

## The implementation of conjugate gradient methods for data fitting

N. Zullpakkal<sup>1,\*</sup>, N. Shapiee<sup>2</sup>, S.F. Husin<sup>3</sup>, W. Khadijah<sup>4</sup>, N.A. Salahudin<sup>1</sup>, S.M. Zokri<sup>1</sup> and M.N. Rashidi<sup>1</sup>

<sup>1</sup>Mathematical Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Terengganu Branch Kuala Terengganu Campus, Kuala Terengganu, Terengganu, Malaysia

<sup>2</sup>Tamhidi Centre, Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia

<sup>3</sup>Centre of Foundation Studies, Universiti Teknologi MARA, Cawangan Selangor, Kampus Dengkil, Selangor, Malaysia

<sup>4</sup>School of Mathematical Sciences, College of Computing, Informatics and Media, Universiti Teknologi MARA Shah Alam Campus, Shah Alam, Selangor, Malaysia

**ABSTRACT** – The conjugate gradient (CG) method is widely used to solve the unconstrained optimization problem by finding the optimal solution. This problem can be solved by an iterative method. CG method can be classified into classical, modified, spectral, three terms, and hybrid. In this research, Polak-Ribiere-Polyak (PRP), Rivaie-Mustafa-Ismail-Leong (RMIL), Nurul Hajar-Mustafa-Rivaie (NMR) and Linda-Aini-Mustafa-Rivaie (LAMR) are the four chosen methods for this comparison study. These methods are tested under the Armijo line search. There are 14 test functions with five initial points and various variables are chosen. This comparison study is tested using MatlabR2022a to evaluate iteration number and CPU time. The performance profiles of the numerical result are plotted using a Sigma plot. Then, a set of data, the ASB dividend rate is used to form a linear model. In conclusion, PRP performs better than any other method since it yields the best numerical results and is applicable for data fitting.

### ARTICLE HISTORY

Received: 16/02/2023

Revised: 26/03/2023

Accepted: 31/03/2023

### KEYWORDS

*Conjugate gradient*

*Iteration*

*CPU time*

*Optimization*

## INTRODUCTION

Optimization is a decision-making tool that involves finding the maximum or minimum value of a function

$$\min_{x \in \mathbb{R}^n} f(x)$$

There are two types of optimization problems which are constrained and unconstrained. This problem can be solved by an iterative method,

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, 3 \quad (1)$$

The step size,  $\alpha_k$  can be solved using exact or inexact line searches. There are three famous inexact line searches which are Wolfe [1], Goldstein [2] and Armijo [3]. Armijo line search is said to be the simplest method where it is easy to be implemented in computation procedures. Armijo line search is much more popular in image restoration and compressed sensing challenges than the Wolfe line search. The formulation of Armijo line search is written as,

$$\beta \in (0, 1), s > 0, \mu, \mu \in \left(0, \frac{1}{2}\right) \text{ where } s = \alpha_k \text{ and } f(x_k + \alpha_k d_k) \leq f(x_k) + \mu \alpha_k d_k \quad (2)$$

The search direction in this iterative method can be solved by optimization methods such as conjugate gradient (CG), steepest descent (SD), Newton's, and Quasi-Newton (QN). CG method is proven to be the best method due to its cheap memory requirements and high global convergence properties, which allow the method to deal with large-scale nonlinear equations quickly and effectively [4]. Different method leads to different search direction. The principal search direction of the CG technique is designated as,

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

The  $\beta_k$  in (3) refers to CG coefficient. The examples of CG coefficients by Polak-Ribiere-Polyak (PRP) [5], Rivaie-Mustafa-Ismail-Leong (RMIL) [6], Nurul Hajar-Mustafa-Rivaie (NMR) [7] and Linda-Aini-Mustafa-Rivaie (LAMR) [8] are listed respectively below

