 : Update new point, $x_{k+1} = x_k + \alpha_k d_k$.
- Step 6: Convergence test and stopping criteria.
If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| \leq \varepsilon$ then stop.
Otherwise, go to Step 3 with $k = k + 1$.

4. RESULT AND DISCUSSION

In this section, the outcome of the numerical experiment is presented. The performance of three-term CG methods were compared based on number of iterations (NOI) and CPU time using performance profile method proposed by [15]. List of test functions is presented in Table 1 with dimensions of 1200,

3000, 6000, and 9000. The total test run involved in the numerical experiment was 800 that involved 5 test functions with 4 large dimensions, 4 initial points, two existing three-term CG methods and its modified version with scaling of $\gamma = 0.2, 0.4, 0.6, 0.8$. All algorithms were implemented under Strong Wolfe line search and using MATLAB R2016a software and ran on an Intel® Core i3 with 4 GB RAM memory.

Table 1: List of Test Functions and Initial Points

Test Function	Initial Points
Extended Himmelblau	(10, ..., 10), (50, ..., 50), (100, ..., 100)
Extended Rosenbrock	(2, ..., 2), (4, ..., 4), (21, ..., 21), (36, ..., 36)
Extended White and Holst	(-3, ..., -3), (3, ..., 3), (6, ..., 6), (9, ..., 9)
Extended Beale	(1, ..., 1), (3, ..., 3), (8, ..., 8), (10, ..., 10)
Shallow	(10, ..., 10), (25, ..., 25), (50, ..., 50), (100, ..., 100)

