

New Three-Term Conjugate Gradient Method with Exact Line Search

¹Nurul Hafawati Fadhilah*, ²Mohd Rivaie, ³Fuziyah Ishak and ⁴Nur Idalisa

^{1,2,4}Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA Cawangan Terengganu
Terengganu, Malaysia

^{1,3}Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Shah Alam, Malaysia

*Corresponding author: nurulhafawatifadhilah@gmail.com

Article history

Received: 17 May 2019

Received in revised form: 25 September 2020

Accepted: 28 September 2020

Published online: 1 December 2020

Abstract Conjugate Gradient (CG) methods have an important role in solving large scale unconstrained optimization problems. Nowadays, the Three-Term CG method has become a research trend of the CG methods. However, the existing Three-Term CG methods could only be used with the inexact line search. When the exact line search is applied, this Three-Term CG method will be reduced to the standard CG method. Hence in this paper, a new Three-Term CG method that could be used with the exact line search is proposed. This new Three-Term CG method satisfies the descent condition using the exact line search. Performance profile based on numerical results show that this proposed method outperforms the well-known classical CG method and some related hybrid methods. In addition, the proposed method is also robust in term of number of iterations and CPU time.

Keywords Conjugate gradient; Three-Term CG method; exact line search; unconstrained optimization

Mathematics Subject Classification 65K10, 90C30

1 Introduction

Consider the nonlinear unconstrained optimization problem as in the following form:

$$\min f(x), x \in \mathbb{R}^n \quad (1)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuously differentiable function. In order to solve (1), the iterative method is applied. The iterative method is given by

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \dots \quad (2)$$

where d_k is a search direction and α_k is a positive step size. The step size α_k is obtained by exact line search [1] which can be determined by

$$\alpha_k = \arg \min_{\alpha \geq 0} f(x_k + \alpha d_k) \tag{3}$$

where argmin stands for argument of minimum that gives minimal function value.

The general search direction d_k of the conjugate gradient (CG) method can be defined as

$$d_k = \begin{cases} -g_k & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 1, \end{cases} \tag{4}$$

where β_k is called the CG coefficient and g_k is the gradient of $f(x)$ at x_k . Some of the well-known classical formulas for β_k are Fletcher and Reeves (FR) [2], Hestenes and Stiefel (HS) [3], and Polak, Ribiere and Polyak (PRP) [4] which are given by

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \tag{5}$$

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})}, \tag{6}$$

$$\beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}, \tag{7}$$

respectively. The notation $\|\cdot\|$ refers to the Euclidean norm for vectors. Recently, there are many new formulas for β_k which contribute to accurate and efficient numerical results. Rivaie *et al.* [5] proposed a new simple formula for β_k based on the original HS method. The formula named as Rivaie, Mustafa, Ismail and Leong (RMIL) is defined as follows:

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

Since Rivaie, Mustafa, Ismail and Leong (RMIL) propose (8) by modifying the already proven Hestenes-Steifel (HS) formula thus, the RMIL method is classified as the modified CG methods [Rivaie *et al.* 6].

The classical CG methods could be improved by using hybrid CG methods. The performances of hybrid CG methods are better than classical CG methods. However, the formulas for hybrid CG methods are quite complicated and difficult to apply in the algorithm because the formulas are combinations of two or more classical methods. Some simple formulas of hybrid methods are Hu and Storey (HuS) [7] and Gilbert and Nocedal (GN) [8] given by,

$$\beta_k^{HuS} = \max \{0, \min \{\beta_k^{PRP}, \beta_k^{FR}\}\}, \tag{9}$$

$$\beta_k^{GN} = \max \{-\beta_k^{FR}, \min \{\beta_k^{PRP}, \beta_k^{FR}\}\} \tag{10}$$

respectively. In order to increase the efficiency of general search direction, Three-Term methods have been widely studied. Beale [9] was the first researcher that proposed the general Three-Term CG method. According to Baluch *et al.* [10], Three-Term CG algorithms are efficient, reliable and robust as compared to the general CG algorithms. Zhang *et al.* [11] proposed a

modified PRP method under Armijo line search. Unfortunately, it cannot be established under weak Wolfe-Powell line search. The formula is obtained by

$$d_k = -g_k + \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} d_{k-1} - \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} (g_k - g_{k-1}). \quad (11)$$

