ISSN 2278-3091
Volume 8, No.1.5, 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse2181.52019.pdf
https://doi.org/10.30534/ijatcse/2019/2181.52019

On Solving Unconstrained Optimization Using Three Term Conjugate Gradient Method

Nur Idalisa[1], Nor Amila Sofiya Abdullah[2], Nurul Nadia Mohd Jalil[3], Nurul Hafawati Fadhilah[4], Siti Aminah Abdullah[5]
[1]Universiti Teknologi MARA, Malaysia, nuridalisa@uitm.edu.my
[2]Universiti Teknologi MARA, Malaysia, amilasofiya13@gmail.com
[3]Universiti Teknologi MARA, Malaysia, nurulnadia9709@gmail.com
[4]Universiti Teknologi MARA, Malaysia, nurulhafawatifadhilah@gmail.com
[5]Universiti Teknologi MARA, Malaysia, sitiaminah9707@gmail.com

ABSTRACT

Conjugate Gradient (CG) method is well-known for its successes in solving unconstrained optimization problem (UOP). This paper is aimed to further investigate the existing three-term CG methods using Strong Wolfe line search. The three-term method is the extension of established classical CG method of Hestenes and Stiefel (HS) in 1952 and by Rivaie, Mustafa, Ismail and Leong (RMIL) in 2012. The standard UOP comprising Extended Rosenbrock, Extended Himmeblau, Extended White & Holst, Extended Extended Denschnb, and Diagonal 4 functions is considered in this paper. All of UOP with dimension of 2, 100, 500, and 1000. The result is then analyzed through the number of iteration and CPU time. Experimental results provide evidence that the three-term HS is better than three-term RMIL.

Key words: Conjugate gradient method, unconstrained optimization, inexact line search, three-term.

1. INTRODUCTION

The unconstrained optimization problems (UOP) emerge in numerous practical applications. The following is the general UOP:

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

where $f : R^n \to R$ is continuously differentiable. The iterative equation for CG method to generate solution {x_k } is as follows:

$$x_{k+1} = x_k + a_k d_k \qquad k = 0,1,2,\ldots \tag{2}$$

where x_k is the current iterate, $\alpha_k > 0$ is the step-size and d_k is CG search direction computed as below:

$$d_k = \begin{cases} -g_k & \text{if } k = 0 \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 1 \end{cases} \tag{3}$$

where g_k denotes the gradient of $f(x)$ .The β_k is a CG parameter. Reference [1] posited that the CG search direction can also take the form as follows:

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

Equation (4) obtained is by entrenching a parameter $\gamma \in (0,1)$ on the initial direction of (3).

There are variations of CG method comprising classical, hybrid, scaled and the three-term CG methods. The following are classical CG parameters used in this paper based on [2], [3] and [4], respectively:

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

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

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

The variation exists to serve as a motivation to achieve the most ideal CG method by imposing strong global convergence and some other properties [4]. In this conjunction, some scholars are particularly interested in the three-term CG methods. Therefore, this study is conducted to decide the best CG methods based on [5], [6], [7] and then validate the efficiency by using performance profile method of [8].

Below is the list of CG method in this paper:

a) Three-term CG based on [5] known as TTHS:

$$d_k = \begin{cases} -g_k & \text{if } k = 0 \\ -g_k + \beta_k d_{k-1} - \theta_k y_{k-1}, & \text{if } k > 1 \end{cases} \tag{8}$$

where β_k of (5), $\theta_k = \dfrac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}}$ and $y_{k-1} = g_k - g_{k-1}$.

b) Three-term CG based on [6] named Method 1:

$$d_k = \begin{cases} -g_k + \gamma g_k & \text{if} \quad k = 0 \\ -g_k + \beta_k d_{k-1} + \theta_k y_{k-1} & \text{if} \quad k \geq 1 \end{cases} \qquad (9)$$

where $\gamma \in (0,1)$, β_k of (6), and $\theta_k = \dfrac{-g_k^T d_{k-1}}{\|g_{k-1}\|^2}$.

c) Three-term CG based on [7] known as TTRMIL:

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

where β_k of (7), and $\theta_k = \dfrac{-g_k^T d_{k-1}}{\|d_{k-1}\|^2}$.

