International Journal of Engineering & Technology, 7 (3.28) (2018) 316-320

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Comparative Study of the Spectral Conjugate Gradient Methods in Regression Analysis

Norhaslinda Zull[1], Wan Khadijah[1], Ummi Khalthum Mohd Yusof[1], Syazni Shoid[1], Mohd Rivaie[2], Mustafa Mamat [1]*

[1]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Besut Campus, Terengganu, Malaysia
[2]Department of Computer Sciences and Mathematics, Universiti Teknologi MARA (UiTM) Terengganu, Campus Kuala Terengganu, Malaysia
**Corresponding author Email: must@unisza.edu.my*

Abstract

In this paper, a spectral conjugate gradient (CG) method is introduced for solving the unconstrained optimization problems. This method are compared with others spectral CG coefficients. Several different dimensions for eighteen types of optimization problems are used to test the efficiency and robustness of spectral conjugate gradient methods by using Matlab subroutine programming. The method should possess the convergence analysis under strong Wolfe line search. The numerical results based on iteration number and CPU time are interpreted into performance profile by using *SigmaPlot*. This method will be implemented in regression analysis to validate its capability on estimating the data.

Keywords: *spectral; convergence analysis; strong Wolfe; regression.*

1. Introduction

The standard unconstrained optimization problem is stated by the following rule,

$$\min_{x\in R^n} f(x) \tag{1.1}$$

It is usually solved iteratively from an initial point, $x_o \in R^n$. According to the recurrence formula, the CG method follows as below,

$$x_{k+1} = x_k + \alpha_k d_k \quad k = 1,2,... \tag{1.2}$$

where x_k is the current iteration point, $\alpha_k > 0$ is a stepsize. Any line search is needed in order to find the stepsize. It is either exact or inexact line search. The difference between exact and inexact line searches is on how to obtain the stepsize. The exact line search will provide the exact stepsize, while the inexact line search is known to produce an approximate stepsize. As stated by [2], the exact line search is always costly and impractical. The inexact line searches that are mostly used by researchers are Goldstein [3], Armijo [12] and Wolfe [23]. Wolfe is divided into strong and weak Wolfe.
The strong Wolfe line search is defined by these following conditions,

$$f(x_k + \alpha_k d_k) \le f(x_k) + \mu\alpha_k g_k^T d_k \tag{1.3}$$
$$\left|g_{k+1}^T d_k\right| \le -\sigma g_g^T d_k.$$

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

The basic search direction, d_k is described as,

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

which is $g_k = \nabla f(x_k)$, while parameters of $\beta_k \in R$ are the coefficient which determine the difference of CG methods.
The conjugate gradient method can be classified into several types such as classical CG, spectral CG and hybrid CG. The classical CG method is established by Hestenes and Stiefel (HS). After few years, another CG method developed by Fletcher Reeves (FR). A new CG method has been introduced by Polak-Ribiere (PR). Then, Fletcher again proposed another method, Conjugate Descent (CD). All of these classical CG methods can be read through [6, 15, 24-25].
Nowadays, the CG methods are developed rapidly by many researchers such as [1, 20]. The β_k that has been modified for this paper is cited from [14]. All the formula of these researchers are stated below,

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} \tag{1.4}$$

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

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

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}, \tag{1.5}$$

$$\beta_k^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{\|d_{k-1}\|^2}, \tag{1.6}$$

where $g_{k-1} = \nabla f(x_{k-1})$ and $\|.\|$ indicates the Euclidian norm of vectors.

The spectral CG method is introduced by Barzilai and Borwein [9] in 1988. This method is developed by Raydan [13] for solving large scale optimization problems. Then, it is modified by many researchers as [7, 10, 19, 26]. Meanwhile, the researchers that have been working on the hybrid CG methods are [4, 16, 21, 27].

The structure of this paper is organized as follows. The algorithm used in this paper is stated in Section 2. In section 3, the global convergence analysis of CG method is presented. The numerical results of these methods are compared and elaborated in the section 4. The implementation of the tested methods is discussed in section 5. Lastly, conclusion is presented in section 6.

