International Journal of Engineering & Technology, 7 (3.28) (2018) 36-43

International Journal of Engineering & Technology

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

Research paper

A new family of conjugate gradient coefficient with application

Norrlaili Shapiee[1], Mohd Rivaie[2], Mustafa Mamat[1*], Puspa Liza Ghazali[3]

[1]*Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia*
[2]*Department of Computer Sciences and Mathematic, Universiti Teknologi MARA, Terengganu, Malaysia*
[3]*Faculty of Economic and Management Science, Universiti Sultan Zainal Abidin, Terengganu, Malaysia*
**Corresponding author E-mail: must@unisza.edu.my*

Abstract

Conjugate gradient (CG) methods are famous for their utilization in solving unconstrained optimization problems, particularly for large scale problems and have become more intriguing such as in engineering field. In this paper, we propose a new family of CG coefficient and apply in regression analysis. The global convergence is established by using exact and inexact line search. Numerical results are presented based on the number of iterations and CPU time. The findings show that our method is more efficient in comparison to some of the previous CG methods for a given standard test problems and successfully solve the real life problem.

Keywords: *conjugate gradient method; conjugate gradient coefficient; global convergence; line search; regression analysis.*

1. Introduction

The conjugate gradient (CG) method plays an important role in solving unconstrained optimization problems. Generally, an unconstrained optimization problem is written as

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

where $f: R^n \to R$ is continuously differentiable. The CG method is an iterative method of the form

$$x_{k+1} = x_k + \alpha_k d_k,\ k = 0,\ 1,\ 2,... \tag{1.2}$$

where x_{k+1} is the new iterate point, x_k is the current iterate point, $\alpha_k \geq 0$ is the stepsize and d_k is the search direction. The equation for calculating d_k is defined by

$$d_k = \begin{cases} -g_k & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 1, \end{cases} \tag{1.3}$$

where g_k is the gradient of $f(x)$ at the point x_k. The parameter $\beta_k \in R$ is known as the CG coefficient. Some examples of known β_k are Hestenes-Stiefel (*HS*) [14], Dai-Yuan (*DY*) [32], Rivaie, Mustafa, Ismail and Leong (*RMIL*) [17], Polak and Ribiere (PR) [5], Liu and Storey (LS) [37], Conjugate Descent (CD) [29], Fletcher and Reeves (FR) [30]. The corresponding formulas for the β_k mentioned are written as follows:

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

$$\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}},\quad \beta_k^{PR} = \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T g_{k-1}},$$

$$\beta_k^{LS} = \frac{g_k^T(g_k - g_{k-1})}{-d_{k-1}^T g_{k-1}},\quad \beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}},\quad \beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}},$$

These CG coefficients could likewise be divided into two groups. The first group comprises of PR, HS, RMIL and LS while in the second group we have FR, DY and CD. It is easy to see that the first group possesses the restart properties, whereas the second group does not have this characteristic [17-18]. In [20] has arranged the CG method into three distinct groups; the classical CG method, the scaled CG method and lastly the hybrid and parameterized CG methods. The classical CG method is the simplest and most straightforward to apply. However, it is difficult to find and produces a new CG method of this type [16].
As indicated to [34-35], all these methods are equivalent if the objective function is strictly convex quadratic. However, they behave distinctively when applied to general non quadratic functions. As mentioned in [33], the history of CG methods starts in [14] who initially proposed a CG method to solve a linear system of equation with a symmetric positive definite matrix. After that, in [30] applied the CG method to general unconstrained optimization problems. Nowadays, the CG methods are noted to be exceedingly valuable for solving large-scale unconstrained optimization problems since it needn't the storage of matrices [1-2, 8, 10-13, 15, 21-28, 31, 38, 39].
In this decade, a few new CG methods have been proposed. Some of the recent studies aim at creating a d_k that satisfies the sufficient descent condition and possess global convergence property. The earliest and well-known research on global convergence of CG methods is done by [6]. In that paper, the global convergence of FR method with exact line search is proven for general function. How-

ever, the PR method with exact line search is not globally convergent [7, 36].
It is well known that regression analysis regularly emerges in economics, finance, trade, meteorology, medicine biology, chemistry physics and etc. [39-43]. The classical regression model is defined by