The type of line search used in this paper is inexact line search which is Strong Wolfe line search given as:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq c_1 \alpha_k g_k^T d_k \qquad (11)$$

$$g(x_k + \alpha_k d_k)^T d_k \leq c_2 g_k^T d_k \qquad (12)$$

Where, $0 < c_1 < c_2 < 1$. The parameters of c_1 and c_2 are varied based on cases displayed in Table 1.

Table 1: Parameters of Strong Wolfe

Case	Value of Strong Wolfe parameters
1	$c_1 = 0.0004, c_2 = 0.001$
2	$c_1 = 0.0004, c_2 = 0.002$
3	$c_1 = 0.0004, c_2 = 0.003$
4	$c_1 = 0.0004, c_2 = 0.004$

2. ALGORITHM

The general CG algorithm given is as follows:

Step 1: Initialization.
Given x_0, set $k=0$.
Step 2: Compute CG coefficient.
Compute β_0 based on (5), (6) or (7).
Step 3: Computing search direction.
Compute d_k based on (8),(9) or (10).
If $g_k = 0$, then stop.
Step 4: Computing step size, α_k using (11) – (12)
Step 5: Updating new point.
$x_{k+1} = x_k + \alpha_k d_k$
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 1 with $k = k+1$.

3. RESULTS

Further summarization of the numerical results which focuses on the average number of iteration (NOI) and CPU time is provided in Table 2.

Table 2: Average CPU time per iteration

Method / Test Function	Method 1	TTRMIL	TTHS
Extended Rosenbrock	0.31373	0.31362	0.38496
Extended Himmelblau	0.08812	0.08323	0.08680
Extended White & Holst	0.44379	0.36009	0.44547
Extended Denschnb	0.15398	0.32890	0.33054
Diagonal 4	0.16471	0.28804	0.29887
TOTAL	1.16433	1.37388	1.54637

Based on Table 2, Method 1 has the best efficiency among others. The efficiency performance of all methods is then extended using the most preferred method of performance profile graph provided by [8].