2. The Spectral Conjugate Gradient Method

In this paper, the RMIL, CD and HS methods are computed by using the spectral gradient method and are denoted as SRMIL, SCD and SHS respectively. The search direction of spectral gradient method, d_k are taken based on [7] and defined by the following rule:

$$d_k = \begin{cases} -g_k & \text{if} \quad k = 0 \\ -\theta_k g_k + \beta_k d_{k-1} & \text{if} \quad k \geq 1 \end{cases} \tag{2.1}$$

where θ_k is cited from [10],

$$\theta_k = 1 - \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}. \tag{2.2}$$

These CG methods are tested by using strong Wolfe line search based on the algorithm below. The following algorithm is referred from the basic general algorithm and used for computing the spectral CG:

Step 1: Initialization. Given an initial point x_0 and set $k = 0$.

Step 2: Computing (1.6), (1.5), (1.4).

Step 3: Computing search direction based on (2.1) and (2.2). If $g_k = 0$, terminate the execution of algorithm.

Step 4: Computing step size α_k by (1.3).

Step 5: Updating a new point as shown in (1.2).

Step 6: Convergent test and stopping criteria. If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| < \varepsilon$, then terminate. Otherwise, go to Step 1 with $k = k + 1$.

3. Global Convergence Properties

The convergent properties of spectral β_k^{RMIL} using inexact line search will be furthered studied. This paper is focusing on the global convergence properties as it is significant to ensure any algorithm to converge.

The β_k^{RMIL} is simplified to make it easier for proving. By simplifying the coefficient of β_k^{RMIL}, it shows that

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

$$0 \leq \beta_k^{RMIL} \leq \frac{\|g_k\|^2}{\|d_{k-1}\|^2} \tag{3.1}$$

The following assumptions are used to analyze the global convergence properties of CG methods.

Assumption 1

i. f is bounded below on the level set R^n and its continuous and differentiable in a neighbourhood N of the level set $\ell = \{x \in R^n \mid f(x) \leq f(x_0)\}$ at the starting point x_0.

ii. The gradient is Lipschitz continuous, where existence of constant $L > 0$ such that $\|g(x) - g(y)\| \leq L\|x - y\|$, for $\forall x, y \in N$.

Based on the assumption above, in [8] has proven the following Lemma. The following Theorem 2 shows that Algorithm 1 possesses the global convergence properties by any line search.

Lemma 1: Assume that Assumption 1 is satisfied. Consider any iteration method (1.2), where d_k satisfies descent search direction, where α_k is obtained by Wolfe line search.

Then

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty \tag{3.2}$$

Theorem 2: Assume that Assumption 1 holds. Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm 1 and let the stepsize α_k determined by Strong Wolfe line search. Then,

$$\lim_{k \to \infty} \|g_k\| = 0.$$

Proof: This proof has been referred from [26]. According to the given conditions, Lemma 1 all hold. In the following, the condition in Theorem 2 will be developed by contradiction. By contradiction, there is a positive constant $c > 0$ such that

$$\|g_k\| \geq c, \tag{3.3}$$

Holds for $\forall k \geq 1$, then, rewrite (2.1) as below

$$d_k + \theta_k g_k = \beta_k^{RMIL} d_{k-1},$$

After squaring both sides of it, we obtain

$$\|d_k\|^2 = (\beta_k^{RMIL})^2 \|d_{k-1}\|^2 - 2\theta_k g_k^T d_k - \theta_k^2 \|g_k\|^2 \tag{3.4}$$

Dividing both sides of (3.4) by $\left(g_k^T d_k\right)^2$, we get

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} = \frac{(\beta_k^{RMIL})^2 \|d_{k-1}\|^2}{\left(g_k^T d_k\right)^2} - \frac{2\theta_k g_k^T d_k}{\left(g_k^T d_k\right)^2} - \frac{\theta_k^2 \|g_k\|^2}{\left(g_k^T d_k\right)^2}$$