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}, \tag{4}$$

$$\beta_k^{RMIL} = \frac{g_k^T(g_k - g_{k-1})}{\|d_{k-1}\|^2}, \tag{5}$$

$$\beta_k^{NMR} = \frac{\left(\frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2} + \frac{g_k^T(g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}\right)}{2} \tag{6}$$

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

The improvement of CG method can be done by modifying the search direction or CG coefficient. Nowadays, most of researchers introduce new hybrid CG coefficients by combining two CG coefficients such as PRP-FR [9] and HS-LAMR [10]. Besides, the research in this field is still relevant since the CG coefficients are applicable for solving real life problems such as regression analysis [11, 12], image restoration [13], portfolio selection [14] and whale optimization algorithm [15].

### ALGORITHM

The algorithm of the CG method is given as follows:

- Step 1: Initialization. Given  $x_0$ , set  $k = 0$ .
- Step 2: Compute the search direction,  $d_k$  as in (3).
- Step 3: Compute all coefficients as in (4), (5), (6) and (7).
- Step 4: Compute the step size,  $\alpha_k$  by using the Armijo line search as in (2).
- Step 5: Update a new point based on an iterative formula in (1).
- Step 6: Convergence test and stopping criterion.

### NUMERICAL RESULTS

All of the CG coefficients are tested for their robustness and efficiency. These two criteria are important in order to determine the best method based on the rate of solving test functions, least CPU time, and least NOI. Table 1 displays the fourteen test functions and five initial points used to compare each CG coefficient under the Armijo line search. Five initial points of the function are chosen randomly, and each function is evaluated using five different dimensions which are 2,4,10,100, and 1000.

**Table 1.** List of test functions

No	Test Function	Variables
1	Generalized Quartic Function	2,4,10,100,1000
2	Quadratic QF2 function	2,4,10,100,1000
3	Generalized Tridiagonal 1 function	2,4,10,100,1000
4	Quadratic QF1	4,10,100,1000
5	Extended Quadratic Penalty QP2	2,4,10,100,1000
6	Hager function	2,4,10,100,1000
7	Extended Powell Function	4,10,100,1000
8	ARWHEAD	2,4,10
9	Diagonal4 Function	2,4,10,100,1000
10	FLETCHCR Function (CUTE)	2,4,10,100,1000
11	Diagonal 2 function	2,4,10
12	NONSCOMP function	2,4,10,100,1000
13	Extended DENSCHNB function	2,4,10,100,1000
14	Extended quadratic penalty QP1	2,4,10

The numerical results are measured in terms of number of iteration (NOI) and central processing unit (CPU) time and interpreted using performance profile introduced by Dolan and More [16]. The performance profile of NOI and CPU time are illustrated as in Figures 1 and 2 respectively.