Li and Li [12] modified (11) such as

$$d_k = \begin{cases} -g_k & \text{if } k = 0, \\ -g_k + \beta_k^{PRP} d_{k-1} - \vartheta_k (g_k - g_{k-1}) & \text{if } k \geq 1, \end{cases} \quad (12)$$

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

Dong *et al.* [13] proposed another search direction by using the Gram-Schmidt orthogonalization onto d_k and g_k , in which

$$d_k = -g_k + \beta_k d_{k-1} - \beta_k \frac{g_k^T d_{k-1}}{\|g_k\|^2} g_k. \quad (13)$$

Khadijah *et al.* [14] extend the RMIL method to the Three-Term method which is globally convergent under strong Wolfe line search. The RMIL coefficient is known as one of the simple formula and easy implementation. The formula can be written as,

$$d_k = \begin{cases} -g_k & \text{if } k = 0, \\ -g_k + \beta_k^{RMIL} d_{k-1} - \theta_k (g_k - g_{k-1}) & \text{if } k \geq 1, \end{cases} \quad (14)$$

where $\theta_k = \frac{g_k^T d_{k-1}}{d_{k-1}^T (d_{k-1} - g_k)}$.

Other researches also studied Three-Term CG methods where the main purpose was to improve the performance of the current methods in terms of its efficiency and numerical performances. From the referred Three-Term methods, it is known that the studies only focus on the application of Three-Term CG method using inexact line search. However, there is still no research on Three-Term CG method that is based on an exact line search.

Theoretically, if an exact line search is used, then $g_k^T d_{k-1} = 0$. So, it is obvious that the existing Three-Term algorithms will be reduced to the general search direction when the exact line search is applied [15]. Thus, this paper strives to propose a new Three-Term CG method with exact line search. The method will improve the performance of the existing CG methods.

The remainder of this paper is organised as follows: The new Three-Term search direction CG algorithm is presented in Section 2. In Section 3, the sufficient descent condition for the proposed method will be presented. Section 4 presents the numerical experiments by implementing the algorithm to solve benchmark test problems. Lastly, the paper ends with the conclusion in Section 5.

2 New Three-Term CG Method

This section begins with the idea to propose a new Three-Term CG method. The proposed formula is similar to search direction presented by [11, 13] with different formula for parameters

θ_k and β_k . A new numerator for θ_k has been proposed in order to fit the exact line search. The new search direction is applied with the classical RMIL defined by (8) and

$$\theta_k = \frac{g_k^T g_{k-1}}{\|g_{k-1}\|^2} \quad (15)$$

where the proposed Three-Term CG method can be written as,

$$d_k^{3TNRMIL} = \begin{cases} -g_k & \text{if } k = 0, \\ -g_k + \beta_k^{RMIL} d_{k-1} - \beta_k^{RMIL} \theta_k d_{k-1} & \text{if } k \geq 1, \end{cases} \quad (16)$$

For ease of readings, $d_k^{3TNRMIL}$ is now denoted as 3TNRMIL search direction. Where, 3T is short-term of Three-Term and NRMIL is named after the contributors which are Nurul, Rivaie, Mustafa, Ismail and Leong.

The following algorithm describes the steps in order to obtain the solution for the optimization functions. For 3TNRMIL, the propose Three-Term search direction (16) is applied with RMIL method (8). Besides, for the classical and modified CG method, the search direction is computed based on (4) and formula β_k in Step 2 is based on (5), (6) or (8) which denoted as FR, HS and RMIL respectively.