$$Y = h\left(X_1, X_2, ..., X_p + \varepsilon\right) \tag{1.4}$$

where Y is the reaction variable, X_i is the indicator variable, $i = 1,2,.., p, p > 0$ is an integer constant and ε is the error term. The function $h\left(X_1, X_2, ..., X_p\right)$ clarify the type of relationship that exist between Y and $X = \left(X_1, X_2, ..., X_p\right)$. In this manner, we acquire the following linear regression model when h is a linear function

$$Y = a_0 + a_1 X_1 + a_2 X_2 + ... + a_p X_p + \varepsilon \tag{1.5}$$

which is the simplest regression model where $a_0, a_1, ..., a_p$ are the regression parameters. The most important errand in regression analysis is to evaluate the parameter $a = \left(a_0, a_1, ..., a_p\right)$ and the method of least squares is essential method to determine the parameters which is defined by

$$\min_{a \in R^{p+1}} S\left(a\right) = \sum_{i=1}^{m} \left(h_i - a_0 + a_1 X_{i1} + a_2 X_{i2} + ... + a_p X_{ip}\right)^2 \tag{1.6}$$

The contents of this paper are arranged into six sections. Section 2 introduces the new β_k and its algorithm. In Section 3, we show the proof of the sufficient descent condition and the global convergence property of our new method. Description of the problem by utilizing regression is displayed as a part of Section 4. Some fascinating numerical result is exhibited in Section 5, where our new method is compared any other CG method and Himmelblau's function is displayed. Discussions on the result are also included. Finally, a short conclusion is presented in Section 6.

2. New CG coefficient

In this section we propose our new CG coefficient known as β_k^{NRMI}. The NRMI signifies Norrlaili, Rivaie, Mustafa and Ismail. The β_k^{NRMI} is given as

$$\beta_k^{NRMI} = \frac{g_k^T\left(g_k - g_{k-1}\right)}{g_{k-1}^T\left(g_k - d_{k-1}\right)}. \tag{2.1}$$

The general algorithm of CG method utilized as a part of this study is as per the followings:
Step 1: Initialization. Given x_0, set $k = 0$.
Step 2: Compute β_k based on (2.1).
Step 3: Compute d_k based on (1.3). If $\|g_k\| = 0$, then stop.
Step 4: Compute α_k. Based on exact and inexact line search.
For exact line search compute $\alpha_k = \min_{\alpha \geq 0} f\left(x_k + \alpha_k d_k\right)$.
For inexact line search compute
α_k, $f\left(x_k + \alpha_k d_k\right) \leq f\left(x_k\right) + \delta\alpha_k g_k^T d_k$, and
$\left|g\left(x_k + \alpha_k d_k\right)^T d_k\right| \leq -\sigma g_k^T d_k$.
Step 5: Update new point based on (1.2).
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 1 with $k = k+1$.

3. Convergent analysis

In this section, we initially showed the sufficient descent condition and later on the proof of global convergence properties.

3.1. Convergent analysis based on exact line search

First, the convergent properties of β_k^{NRMI} will be concentrated based on the exact line search.

3.1.1. Sufficient descent condition

The convergent properties of β_k^{NRMI} will be studied. Firstly, we assume that every search direction d_k should satisfy the descent condition

$$g_k^T d_k < 0, \tag{3.1}$$

for all $k \geq 0$. If there exists a constant $C > 0$ for all $k \geq 0$, then the search directions satisfy following sufficient descent condition

$$g_k^T d_k \leq -C\|g_k\|^2. \tag{3.2}$$

Theorem 1: *Consider a CG method with the search direction* (1.3) *and* β_k^{NRMI} *given as* (2.1), *then condition* (3.2) *holds for all* $k \geq 0$.