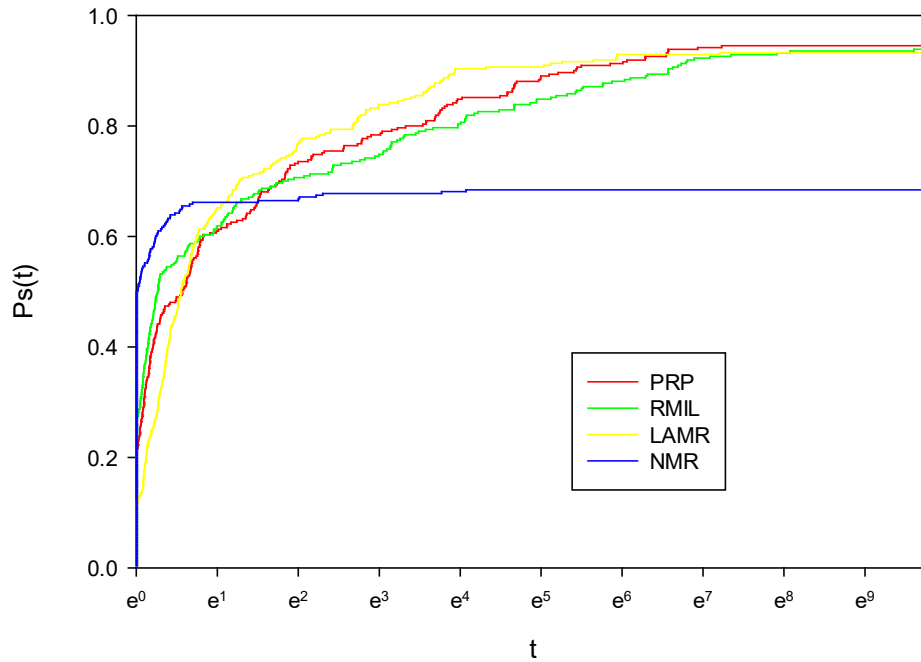


Figure 1. Performance profile based on number of iterations

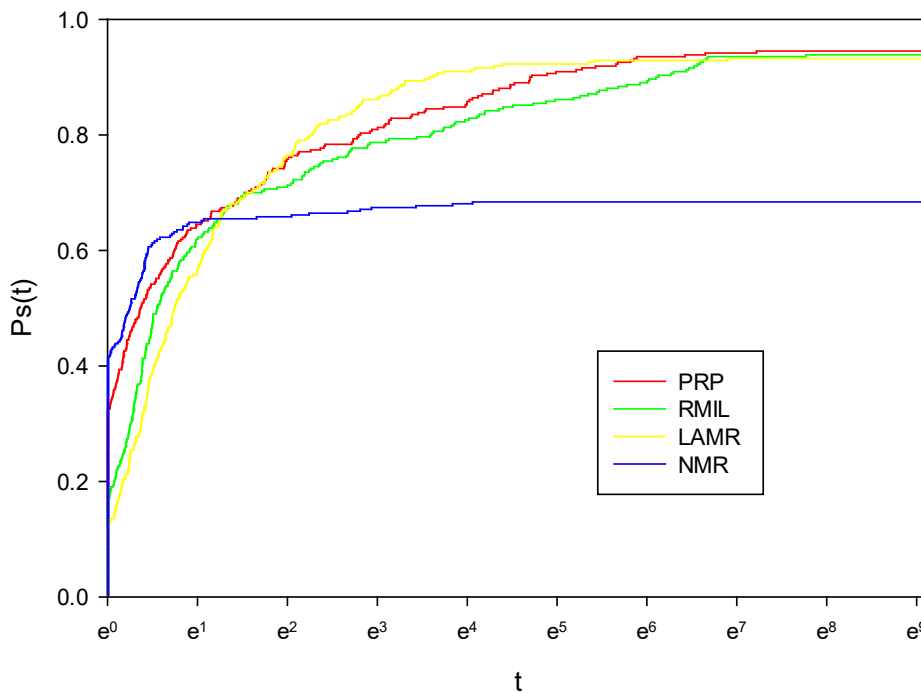


Figure 2. Performance profile based on CPU time

Figure 1 and Figure 2 show that NMR is the fastest method with the least NOI and CPU time, which makes it the most efficient one. The LAMR method is the slowest but at a certain point, it overtakes the other CG method. It means that this method is able to generate less NOI and CPU time for certain functions compared to other methods. Based on both figures, PRP has the highest percentage of solving the test functions which is 94.5% while RMIL, LAMR, and NMR can solve 94.2%, 93.2%, and 68.4% of test functions respectively. Based on the success rate of solving test functions, it can be concluded that PRP is the most robust method.

## IMPLEMENTATION & DISCUSSION

This section discusses on the implementation of CG method for data fitting. The CG method is compared with least square and excel trendline methods by generating a regression model. The real data, ASB dividend rate between 2003 and 2020 are used and listed in Table 2.