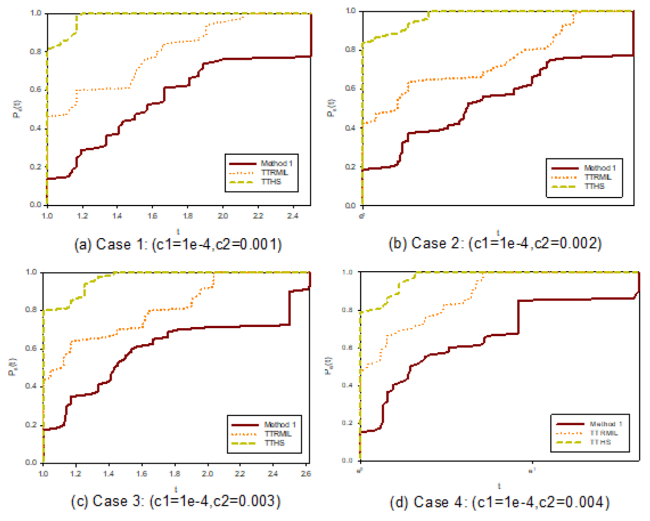

Figure 1: Performance profile based on NOI for all cases

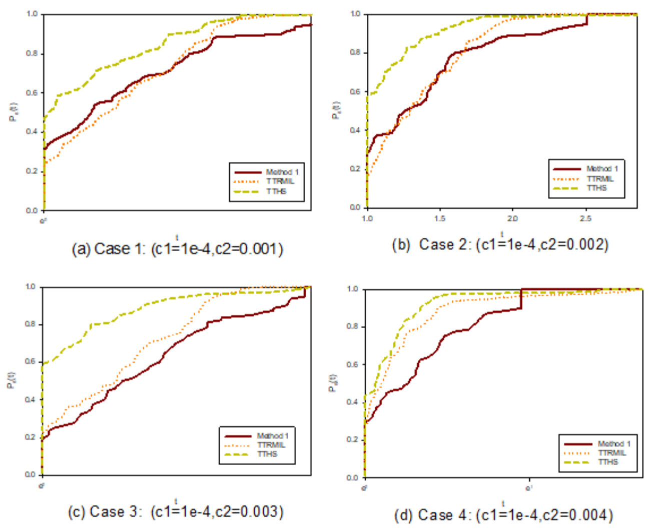

Figure 2: Performance profile based on CPU times for all cases

For all the figures above, it is clearly shown that the top curve is monopolized by TTHS which means that the TTHS converge faster compared to Method 1 and TTRMIL. From the right side of the graph, it shows that all methods are capable to solve all the selected problems. As a conclusion, it can be deduced that the TTHS has better performance in terms of NOI and CPU time compared to other methods. Method 1 is the worst method in terms of NOI, but in terms of CPU time, it lies between the TTHS and TTRMIL methods. Meanwhile, if compared with the three-term RMIL, Method 1 performed better than TTRMIL based on computation time.

FUTURE RESEARCH

The issue of finding global minimum to solve UOP is beneficial and significant in various fields and applications. For instance, the presented study is applicable to machine learning techniques which includes regression, Bayesian learner, decision tree and neural network [9]. The training algorithm for neural network comprising the method of Back propagation, CG, Resilient propagation, quick propagation, and Levenberg-Marquardt [10]. Therefore, the presented study could be extended into the implementation of neural network training algorithm.

ACKNOWLEDGEMENT

The authors would like to offer a sincere thank you to Dr. Mohd Rivaie Mohd Ali who provided guidance and valuable advice in this research.

REFERENCES

1. S. Saha and B. Nath, (2015), **A modified form of conjugate direction for general nonlinear function and its convergence.** 7(1), pp. 25-31.
2. M. Rivaie, M. Mamat, L.W. June and I. Mohd,(2012), **New conjugate gradient coefficient for large scale nonlinear unconstrained optimization**, *International Journal Of Mathematical Analysis*. Vol. 6 (21-24) , pp 1131-1146.
3. M. Rivaie, M. Mamat, L.W. June and I. Mohd, (2012), **A new class of nonlinear conjugate gradient coefficients with global convergence properties.** *Appl. Math. Comp.* Vol 218, pp 11323 – 11332. https://doi.org/10.1016/j.amc.2012.05.030
4. M. Rivaie, M. Mamat, and M. Abashar, (2015), **A new class of nonlinear conjugate gradient coefficients with exact and inexact line searches.** *Appl. Math. Comp.* Vol 268, , pp 1152 – 1163. https://doi.org/10.1016/j.amc.2015.07.019
5. L.Zhang, W.Zhou, and D. H. Li,(2007), **Some descent three-term conjugate gradient methods and their global convergence**, Optim. Methods. Softw., Vol. 22, pp. 697-711. https://doi.org/10.1080/10556780701223293
6. N. I. Norddin, M. Rivaie, N. H. Fadhilah, and M. A. S. Nasir, (2018), **A New Sufficient Descent Conjugate Gradient Method with Exact Line Search**. in *International Conf. on Mathematical Sciences and Technology (MathTech)* , Penang.
7. L. M. Zou, Y. M. Feng and J. K. Liu, (2018), **Some three-term conjugate gradient methods with the inexact line search condition**, *Calcolo*, Vol. 55, no. 16. https://doi.org/10.1007/s10092-018-0258-3
8. E.D. Dolan and J.J. More, (2002), **Benchmarking optimization software with performance profile**, *Math. Prog*. Vol. 91, pp. 201-213. https://doi.org/10.1007/s101070100263
9. J. Goyal and B. Kishan, (2019), **Progress on machine learning techniques for software fault prediction,** *International Journal of Advanced Trends in Computer Science and Engineering.* Vol 8, No. 2, March – April. https://doi.org/10.30534/ijatcse/2019/33822019
10. L. C. P. Velasco, R.P. Serquina, M. S. A. A. Zamad, B. F. Juanico, and J. C. Lomocso, (2019), **Performance analysis of multilayer perceptron neural network models in week-ahead rainfall forecasting,** *International Journal of Advanced Computer Science and Applications.* Vol. 10, No. 3. https://doi.org/10.14569/IJACSA.2019.0100374