Proof: If $k = 0$, then it is clear that $g_0^T d_0 = -C\|g_0\|^2$. Hence, condition (3.2) holds true. We also need to show that for $k \geq 1$, condition (3.2) will also hold true. From (1.3), multiply by g_{k+1}^T, then,

$$g_{k+1}^T d_{k+1} = g_{k+1}^T(-g_{k+1} + \beta_{k+1}^{NRMI} d_k) = -\|g_{k+1}\|^2 + \beta_{k+1}^{NRMI} g_{k+1}^T d_k \tag{3.3}$$

For exact line search, we realize that $g_{k+1}^T d_k = 0$. Thus,

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

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

3.1.2. Global convergence properties

Next, we will show that CG methods with β_k^{NRMI} converge globally. However, we first need to simplify our new β_k^{NRMI}, so that our convergence proof will be markedly easier [16]. From (2.1), we realize that

$$\beta_{k+1}^{NRMI} = \frac{g_{k+1}^T\left(g_{k+1} - g_k\right)}{g_k^T\left(g_{k+1} - d_k\right)} = \frac{\|g_{k+1}\|^2 - g_{k+1}^T g_k}{g_k^T g_{k+1} - g_k^T d_k}.$$

Hence, we get

$$\beta_{k+1}^{NRMI} \leq \frac{\|g_{k+1}\|^2}{\|g_k\|^2}. \tag{3.4}$$

The following basic assumptions are always needed in the analysis of global convergence properties of CG methods.

Assumption 1:

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

(ii) The gradient $g(x)$ is Lipschitz continuous in N, so a constant $L > 0$ exists such that $\|g(x) - g(y)\| \le L\|x - y\|$, for any $x, y \in N$.

Under this assumption, we have the following lemma which was proven by [6]. This lemma also holds for the exact minimization rule, the Goldstein and the Wolfe rule as shown in [10].

Lemma 1: *Suppose that Assumption 1 holds true. Consider any CG method of the form (1.3), where d_k is a descent search direction and α_k satisfies the exact minimization rule. Then, the following condition also known as the Zoutendijk condition holds*

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

The proof of this lemma can be seen from [9]. By using Lemma 1, we can obtain the following convergent theorem of the CG method using (3.4).

Theorem 2: *Suppose that Assumption 1 holds true. Consider any CG method in the form of (1.3) and (1.2,), where α_k is obtained by the exact minimization rule. Also, suppose that Assumption 1 and the descent condition hold true. Then, either*

$$\lim_{k\to\infty} \|g_k\| = 0 \quad \text{or} \quad \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty.$$

Proof: To prove Theorem 2, we use the contradiction approach. That is if Theorem 2 is not true, then a constant $c > 0$ exists such that

$$\|g_k\| \ge c. \tag{3.5}$$

Rewriting (1.3) as

$$d_{k+1} + g_{k+1} = \beta_{k+1} d_k$$

and squaring both sides of the equation, we obtain

$$\|d_{k+1}\|^2 = (\beta_{k+1})^2 \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2. \tag{3.6}$$

Dividing both side by $(g_{k+1}^T d_{k+1})^2$, then,

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} = \frac{(\beta_{k+1})^2 \|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} - \frac{2}{g_{k+1}^T d_{k+1}} - \frac{\|g_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2}$$

By completing the square,

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

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \le \frac{(\beta_{k+1})^2 \|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2}. \tag{3.7}$$

Applying (3.4) in (3.7) yields

$$\begin{aligned}\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} &\le \left(\frac{\|g_{k+1}\|^2}{\|g_k\|^2}\right)^2 \frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2}.\\ &\le \left(\frac{\|g_{k+1}\|^4}{\|g_k\|^4}\right) \frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2}.\\ &\le \frac{\|d_k\|^2}{\|g_k\|^4} + \frac{1}{\|g_{k+1}\|^2}.\end{aligned}$$

Therefore, we concluded that

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \le \frac{1}{\|g_{k+1}\|^2} \tag{3.8}$$

Hence,

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \le \frac{k}{c^2}$$

Therefore,

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \frac{c^2}{k} \tag{3.9}$$

This implies

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge c^2 \sum_{k=0}^{\infty} \frac{1}{k} = \infty.$$

Therefore, from (3.9) and (3.5), it follows that

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

This contradicts the Zoutendjik condition in Lemma 1. Therefore, the proof is completed.