**Table 2.** ASB dividend rate from 2003 to 2020

Year	ASB Dividend rate from 2003 to 2020
2003	7.25
2004	7.25
2005	7.25
2006	7.3
2007	8.0
2008	7.0
2009	7.3
2010	7.5
2011	7.65
2012	7.75
2013	7.7
2014	7.5
2015	7.25
2016	6.75
2017	7.0
2018	6.5
2019	5.0
2020	3.75

A linear regression model is formed based on the following equation,

$$y = \alpha_0 + \alpha_1 x \quad (8)$$

As shown in the formula,  $x$  is the explanatory variable and  $y$  is the dependent variable. The slope of the line is  $a_0$  and  $a_1$  is the intercept. The result is analyzed by calculating the relative error using the following formula,

$$\text{relative error} = \frac{|\text{exact value} - \text{approximate value}|}{|\text{exact value}|} \quad (9)$$

### Conjugate gradient method

The real data in Table 2 is coded to form a linear regression model which can be used in MatlabR2022a. Then, all the CG coefficients are implemented for data fitting using the same initial points. Table 3 displays the outcomes for NOI, CPU time,  $a_0$  and  $a_1$  which are generated in MatlabR2022a.

**Table 3.** Results of CG method

Method	Initial Point	NOI	CPU Time	$a_0$	$a_1$
RMIL	(2,2)	58	0.0028	8.1020	-0.1178
PRP	(2,2)	62	0.0031	8.1020	-0.1178
LAMR	(2,2)	80	0.0518	Fail	Fail
NMR	(2,2)	9	0.0021	Fail	Fail

### Least squares method

The least squares method is a way to figure out which straight line fits a set of points the best. To get the best straight line, the sum of the data's errors should be as small as possible. For the best fit, the following formula is used as a criterion.

$$\min E(a) = \sum_{i=1}^m (y_i - (a_0 + a_1 x_i))^2$$

The values of  $a_0$  and  $a_1$  are obtained manually using the following formula.

$$\begin{bmatrix} m & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m x_i y_i \end{bmatrix}$$

By using the ASB dividend rate data, the value of the element in the matrix two by two above is obtained. The total number of data is assigned as  $m$  and  $x$  is the total sum of  $m$ . The dividend rate for each year is  $y$ . Table 4 shows the element of the matrix based on the data.

**Table 4.** The element values of the matrix based on the provided data

Year	$x$	ASB Dividend rate, $y$	$x^2$	$xy$
2003	1	7.25	1	7.25
2004	2	7.25	4	14.5
2005	3	7.25	9	21.75
2006	4	7.3	16	29.2
2007	5	8.0	25	40
2008	6	7.0	36	42
2009	7	7.3	49	51.1
2010	8	7.5	64	60
2011	9	7.65	81	68.85
2012	10	7.75	100	77.5
2013	11	7.7	121	84.7
2014	12	7.5	144	90
2015	13	7.25	169	94.25
2016	14	6.75	196	94.5
2017	15	7.0	225	105
2018	16	6.5	256	104
2019	17	5.0	289	85
2020	18	3.75	324	67.5
Total	171	125.7	2109	1137.1

The summation values from Table 4 are substituted into the matrix formula and solved to obtain  $a_0$  and  $a_1$ .

$$\begin{bmatrix} 18 & 171 \\ 171 & 2109 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 125.7 \\ 1137.1 \end{bmatrix}$$