Algorithm 1

- Step 1:** Initialization. Given x_0 set $k = 0$.
- Step 2:** Compute β_k based on (5), (6) or (8).
- Step 3:** Compute search direction d_k based on formula (4) or formula (16) and θ_k based on formula (15). If $g_k = 0$, then stop.
- Step 4:** Determine step size α_k by solving (3).
- Step 5:** Update new point based on (2).
- Step 6:** Convergent test and stopping criteria: If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| \leq \varepsilon$ then stop. Otherwise go to Step 2 with $k = k + 1$.

3 Convergence Analysis

In this section, the convergence of the new method is studied. For an algorithm to converge, one of the requirements is to fulfil the sufficient descent condition. The proof of the proposed method that possesses sufficient descent condition with exact line search is shown as follows.

Theorem 1

Consider a CG method with the search direction d_k as given by (16). Then the sufficient descent condition

$$g_k^T d_k \leq -C \|g_k\|^2 \text{ for } k \geq 0, \quad 0 < C \leq 1 \quad (17)$$

holds for all $k \geq 0$.

Proof

The proof is by induction. If $k = 0$, then $g_0^T d_0 = -\|g_0\|^2 \leq -C \|g_0\|^2$. Hence, condition (17) holds true. Then, it is necessary to show that for $k \geq 1$, condition (17) will also holds true.

Multiply (16) by g_{k+1}^T . Thus,

$$\begin{aligned} g_{k+1}^T d_{k+1} &= g_{k+1}^T (-g_{k+1} + \beta_{k+1} d_k - \beta_{k+1} \theta_{k+1} d_k) \\ &= -\|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k - \beta_{k+1} \theta_{k+1} g_{k+1}^T d_k. \end{aligned}$$

For exact line search, $g_{k+1}^T d_k = 0$. Then,

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

This implies that d_{k+1} is sufficient descent direction. Hence, the $g_k^T d_k \leq -C \|g_k\|^2$ holds true. The proof is completed. \square

4 Numerical Experiments

In order to support theoretical proof, the proposed method is tested using different test problems with various initial points. The list of different standard test problems with their global minimum points and function values are tabulated in Table 1. In this paper, the proposed method 3TNRMIL will be compared with the classical (FR, HS), modified CG (RMIL) and hybrid CG methods (HuS and GN). The stopping criterion is $\|g_k\| \leq 10^{-6}$ as suggested by Andrei [16]. The numerical experiment was conducted using a computer with processor Intel®Core™ i3-3217U CPU @ 1.80GHz with RAM 4GB and all the programming codes are typewritten in Maple 16. The efficiency of the methods was analysed based on the number of iterations and CPU time.

The initial points were identified from four different geometrical quadrants in order to test the behaviour of the method; whether the method converges to global solution or local solution point. For example, Extended Himmelblau function which is a multimodal function that has four different global solution points in different quadrants. The graph and contour plot of Extended Himmelblau function is presented in Figure 1. Other than that, the initial points are chosen from point closer to the solution point to the one that is furthest from it [17]. The initial points from four quadrants are as follow:

Quadrant 1: (1.25, 1.25), (10, 10), (100, 100)

Quadrant 2: (-1.25, 1.25), (-10, 10), (-100, 100)

Quadrant 3: (-1.25, -1.25), (-10, -10), (-100, -100)

Quadrant 4: (1.25, -1.25), (10, -10), (100, -100)

The results were collected by testing all selected methods on the test problem with different initial points. The classical CG method FR, HS and RMIL were applied with the standard search direction of CG method (4) and RMIL method which had been applied with the new proposed Three-Term search direction (16) is denoted as 3TNRMIL.

The numerical results are presented graphically using the performance profile that is introduced by Dolan and Moré [22]. In this performance profile, they introduced the notion of means to evaluate and compare the performance of the set solver S on a test set P . Assume that η_s solvers and η_p problems exists, for each problem p and solver s , define

$t_{p,s}$ = Computing time (the number of iterations or CPU time or others) required to solve problems p by solver s .