Theorem 3: *Suppose that Assumption 1 holds true. Consider any CG methods in the form of* (1.3) *and* (1.2)*, where α_k is obtained by the exact minimization rule. Also, suppose that Assumption* 1 *and the descent condition hold true. Then, either*

$$\lim_{k\to\infty} \|g_k\| = 0 \quad \text{or} \quad \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|} < \infty.$$

Proof: From (3.6) and (3.4)

$$\begin{aligned}\|d_{k+1}\|^2 &= \left(\frac{\|g_{k+1}\|^2}{\|g_k\|^2}\right)^2 \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2.\\ &= \left(\frac{\|g_{k+1}\|^4}{\|g_k\|^4}\right) \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2.\end{aligned} \tag{3.10}$$

We have already proved that condition (3.2) holds. Therefore, we know that

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

Hence, from (3.10)

$$\|d_{k+1}\|^2=\frac{\|g_{k+1}\|^4\|d_k\|^2}{\|g_k\|^4}+2C\|g_{k+1}\|^2-\|g_{k+1}\|^2$$
$$=\frac{\|g_{k+1}\|^4\|d_k\|^2}{\|g_k\|^4}-\|g_{k+1}\|^2(1-2C) \quad (3.11)$$

Multiply both sides of (3.11) with $\frac{\|g_{k+1}\|^2}{\|d_{k+1}\|^2}$, then

$$\|d_{k+1}\|^2\frac{\|g_{k+1}\|^2}{\|d_{k+1}\|^2}=\frac{\|g_{k+1}\|^2}{\|d_{k+1}\|^2}\left(\frac{\|g_{k+1}\|^4\|d_k\|^2}{\|g_k\|^4}-\|g_{k+1}\|^2(1-2C)\right)$$
$$=\frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2}\left((2C-1)+\frac{\|g_{k+1}\|^2\|d_k\|^2}{\|g_k\|^4}\right)$$
$$\frac{\|d_{k+1}\|^2\|g_{k+1}\|^2}{\|d_{k+1}\|^2}\le\frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \quad (3.12)$$

Based on Theorem 2, we know that $\lim_{k\to\infty}\frac{\left(g_{k+1}^Td_{k+1}\right)^2}{\|d_{k+1}\|^2}<0.$ This will imply that if Theorem 3 is not true, then we have $\lim_{k\to\infty}\frac{\left(g_{k+1}^Td_{k+1}\right)^2}{\|d_{k+1}\|^2}=\infty$. From (3.12), we get $\infty\le\frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2}$ and $\infty\le\frac{\|g_{k+1}\|^6}{\|d_{k+1}\|^2}$ respectively. Hence, Theorem 3 holds true for sufficiently large k .

3.2. Convergent analysis based on inexact line search

In this section, the convergent properties of β_k^{NRMI} will be studied based on the inexact line search by means of strong Wolfe line search. We will also show that this CG coefficient will possess sufficient descent conditions under this line search. Under this inexact line search, we have

$$f(x_k+\alpha_kd_k)\le f(x_k)+\delta\alpha_kg_k^Td_k, \quad (3.13)$$

and

$$\left|g(x_k+\alpha_kd_k)^Td_k\right|\le-\sigma g_k^Td_k. \quad (3.14)$$

The following theorem which illustrates that the formula possesses the sufficient descent condition under the strong Wolfe line search. From (2.1), we realize that

$$\beta_{k+1}^{NRMI}=\frac{g_{k+1}^T(g_{k+1}-g_k)}{g_k^T(g_{k+1}-d_k)}=\frac{\|g_{k+1}\|^2-g_{k+1}^Tg_k}{g_k^Tg_{k+1}-g_k^Td_k}.$$

Hence, we get

$$\beta_{k+1}^{NRMI}\le\frac{\|g_{k+1}\|^2}{g_k^Tg_{k+1}}. \quad (3.15)$$

Theorem 4: Let the sequences $\{g_k\}$ and $\{d_k\}$ in the general algortihm, and let the stepsize α_k be determined by the SWP line search (3.13) and (3.14). If $\sigma\in(\delta,1)$, then the sufficient descent condition (3.2) holds.