Note that, $g_k^T d_k = -\|g_k\|^2$. By applying this condition, it shows that,

$$= \frac{(\beta_k^{RMIL})^2 \|d_{k-1}\|^2}{\|g_k^T d_k\|^2} - \frac{2\theta_k g_k^T d_k}{\|g_k^T d_k\|^2} - \frac{\theta_k^2 \|g_k\|^2}{\|g_k^T d_k\|^2}$$

$$= \frac{(\beta_k^{RMIL})^2 \|d_{k-1}\|^2}{\|g_k^T d_k\|^2} - \frac{1}{\|g_k\|^2}\left(\theta_k^2 - 2\theta_k\right).$$

After completing square, yields,

$$= \frac{(\beta_k^{RMIL})^2 \|d_{k-1}\|^2}{\|g_k^T d_k\|^2} - \frac{1}{\|g_k\|^2}\left(\theta_k^2 - 2\theta_k + 1 - 1\right)$$

$$= \frac{(\beta_k^{RMIL})^2 \|d_{k-1}\|^2}{\left(g_k^T d_k\right)^2} - \frac{(\theta_k - 1)^2}{\|g_k\|^2} + \frac{1}{\|g_k\|^2} \le \frac{(\beta_k^{RMIL})^2 \|d_{k-1}\|^2}{\left(g_k^T d_k\right)^2} + \frac{1}{\|g_k\|^2}$$

Applying (3.1), yields

$$= \left(\frac{\|g_k\|^2}{\|d_{k-1}\|^2}\right)^2 \frac{\|d_{k-1}\|^2}{\left(g_k^T d_k\right)^2} + \frac{1}{\|g_k\|^2} = \frac{\|g_k\|^4}{\|d_{k-1}\|^4} \frac{\|d_{k-1}\|^2}{\left(g_k^T d_k\right)^2} + \frac{1}{\|g_k\|^2}$$

$$= \frac{\|g_k\|^2}{\|d_{k-1}\|^2 \|d_k\|^2} + \frac{1}{\|g_k\|^2} \le \frac{1}{\|d_{k-1}\|^2} + \frac{1}{\|g_k\|^2} \qquad (3.5)$$

By noting that $\frac{1}{\|d_0\|^2} = \frac{1}{\|g_0\|^2}$, then from (3.5), we obtain

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} = \frac{1}{\|g_0\|^2} + \frac{1}{\|g_k\|^2}.$$

Hence, $\frac{\|d_k\|^2}{(g_k^T d_k)^2} \le \sum_{i=0}^{k-1} \frac{1}{\|g_i\|^2}$

Taking summation on the both sides,

$$\sum_{i=0}^{k-1} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} \ge \sum_{i=0}^{k-1} \frac{c^2}{k}. \qquad (3.6)$$

Thus, from (3.6) and (3.3), it follows

$$\sum_{k=0}^{k-1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = +\infty$$

This contradicts the Zoutendjik condition in Lemma 1. Thus, it proves the condition in Theorem 2.

4. Results and Discussion

SRMIL is compared to the spectral CD (SCD) and spectral HS (SHS). Then, the performances of these methods are evaluated based on the comparison of the iteration numbers and Central Processing Unit (CPU) time to reach the solution point. All these problems are established with few dimensions of small and large scales. Four different initial points have been tested, starting from a point which is closer to the solution point till to the point further away from the solution point. The choices of the initial points are randomize. For further reading, refer to [11].

We considered $\varepsilon = 10^{-6}$ and all these methods are terminated when the stopping criteria $\|g_k\| < 10^{-6}$ has been fulfilled. All the problems stated below are computed by MatlabR2012 subroutine programming. The number of iteration and CPU time are noted in order to compare the performance of these spectral CG methods. All these methods are tested by using standard test problems based on [17]. Table 1 shows the standard test problems used in this article.