Table 1: Standard Test Problems

Test problem	Global minimum point	Minimum function value
Strait function with $n = 2$ [18] $f(x) = (x_2 - x_1^2)^2 + 100(1 - x_1)^2$	$x^* = (1, 1)$	$f(x^*) = 0$
Zettl function with $n = 2$ [19] $f(x) = (x_1^2 - x_2^2 - 2x_1)^2 + 0.25x_1$	$x^* = (-0.0299, 0)$	$f(x^*) = -0.0037$
Threehump function with $n = 2$ [20] $f(x) = 2x_1^2 - 1.05x_1^4 + \frac{x_1^6}{6} + x_1x_2 + x_2^2$	$x^* = (0, 0)$	$f(x^*) = 0$
Rosenbrock function with $n = 2$ [21] $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$	$x^* = (1, 1)$	$f(x^*) = 0$
Extended Beale function with $n = 2$ [21] $f(x) = (1.5 - x_1(1 - x_2))^2 + (2.25 - x_1(1 - x_2^2))^2 + (2.625 - x_1(1 - x_2^3))^2$	$x^* = (3, 0.5)$	$f(x^*) = 0$
Extended Himmelblau function with $n = 2$ [21] $f(x) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$	$x^* = (3, 2),$ $x^* = (-2.8051, 3.131),$ $x^* = (-3.779, -3.283),$ $x^* = (3.584, -1.848)$	$f(x^*) = 0$

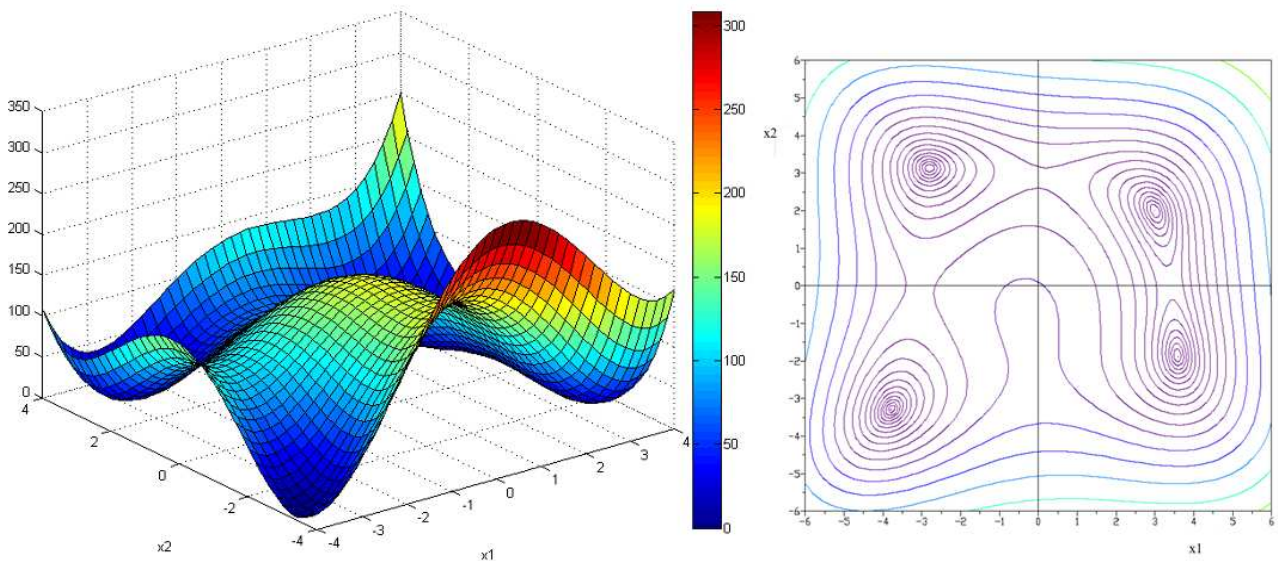


Figure 1: Graph and Contour Plot of Extended Himmelblau Function

Requiring a baseline for comparisons, they compared the performance on problem p by solver s with the best performance by any solver on this problem using the performance ratio