Proof: From (3.15) and using (3.14), we get

$$\left|\beta_k^{NRMI}g_{k+1}^Td_k\right|\le\frac{\|g_{k+1}\|^2}{g_k^Tg_{k+1}}\sigma\left|g_k^Td_k\right|. \quad (3.16)$$

By (1.3), we have

$$\frac{g_{k+1}^Td_{k+1}}{\|g_{k+1}\|^2}=-1+\beta_k^{NRMI}\frac{g_{k+1}^Td_k}{\|g_{k+1}\|^2}. \quad (3.17)$$

We prove the descent property of $\{d_k\}$ by induction. Since $g_0^Td_0=-\|g_0\|^2<0$, if $g_0\ne0$, now we suppose that d_i , $i=1,2,...,k$ are all descent directions, for example $d_i^Tg_i<0$. By (3.16), we get

$$\left|\beta_k^{NRMI}g_{k+1}^Td_k\right|\le\frac{\|g_{k+1}\|^2}{g_k^Tg_{k+1}}\sigma\left(-g_k^Td_k\right). \quad (3.18)$$

that is

$$\frac{\|g_{k+1}\|^2}{g_k^Tg_{k+1}}\sigma g_k^Td_k\le\left|\beta_k^{NRMI}g_{k+1}^Td_k\right|\le-\frac{\|g_{k+1}\|^2}{g_k^Tg_{k+1}}\sigma g_k^Td_k. \quad (3.19)$$

However, from (3.17) together with (3.19), we deduce

$$-1+\sigma\frac{g_k^Td_k}{g_k^Tg_{k+1}}\le\frac{g_{k+1}^Td_{k+1}}{\|g_{k+1}\|^2}\le-1-\sigma\frac{g_k^Td_k}{g_k^Tg_{k+1}}, \quad (3.20)$$

and we know $\frac{1}{\|g_k\|^2}\le\frac{1}{g_k^Tg_{k+1}}.$

Replace (3.20), we get

$$-1+\sigma\frac{g_k^Td_k}{\|g_k\|^2}\le\frac{g_{k+1}^Td_{k+1}}{\|g_{k+1}\|^2}\le-1-\sigma\frac{g_k^Td_k}{\|g_k\|^2}. \quad (3.21)$$

Repeating this process and using the fact $d_0^Tg_0=-\|g_0\|^2$ imply

$$-\sum_{i=0}^{k}[\sigma]\le\frac{g_{k+1}^Td_{k+1}}{\|g_{k+1}\|^2}\le-2+\sum_{i=0}^{k}[\sigma]. \quad (3.22)$$

Then, (3.22) can be written as

$$-\frac{1}{1-\sigma}\le\frac{g_{k+1}^Td_{k+1}}{\|g_{k+1}\|^2}\le-2+\frac{1}{1-\sigma}. \quad (3.23)$$

Thus, by induction $g_k^Td_k<0$, holds for all $k\ge0.$ Denote $c=2-1/(1-\sigma)$, then $c\in(0,1)$ and (3.23) turns out to be

$$c-2\le\frac{g_k^Td_k}{\|g_k\|^2}\le-c \quad (3.24)$$

This implies that (3.2) holds. The proof is complete.

4. Description of the problem

In this section, the detailed description of the problem considered is given below.

Problem 1: In Table 1, there is data of some kind of commodity between year and index of road deaths in 2004 until 2014:

Table 1: Year and index of road deaths (2004-2014)

No.	Year (p)	Index of Road Deaths (d)
1	2004	4.51
2	2005	4.18
3	2006	3.98
4	2007	3.73
5	2008	3.63
6	2009	3.55
7	2010	3.40
8	2011	3.21
9	2012	3.05
10	2013	2.90
11	2014	2.66

From a statistical point of view, we can conclude that there will be possible changes in the index of road deaths although will occur road deaths per year. In summary, there will be an abatement in the index of road deaths for the years ahead and our main objective is to determine the function between the index of road deaths and the year which d is the regression equation for p.
From the given data above, one can observe that there exists a linear relationship between the year and the index of road deaths with the regression equation given by $\hat{d} = a_0 + a_1 p$, where a_0 and a_1 signifying the regression parameters. Solving the above regression equation includes finding the value of a_0 and a_1 by the method of least squares that minimized the problem

$$\min Q = \sum_{i=0}^{n} \left[d_i - \left(a_0 + a_1 p_i \right) \right]^2, \tag{4.1}$$

We can now change the above least squares problem into an unconstrained optimization problem as

$$\min_{x \in R^2} f(a) = \sum_{i=1}^{n} \left[d_i - a\left(1, p_i\right)^T \right]^2. \tag{4.2}$$

Solving the above problem (4.2) utilizing the linear least squares method yields the solution a = (4.5325,-0.1675). In this context, we utilize our proposed a new conjugate gradient to solve this problem solving regression problem compare with the linear least squares method [3].