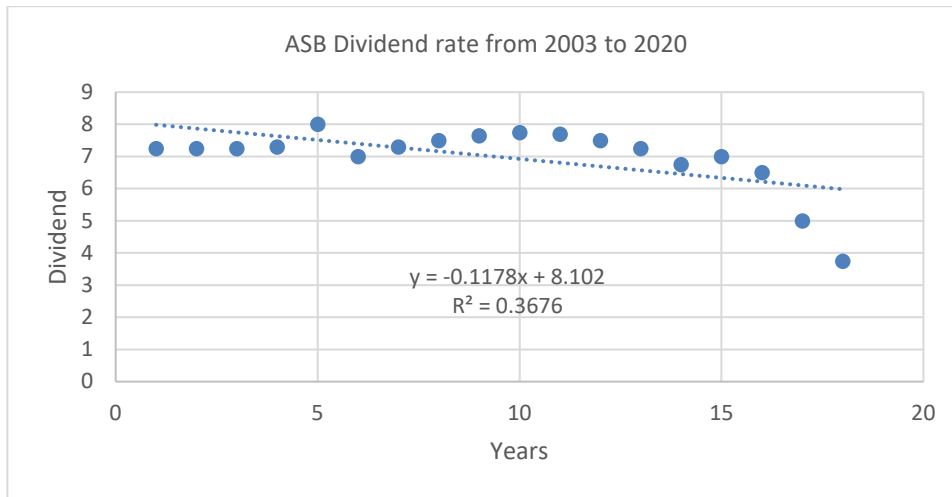
$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 18 & 171 \\ 171 & 2109 \end{bmatrix}^{-1} \begin{bmatrix} 125.7 \\ 1137.1 \end{bmatrix}$$

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \frac{1}{8721} \begin{bmatrix} 2109 & -171 \\ -171 & 18 \end{bmatrix} \begin{bmatrix} 125.7 \\ 1137.1 \end{bmatrix}$$

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 8.10196 \\ -0.11775 \end{bmatrix}$$

### Excel trendline method

Microsoft Excel is used to generate the trendline as well as the values of  $a_0$  and  $a_1$ . The trendline is based on the relationship between dividends paid by ASB and years which is shown in Figure 3.



**Figure 3.** Trendline of ASB dividend rate from 2003 to 2020

In conclusion, RMIL and PRP yield the same values of  $a_0$  and  $a_1$  as least square and excel trendline method. Thus, the values of  $a_0$  and  $a_1$  may be represented in the form linear equation as follows,

$$y = 8.102 - 0.1178x$$

Thus, the relative errors of each year are computed using (9) and listed in Table 5.

**Table 5.** Relative error computed by year

Year	Actual Value	Approximate Value	Relative Error
2003	7.25	7.9842	0.1013
2004	7.25	7.8664	0.0850
2005	7.25	7.7486	0.0688
2006	7.3	7.6308	0.0453
2007	8.0	7.513	0.0609
2008	7.0	7.3952	0.0565
2009	7.3	7.2774	0.0031
2010	7.5	7.1596	0.0454
2011	7.65	7.0418	0.0795
2012	7.75	6.9240	0.1066
2013	7.7	6.8062	0.1161
2014	7.5	6.6884	0.1082
2015	7.25	6.5706	0.0937
2016	6.75	6.4528	0.0440
2017	7.0	6.335	0.0950
2018	6.5	6.2172	0.0435
2019	5.0	6.0994	0.2199
2020	3.75	5.9816	0.5951
Average			0.1093

Table 6 shows the summarization of the CG coefficients based on the numerical results where the least value of NOI and CPU time of each method are calculated by percentage. The applicability column shows the implementation of each method for data fitting.

**Table 6.** Numerical results

Methods	NOI (%)	CPU (%)	Test Function (%)	Applicability
PRP	21.3	32.6	94.5	Applicable
RMIL	22.9	16.8	94.2	Applicable
LAMR	11.9	11.9	93.2	Not Applicable
NMR	50.0	41.3	68.4	Not Applicable

Based on Table 6, PRP is considered to be the best method since it has the highest percentage for solving 293 test functions. PRP can also be implemented in regression analysis.

## CONCLUSION

The results show that PRP has a better performance method compared to the other methods. It is able to solve most of the test functions. The graphs in the performance profile are based on NOI and CPU generated and show that PRP is the most robust method. Besides, the application result of CG method is also stated in this chapter. The value for  $a_0$  and  $a_1$  of certain CG is the same as the least square and excel trendline method. Therefore, it has been proved that the CG method can also be applied using real data.