Table 1: List of standard test problems

No.	Test Problems	Dimensions	Initial Points
1	BIGGSB1	2	(10,10),(20,20), (30,30),(40,40)
2	Booth	2	(3,3),(6,6),(9,9), (12,12)
3	FLETCHR	2	(2,2),(6,6),(8,8), (10,10)
4	Generalized Quartic	2	(3,3),(6,6),(12,12),(22,22)
5	NONSCOMP	2	(2,2),(6,6),(9,9),(11,11)
6	Six Hump	2	(5,5),(9,9),(14,14),(22,22)
7	Three Hump	2	(-8,-8),(2,2),(8,8),(12,12)
8	Trecanni	2	(2,2),(4,4),(8,8),(11,11)
9	Zettl	2	(13,13),(18,18), (24,24),(28,28)
10	Extended Penalty	2,10	(3,…,3),(6,…,6), (10,…,10),(16,…,16)
11	Extended Tridiagonal	2,10	(4,…,4),(14,…,14), (24,…,24),(34,…,34)
12	Extended White and Holst	2,10,100	(-9,…,-9),(-2,…,-2), (6,…,6),(11,…,11)
13	Hager	2,10,100	(2,…,2),(5,…,5), (9,…,9),(12,…,12)
14	Raydan	2,10,100	(2,…,2),(3,…,3), (4,…,4), (5,…,5)
15	Extended Ros-enbrock	2, 10,100, 1000	(2,…,2),(7,…,7), (15,…,15),(22,…,22)
16	Shalow	2,10,100, 1000	(11,…,11),(16,…,16), (60,…,60),(64,…,64)
17	Diagonal 4	2, 10,100, 1000,10000	(10,…,10),(15,…,15), (20,…,20),(25,…,25)
18	Extended Himmelblau	2,10,100, 1000,10000	(4,…,4),(9,…,9), (14,…,14),(18,…,18)

All the results stated in the table above are presented by Performance Profile which has been introduced by [5]. This performance profile will define the robustness and efficiency of all the algorithms at one time. Comparisons can be made through the plotted graphs which are presented by the performance profile in term of iteration number and CPU time.

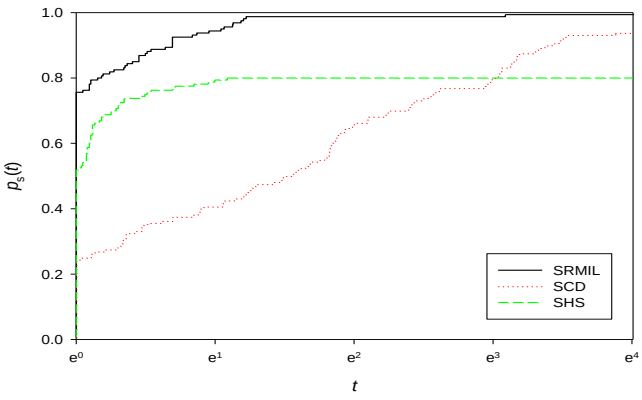

Fig. 1: The performance profile based on Iteration Number

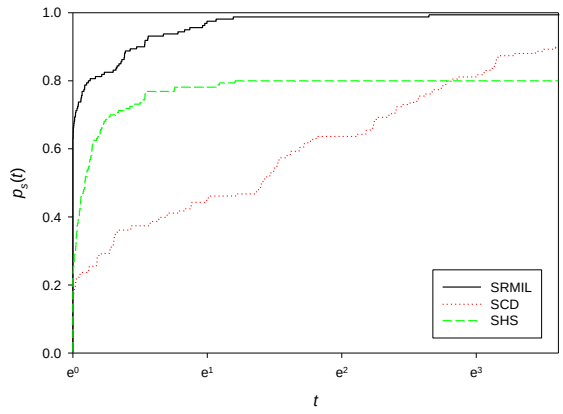

Fig. 2: The performance profile based on CPU Time

By referring on both Figure 1 and 2, the SRMIL is known to be most robust method as it solves almost all the test problems. Then, it is followed by the SCD and SHS respectively. By looking at the left sides of both graphs, the SRMIL has a better performance compared to SCD and SHS. Although SCD is more robust than SHS, but SHS is able to perform better than SCD. In conclusion, it is concluded that the SRMIL is the best method compared to the others.

5. Implementation

Nowadays, optimization has been widely studied for real life problems such as [18, 22]. In this research, all of the tested methods will be implemented in regression analysis. A real data set, Malaysia's bank lending rate taken from https://tradingeconomics.com/malaysia/bank-lending-rate is used to form a regression model. The data in Table 2 shows the rate from January to December 2017.