$$r_{p,s} = \frac{t_{p,s}}{\min \{t_{p,s} : s \in S\}}.$$

Suppose that a parameter $r_M \geq r_{p,s}$ for all p, s is chosen, and $r_{p,s} = r_M$ if and only if solver s does not solve problem p . The performance of solver s on any given problem might be of interest. To obtain an overall assessment of the performance of the solver, it is defined as

$$P_s(t) = \frac{1}{n_p} \text{size} \{p \in P : r_{p,s} \leq t\}.$$

Thus, $P_s(t)$ is the probability for solver $s \in S$ that a performance ratio $r_{p,s}$ is within a factor $t \in R$ of the best possible ration. In other words, this performance profile intends to seek the best possible method that capable to solve all given problems by evaluating the calculation of ratio values. Then, function P_s is the cumulative distribution function for the performance ratio. The performance profile $P_s = R \mapsto [0, 1]$ for a solver is a non-decreasing, piecewise, and continuous from the right. The value of $P_s(1)$ is the probability that the solver will win over the rest of the solvers. In general, a solver with high values of $P(t)$ or at the top right of the figure is preferable or represent the best solver.

The vertical line of the graph also known as right side of the graph shows the capability of method to solve all the selected test problems. This side will show the efficiency of method to solve all selected test problems by observing the value of $P_s(t)$. Otherwise, the horizontal line of the graph is also called as left side of the graph reveals the faster method compared to the rest of the solvers. The left side will expose the robustness of method by focusing on the top curve of the graph. Figure 2 until Figure 5 show the performance profile based on the number of iterations and CPU time.

Figure 2 and Figure 3 show the comparison between classical CG search direction and the new proposed Three-Term CG method in term of number of iteration and CPU time. From the figures, the proposed Three-Term method outperformed the current classical CG method since it can converge faster and able to solve all selected test problems.

Figure 2 and Figure 3 show that the 3TNRMIL is the best method rather than original RMIL, FR and HS methods. It can be clearly seen that 3TNRMIL method is placed at the top of the graph. From the right side, 3TNRMIL is capable of solving all selected test problems since its values of $P_s(t)$ equals to 1. In conclusion, the 3TNRMIL performance is better than the performance of classical CG.

In addition, the proposed search direction is compared with the hybrid CG methods, HuS and GN methods in order to see the efficiency of the newly proposed method. The performance profiles based on the number of iterations is shown in Figure 4 while Figure 5 shows the performance profile based on CPU time.

Figure 4 and Figure 5 reveal that performance curves for the 3TNRMIL method are clearly above the other curves which mean this method converges faster to the solution than HuS and GN methods. In addition, the right side of the graph shows that 3TNRMIL is capable of solving all selected problems. This concludes that the performance of 3TNRMIL is better than the performance of hybrid CG methods.