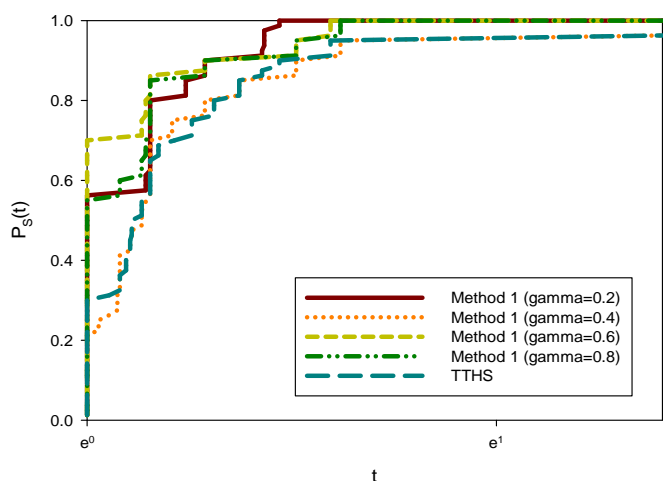


Figure 1: Performance profile of iteration number

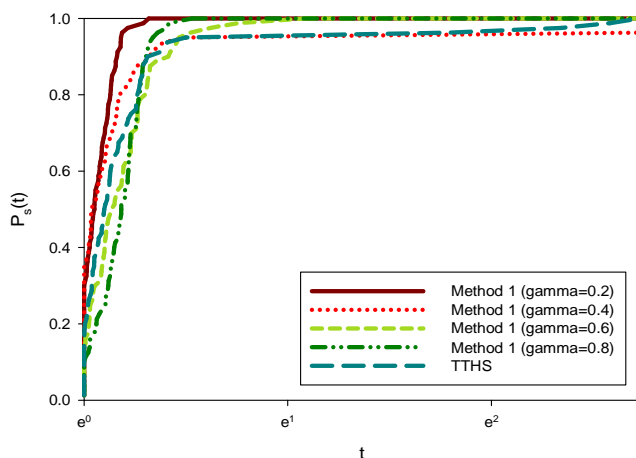


Figure 2: Performance profile of CPU time

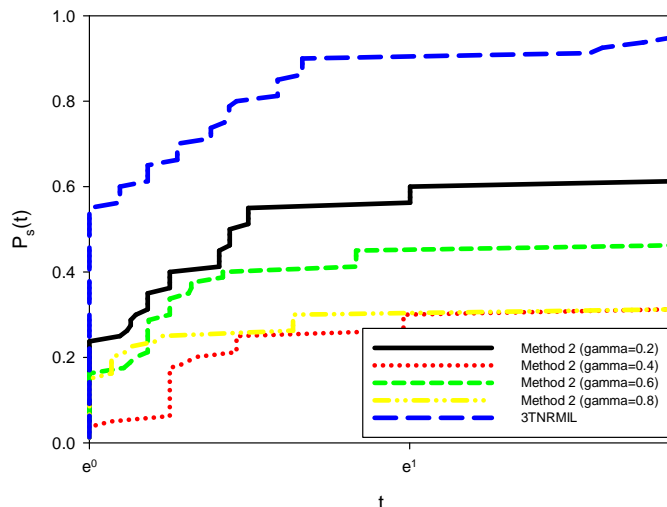


Figure 3: Performance profile of iteration number

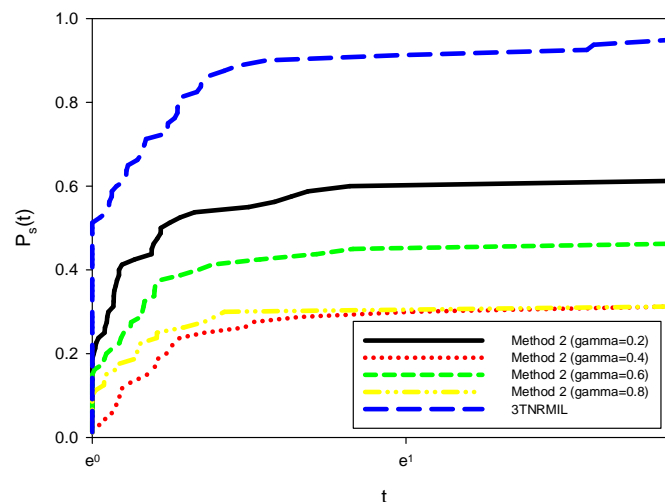


Figure 4: Performance profile of CPU time

From Figure 1 and Figure 2, it is shown that Method 1 which is the method modified from TTHS has the best performance in terms of NOI and CPU times. Figure 3 and Figure 4 show that the 3TNRMIL has better performance compared to the modified 3TNRMIL which is also known as Method 2.

Therefore, it can be concluded that the scaling initial direction is not applicable to 3TNRMIL but surprisingly, successful to improve TTHS. However, based on literature, the performance of HS is better than RMIL. Figure 5 validated this fact and confirmed that Method 1 is superior to the original TTHS as well as 3TNRMIL.

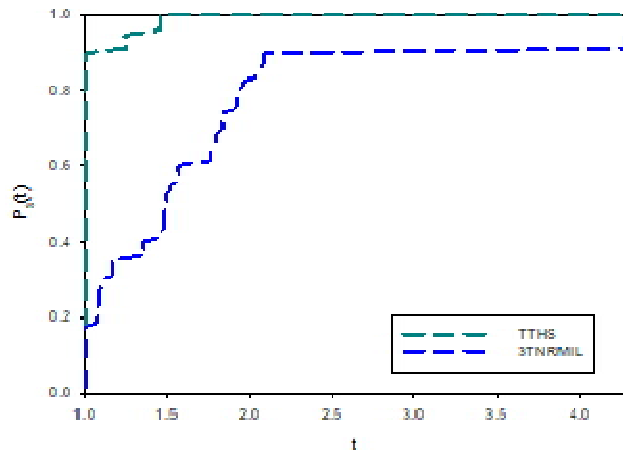


Figure 5: Performance profile between TTHS and 3TNRMIL method based on number of iteration

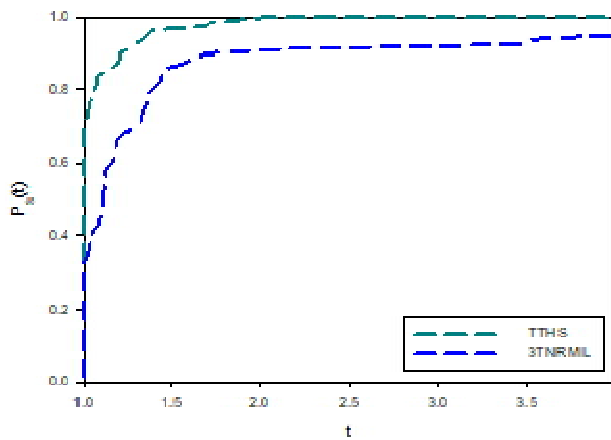


Figure 6: Performance profile between TTHS and 3TNRMIL method based on CPU time

5. CONCLUSION AND FUTURE RESEARCH

The performance of all methods is based on the comparison of efficiency performed by the methods based on number of iteration and CPU time. The total test run involved in the numerical experiment is 800 that involved 5 test functions with 4 large dimensions, 4 initial points, two existing three-term CG methods and its modified version with scaling of $\gamma = 0.2, 0.4, 0.6, 0.8$.