Table 2: Bank Lending Rate of 2017

Months, x_i	Bank Lending Rate, y_i
January	4.5400
February	4.6073
March	4.5975
April	4.5911
May	4.6137
June	4.4745
July	4.6097
August	4.6094
September	4.6233
October	4.6410
November	4.6086
December	4.6077

Based on Table 2, the first until eleventh data are used to form a linear regression model by the subsequent form,

$$y = a_0 + a_1 x_i.$$

The last data in Table 2 is reserved for error calculation. The relative error is obtained by applying the following formula,

$$\text{relative error} = \frac{|\text{exact value - approximate value}|}{|\text{exact value}|} \quad (5.1)$$

The least squares method is used to find the values of α_0 and α_1 obtained by minimizing the problem,

$$\min E(a) = \sum_{i=1}^{m} \left(y_i - \left(a_0 + a_1 x_i \right) \right)^2 . \quad (5.2)$$

The unconstrained problem are formed by converting (5.2) as below,

$$\min_{x \in R^2} E_i^2 = \sum_{i=1}^{n} [y_i - a(1, x_i)^T]^2. \quad (5.3)$$

The algorithm in Section 2 is used to compute the values of α_0 and α_1 for all tested CG methods. An initial point, (-5, 5) is selected to test all the CG methods. The results are stated as in Table 3.

Table 3: The Results for CG Methods

Methods	NOI / CPU Time	α_0	α_1
SRMIL	2 / 0.00278	4.56031090909093	0.0053436363637385
SCD	2 / 0.00279	4.56031090908824	0.0053436363410524
SHS	2 / 0.00287	4.56031090909091	0.0053436363637369

As in Table 3, the iteration number for these methods is 2. SRMIL have the least CPU time. The linear for each tested methods are listed in Table 4.

Table 4: The Linear Models

Methods	Result Approximation Function
Least Squares	$y = 0.0053436363636359x + 4.56031090909091$
SRMIL	$y = 0.0053436363637385x + 4.56031090909093$
SCD	$y = 0.0053436363410524x + 4.56031090908824$
SHS	$y = 0.0053436363637369x + 4.56031090909091$

Based on the approximation function in Table 4, replace $x = 12$ into the equation in order to find the estimation data for December. Then, the formula (5.1) is used to compute the relative error.

Table 5: The Estimated Data and Relative Error

Methods	Estimated Data for December	Relative Error
Least Squares	4.62443454545454	0.003631865237438
SRMIL	4.62443454545579	0.003631865237709
SCD	4.62443454518087	0.003631865178043
SHS	4.62443454545575	0.003631865237701

As in Table 5, SCD has the least relative error. However, the estimated data and relative error are same for all tested methods if they are taken to the ninth decimal points by 4.624434545 and 0.003631865 respectively. Thus, it can be concluded that the new proposed method is as convincing as the others.

6. Conclusion

Based on the numerical results between SRMIL, SCD and SHS, SRMIL has been proven to be the best method. SRMIL is also applicable in solving the real life problem. Hence, the SRMIL method is promising.

Acknowledgement

The authors would like to thank the editors and the referees for their suggestions and comments. The authors are also thankful to The Ministry of Higher Education of Malaysia (MOHE) for the sponsorship of this research via MyPhd Scheme.

References

[1] A. Abashar, M. Mamat, M. Rivaie, I. Mohd, Global convergence properties of a new class of conjugate gradient method for unconstrained optimization, Applied Math. Sci., 65-68 (2014) 3307-3319.

[2] A. Abashar, M. Mamat, M. Rivaie, I. Mohd, O. Omer, The proof of sufficient descent condition for a new type of conjugate gradient methods, AIP Conference Proceedings, 1602 (2014) 296-303.

[3] A.A. Goldstein, On steepest descent, SIAM J. Control., 3 (1965) 147-151.

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

[5] E. Dolan, J.J. More, Benchmarking optimization software with performance profile, Math. Prog., 91 (2002) 201–213.

[6] E. Polak, G. Ribiere, Note sur la convergence de directions conjugees, Rev. Francaise Inform. Recherche Operationelle, 3 (1969) 35–43.