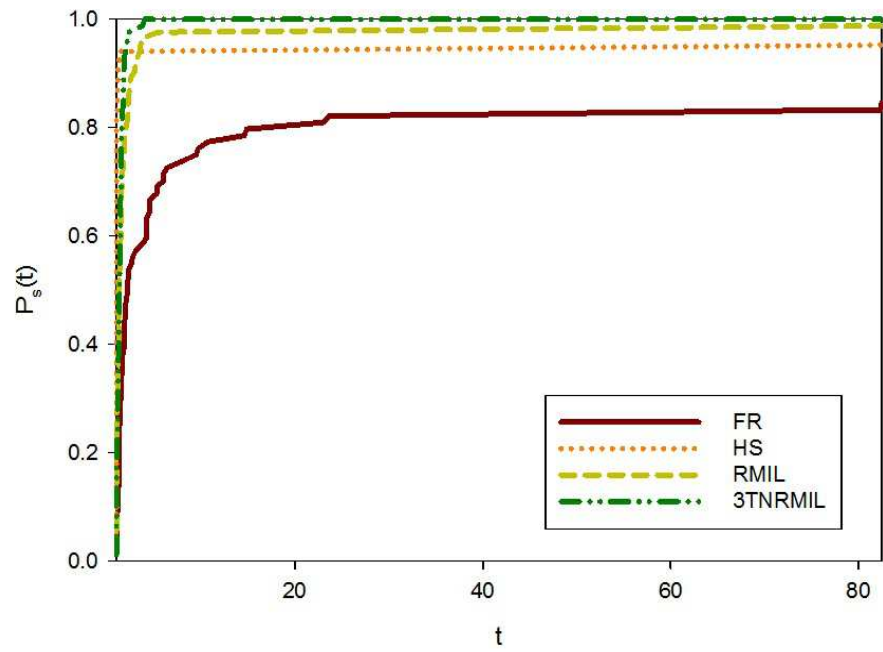


Figure 2: Performance Profile Based on Number of Iterations

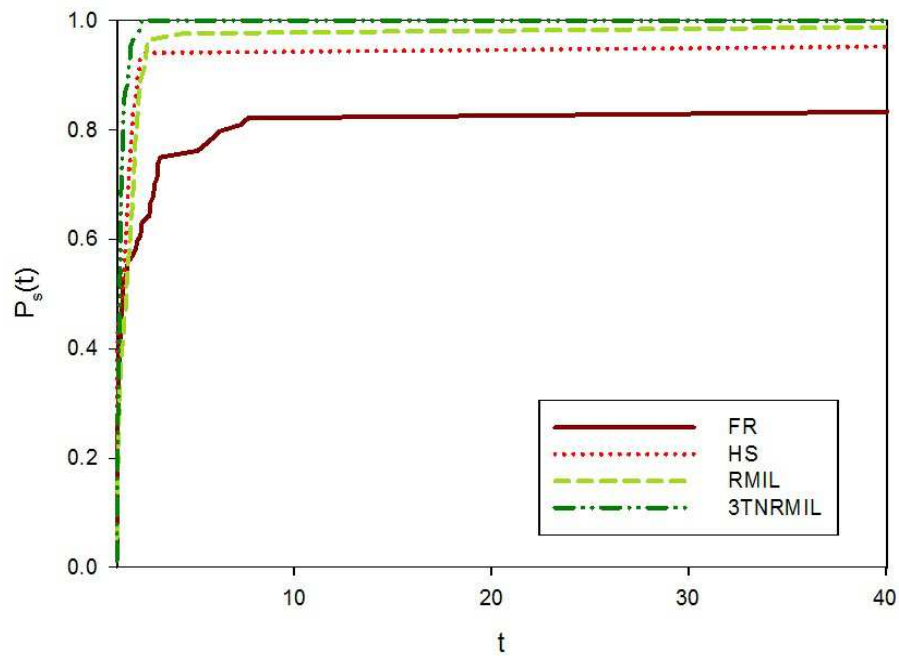


Figure 3: Performance Profile Based on CPU Time

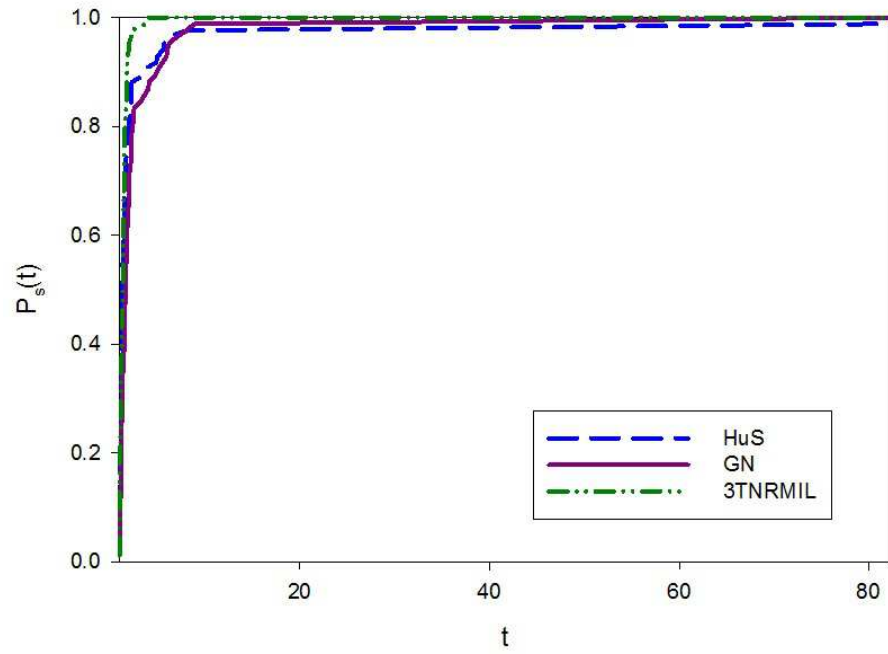


Figure 4: Performance Profile Based on Number of Iterations

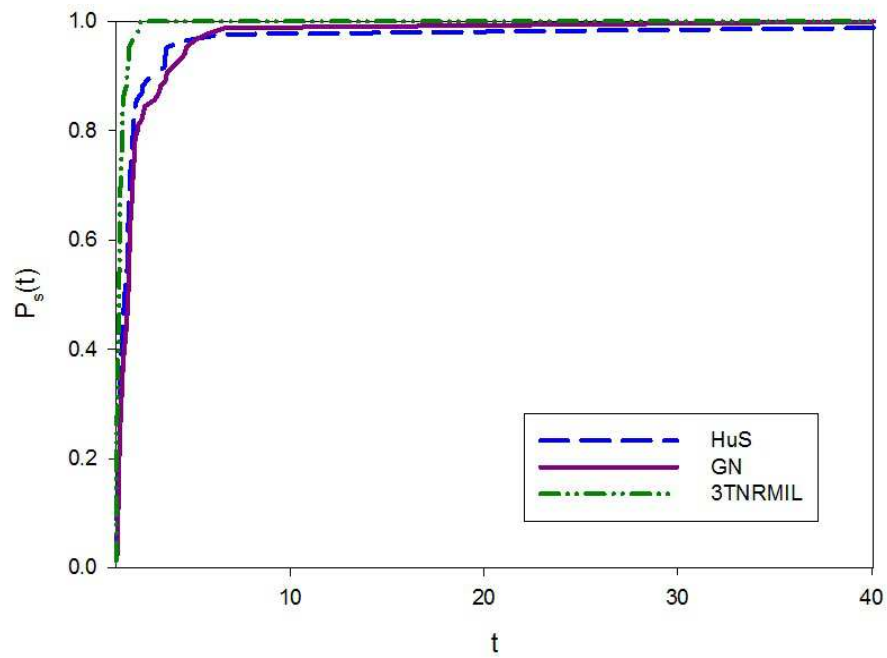


Figure 5: Performance Profile Based on CPU Time

5 Conclusion

The recent studies on CG method have led to the variety of modification on the CG method. In this study, the efficient and simple coefficient RMIL had been implemented in the proposed Three-Term CG method under exact line search which is 3TNRMIL. The numerical results show that 3TNRMIL has the best performance as compared to the classical, modified and hybrid CG methods. In the future, we intend to apply other classical and modified CG methods into the proposed search direction using more test functions and inexact line search.

Acknowledgments

The authors would like to thank The Ministry of Higher Education Malaysia and GE STEM grant with vot no. 07397 for funding this paper. The authors also like to thank the reviewers for improving the presentation of this manuscript. This research is conducted under the fundamental research grant scheme (FRGS) grant number: 600-RMI/FRGS 5/3 (64/2015).