The researchers would like to recommend further studies on the aspect of line search selection and the behaviors of test functions. The need to explore different modified version of three-term RMIL CG methods is also highly recommended as this will help establish an efficient CG method that fulfill the global convergence properties. This is due to the fact that CG method can be employed in various application such as a training algorithm in neural network apart than the back propagation algorithm and Levenberg Marquardt algorithm [16].

ACKNOWLEDGEMENT

The authors would like to express their sincere appreciation to Dr. Mohd Rivaie Mohd Ali who provided guidance and valuable advice in this research.

REFERENCES

1. M. R. Hestenes and E. Stiefel, **Methods of Conjugate Gradients for Solving Linear Systems**, *J. Res. Nat. Bur. Stand.* Vol. 49, pp. 409-436, 1952. <https://doi.org/10.6028/jres.049.044>
2. R. Fletcher and C. M. Reeves, **Function Minimization by conjugate gradients**, *The Computer Journal*. 1964. <https://doi.org/10.1093/comjnl/7.2.149>
3. E. Polak et G. Ribiere, **Note sur la convergence de méthodes de directions conjuguées**, *Rev Francaise Informmat Recherche Opertionelle*. Vol. 16, pp. 35-43, 1969 <https://doi.org/10.1051/m2an/196903R100351>
4. G. Zoutendijk, **Nonlinear programming computational methods**, in *Integer and nonlinear programming*, J. Abadie, Ed. North Holland, Amsterdam, 1970.
5. M.J.D. Powell, **Restart procedures for the conjugate gradient method**, *Math. Program*, Vol. 12, pp. 241 – 254, 1977. <https://doi.org/10.1007/BF01593790>
6. M. Rivaie, M. Mamat, L.W. June and I. Mohd, **New conjugate gradient coefficient for large scale nonlinear unconstrained optimization**, *International Journal Of Mathematical Analysis*. Vol. 6 (21-24), pp 1131-1146, 2012.
7. L. M. Zou, Y. M. Feng and J. K. Liu, **Some three-term conjugate gradient methods with the inexact line search condition**, *Calcolo*, Vol. 55, no. 16, 2018. <https://doi.org/10.1007/s10092-018-0258-3>
8. L. Zhang, W. Zhou, and D. H. Li, **Some descent three-term conjugate gradient methods and their global convergence**, *Optim. Methods. Softw.*, Vol. 22, pp. 697-711, 2007. <https://doi.org/10.1080/10556780701223293>
9. N. H. Fadhilah, M. Rivaie, F. Ishak and N. Idalisa, **New Three Term Conjugate Gradient Method with Exact Line Search**, in *Proc. International Conf. on Mathematical Sciences and Technology (MathTech)*, Penang, 2018.
10. N. I. Norddin, M. Rivaie, N. H. Fadhilah, M. Agos and S. Nasir, **A New Sufficient Descent Conjugate Gradient Method with Exact Line Search**. in *Proc. International Conf. on Mathematical Sciences and Technology (MathTech)*, Penang, 2018.
11. S. Saha and B. Nath, **A modified form of conjugate direction for general nonlinear function and its convergence**. 7(1), 2015, pp. 25-31
12. Z.J. Shi, and J. Guo, **A new family of conjugate gradient method**. *J. Comput. Appl. Math.*, Vol. 224, pp 444-457, 2009. <https://doi.org/10.1016/j.cam.2008.05.012>

13. Z. J. Shi, S. Wang, and Z. Xu, **The convergence of conjugate gradient method with monotone line search**, *Appl.Math.Comp.*, Vol. 217, no. 1921-1932, 2010.
<https://doi.org/10.1016/j.amc.2010.06.047>
14. G. Yuan, X. Lu, and Z. Wei, **A conjugate gradient method with descent direction for unconstrained optimization**, *J. Comp. App. Maths.*, Vol. 233, no. 519-530, 2009.
<https://doi.org/10.1016/j.cam.2009.08.001>
15. E.D. Dolan and J.J. More, **Benchmarking optimization software with performance profile**, *Math. Prog.* Vol. 91, pp. 201-213, 2002.
<https://doi.org/10.1007/s101070100263>
16. M.K.Singla, J. Gupta and P. Nijhawan., **Comparative study on Backpropagation and Levenberg Marquardt algorithm on short term load forecasting**, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol 8, no 2, March- April 2019.
<https://doi.org/10.30534/ijatcse/2019/14822019>