[7] E.G. Birgin, J.M. Martinez, A spectral conjugate gradient method for unconstrained optimization, J. Appl. Maths. Optimization, 43 (2001) 117-128.

[8] G. Zoutendijk, Nonlinear programming computational methods, In J. Abadie (Ed.), Integer and nonlinear programming, Amsterdam: North Holland, pp. 37-86, 1970.

[9] J. Barzilai, J.M Borwein, Two-point step size gradient methods, IMA Journal of Numerical Analysis, 8 (1998) 141-148.

[10] J. Liu, Y. Jiang, Global convergence of a spectral conjugate gradient method for unconstrained optimization, Abstract and Applied Analysis, 2012, 1-12.

[11] K.E. Hilstrom, A simulation test approach to the evaluation of nonlinear optimization algorithms, ACM. Trans. Math. Softw., 3 (1977) 305–315.

[12] L. Armijo, Minimization of functions having Lipschitz continuous partial derivatives, Pacific J. Mathematics, 16 (1966) 1-3.

[13] M. Raydan. The Barzilain and Borwein gradient method for the large unconstrained minimization problem, SIAM Journal on Optimization, 7 (1997) 26-33.

[14] M. Rivaie, M. Mustafa, L.W. June, I. Mohd, A new class of nonlinear conjugate gradient coefficient with global convergence properties, Appl. Math. Comp., 218 (2012) 11323–11332.

[15] M.R. Hestenes, E. Steifel, Method of conjugate gradient for solving linear equations, J. Res. Nat. Bur. Stand., 49 (1952) 409–436.

[16] N. 'Aini, N. Hajar, M. Mamat, N. Zull, M. Rivaie, Hybrid quasi-Newton and conjugate gradient method for solving unconstrained optimization problems, Journal of Engineering and Applied Sciences., 12(18) (2017) 4627-4631.

[17] N. Andrei, An unconstrained optimization test functions collection, Adv. Modell. Optim., 10 (2008) 147–161.

[18] N. Hajar, N. 'Aini, N. Shapiee, Z.Z. Abidin, W. Khadijah, M. Rivaie, M. Mamat, A new modified conjugate gradient coefficient for solving system of linear equations, Journal of Physics: Conference Series, 890 (2017) 1-6.

[19] N. Zull, M. Rivaie, M. Mamat, Z Salleh, Z. Amani, Global convergence properties of a new spectral conjugate gradient by using strong wolfe line search, Applied Math. Sci., 9 (2015) 3105-3117.

[20] N. Zull, N. 'Aini, S. Shoid, N.H.A. Ghani, N.S. Mohamed, M. Rivaie, M. Mamat, A conjugate gradient method with descent properties under strong Wolfe line search, Journal of Physics: Conference Series, 890 (2017) 1-6.

[21] N.H.A. Ghani, N.S. Mohamed, N. Zull, S. Shoid, M. Rivaie, M. Mamat, Performance comparison of a new hybrid conjugate gradient method under exact and inexact line search, Journal of Physics: Conference Series, 890 (2017) 1-6.

[22] N.Z. Abidin, M. Mamat, B. Dangerfield, J.H. Zulkepli, M.A. Baten, A. Wibowo, Combating obesity through healthy eating behavior: A call for system dynamics optimization, Plos One, 9 (2014) 1-17.

[23] P. Wolfe, Convergence conditions for ascent method, SIAM Rev., 11 (1969) 226-235.

[24] R. Fletcher, C. Reeves, Function minimization by conjugate gradients, Comput. J., 7 (1964) 149-154.

[25] R. Fletcher, Practical method of optimization, John Wiley and Sons, 2013.

[26] W. Khadijah, M. Rivaie, M. Mamat, I. Jusoh. A spectral KRMI conjugate gradient method under the strong-Wolfe line search, AIP Conference Proceedings, 1739 (2016) 1-8.

[27] Y.H. Dai, Y. Yuan, An efficient hybrid conjugate gradient method for unconstrained optimization, Ann. Oper. Res., 103 (2002) 33-47.