## ACKNOWLEDGEMENT

The authors would like to express their utmost gratitude to the editors and the referees for the recommendations and valuable comments.

## REFERENCES

- [1] P. Wolfe, "Convergence conditions for ascent methods," *SIAM Review*. vol. 11, no. 2, pp. 226-235, 1969.
- [2] A.A. Goldstein, "On Steepest Descent," *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*. vol. 3, no. 1, pp. 147-151, 1965.
- [3] L. Armijo, "Minimization of functions having Lipschitz continuous first partial derivatives," *Pacific Journal of Mathematics*. vol. 16, no. 1, pp. 1-3, 1966.
- [4] M. Koorapetse and P. Kaelo, "An efficient hybrid conjugate gradient-based projection method for convex constrained nonlinear monotone equations," *Journal of Interdisciplinary Mathematics*. vol. 22, no. 6, pp. 1031-1050, 2019.
- [5] E. Polak and G. Ribiere, "Note sur la convergence de méthodes de directions conjuguées," *Revue française d'informatique et de recherche opérationnelle. Série rouge*. vol. 3, no. 16, pp. 35-43, 1969.
- [6] M. Rivaie, M. Mamat, L. W. June, and I. Mohd, "A New class of nonlinear conjugate gradient coefficients with global convergence properties," *Applied Mathematics and Computation*. vol. 218, no. 22, pp. 11323-11332, 2012.
- [7] N. H. M. Yusoff, M. Mamat, M. Rivaie, and I. Mohd, "A new conjugate gradient method for unconstrained optimization with sufficient descent," in *AIP Conference Proceedings*, 2014, pp. 514-519.
- [8] N. Zull, N. Aini, M. Rivaie, and M. Mamat, "A new gradient method for solving linear regression model," *International Journal of Recent Technology and Engineering*. vol. 7, no. 5, pp. 624-630, 2019.
- [9] D. Touati-Ahmed and C. Storey, "Efficient hybrid conjugate gradient techniques," *Journal of optimization theory and applications*. vol. 64, pp. 379-397, 1990.
- [10] N. Zullpakkal, N. 'Aini, N. Ghani, N. Mohamed, N. Idalisa, and M. Rivaie, "Covid-19 data modelling using hybrid conjugate gradient method," *Journal of Information and Optimization Sciences*. vol. 43, no. 4, pp. 837-853, 2022.
- [11] N.I. Norddin, M.R.M. Ali, N.H. Fadhilah, N. Atikah, A. Shahida, and N.H.N. Noh, "multiple linear regression model of rice production using conjugate gradient methods," *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*. pp. 229-236, 2019.
- [12] N. Idalisa, M. Rivaie, N.H.M. Noh, M.A.S. Nasir, N.H. Fadhilah, and N. Alias, "A new three-term conjugate gradient method with application to regression analysis," *International Journal of Electrical & Computer Engineering*. vol. 12, no. 5, pp. 2022.
- [13] S. Babaie-Kafaki, N. Mirhoseini, and Z. Aminifard, "A descent extension of a modified polak-ribière-polyak method with application in image restoration problem," *Optimization Letters*. vol. 17, no. 2, pp. 351-367, 2023.
- [14] M. Malik, I.M. Sulaiman, A.B. Abubakar, and G. Ardaneswari, "A new family of hybrid three-term conjugate gradient method for unconstrained optimization with application to image restoration and portfolio selection," *AIMS Mathematics*. vol. 8, no. 1, pp. 1-28, 2023.
- [15] L.R. Khaleel and B.A. Mitras, "Hybrid whale optimization algorithm with modified conjugate gradient method to solve global optimization problems," *Open Access Library Journal*. vol. 7, no. 6, pp. 1-18, 2020.
- [16] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical Programming*. vol. 91, pp. 201-213, 2002.