References

- [1] Yuan, G., Lu, S. and Wei, Z. A line search algorithm for unconstrained optimization. *J. Software Engineering & Application*. 2010. 3: 503–509.
- [2] Fletcher, R. and Reeves, C. Function minimization by conjugate gradients. *Comput. J.* 1964. 7: 149–154.
- [3] Hestenes, M. R. and Stiefel, E. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.* 1952. 49: 409–436.
- [4] Polak, E., Ribiere, G. and Polyak. Note sur la convergence de directions conjugees. *Rev. Francaise Inform. Recherche Operationelle*. 1969. 3: 35–43.
- [5] Rivaie, M., Mamat, M., June, L. W. and Mohd, I. A new class of nonlinear conjugate coefficients with global convergence properties. *Applied Mathematics and Computation*. 2012. 218(22): 11323–11332.
- [6] Rivaie, M., Mamat, M., Mohd, I. and Fauzi, M. Modified Hestenes-Steifel conjugate gradient coefficient for unconstrained optimization. *Journal of Interdisciplinary Mathematics*. 2010. 13(3): 241–251.
- [7] Hu, Y. F. and Storey, C. Global convergence result for conjugate gradient methods, *JOTA*. 1991. 2(71): 399–405.
- [8] Gilbert, J. C. and Nocedal, J. Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.* 1992. 2: 21–42.
- [9] Beale, E. M. L. *A derivation of conjugate gradient*, in: F.A. Lootsma (Ed.), *Numerical Methods for Nonlinear Optimization*. Academic Press, London. 1972. 39–43.
- [10] Baluch, B., Salleh, Z., Alhawarat, A. and Roslan, U.A.M. A new modified three-term conjugate gradient method with sufficient descent property and its global convergence. *Journal of Mathematics*. 2017. 1–12.
- [11] Zhang, L., Zhou, W. and Li, D. A descent modified Polak-Ribiere-Polyak conjugate gradient method and its global convergence. *IMA J. Numer. Anal.* 2006. 26: 629–640.

- [12] Li, Q. and Li, D. (2011). A class of derivative-free methods for large-scale nonlinear monotone equations. *IMA J. Numer. Anal.* 2011. 31(4): 1625–1635.
- [13] Dong, X. L., Liu, W. L. and He, Y. B. New Version of the Three-Term conjugate gradient method based on spectral scaling conjugacy condition that generates descent search direction. *Applied Mathematics and Computation.* 2015. 269: 606–617.
- [14] Khadijah, W. Rivaie, M. and Mamat, M. A Three-Term conjugate gradient method under the strong-Wolfe line search. *AIP Conference Proceedings.* 2017. 1–8.
- [15] Narushima, Y., Yabe, H. and Ford, A. A Three-term conjugate gradient method with sufficient descent property for unconstrained optimization. *SIAM Journal on Optimization.* 2008. 21: 212–230.
- [16] Andrei, N. Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization. *Bull. Malays. Math. Sci. Soc.* 2011. 34(2): 319–330.
- [17] Hillstom, K. E. A simulation test approach to the evaluation of nonlinear optimization algorithms. *ACM Trans. Math. Softw.* 1977. 3(4): 305–315.
- [18] Witte, B. F. and Holst, W. R. Two new direct minimum search procedures for functions of several variables. *Proceedings of Spring Joint Computer Conference.* 1964. 27.
- [19] Nagendra, S. *Catalogue of Test Problems for Optimization Algorithm Verification.* Technical Report 97-CRD-110. 1997.
- [20] Silagadze, Z. K. Finding two-dimensional peaks. *Physics of Particles and Nuclei Letters.* 2007. 4(1): 73–80.
- [21] Andrei, N. An unconstrained optimization test functions collection. *Advanced Modelling and Optimization.* 2008. 10: 147–161.
- [22] Dolan, E. D. and Moré, J. J. Benchmarking optimization software with performance profile. *Math. Prog.* 2002. 91(2): 201–213.