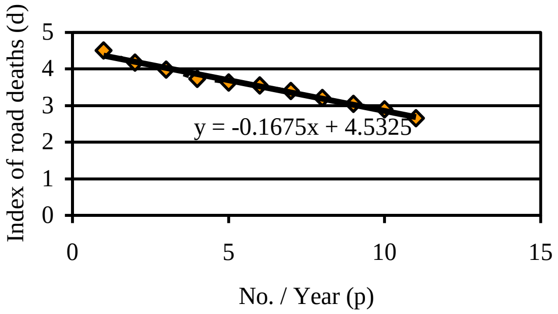

Fig. 1: Index of road deaths versus year for linear

Fig. 1 shows the graph plot for the index of road deaths by using trend line in Microsoft Excel. The trend line here shows a linear approximation function as $y = -0.1675x + 4.5325$.

5. Results and discussion

In this section, we carry out some numerical tests for FR, HS, NRMI and RMIL methods. Some test problems in [19] are selected to analyze the efficiency of β_k^{NRMI} based on the number of iterations and CPU time. The stopping criteria are set to be when $\|g_k\| \le 10^{-6}$. As suggested by [9], a test point should not be restricted to a point that is too close to the solution point. The best selected initial points should be based on the random number generator. However, we believe that this approach, will add to the complexity of the computer programming hence leading to high CPU time. Therefore, in this research we have selected four different initial points for each of the test problems. We start from a point that is close to the solution point and then move to the one that is furthest from it. These four initial points will also allow us to test the global convergence properties and the robustness of our method at the same time [17]. A list of problem functions and the initial points selected are shown in Table 2.
All the problems mentioned in Table 2 are solved using MATLAB2011b subroutine programming by using the exact and inexact line searches to obtain the stepsize. In some cases, the computation stops due to the line search failing to find a positive stepsize, thus the test is considered a failure. In order to describe the method's performance and to determine the best method, we use the Performance Profile introduced by [4].
We present Himmelblau's function (Fig. 2), which is a multimodal function to test the performance of the optimization algorithms. The function is defined as follows:

$$f\left(x_1, x_2\right) = \left(x_1^2 + x_2 - 11\right)^2 + \left(x_1 + x_2^2 - 7\right)^2.$$

This function is a summation of two squared terms. Each term inside the bracket can be considered as an error term. The first term calculate the distinction between the term $\left(x_1^2 + x_2\right)$ and 11 and the second term calculates the contrast between the term $\left(x_1 + x_2^2\right)$ and 7. Since the goal is to minimize these squared contrasts (or deviations of the two terms from 11 and 7 separately), the optimum solution will be a set of values of x_1 and x_2, fulfilling the following two equations.

$$x_1^2 + x_2 = 11, \qquad x_1 + x_2^2 = 7.$$

It has four indistinguishable four minima in this range (one in each quadrant). The locations of all the minima can be discovered logically. However, because they are roots of cubic polynomials, when written in terms of radicals, the expressions are fairly complicated.
Many engineering design problems expect to find a set of design parameters fulfilling various objectives simultaneously. In these problems, a mathematical expression for each objective is typically written and the distinction of the expression from the objective is calculated. The distinctions are then squared and included to form an overall objective function, which must be minimized. In this manner, the above Himmelblau's function work looks like the mathematical expression of an objective function in many engineering design problems. We utilized different initial points with general algorithm of CG method and Himmelblau's function. Every initial point gave a different solution point for every CG method, as shown in Table 2.

Table 2: A list of problem functions

No.	Function	n	Initial Points
1	Six hump	2	(2,2), (10,10), (36,36), (52,52)
2	Three hump	2	(5,5), (31,31), (41,41), (55,55)
3	Booth	2	(3,3), (27,27), (73,73), (100,100)
4	Treccani	2	(1,1), (42,42), (62,62), (91,91)
5	Zettl	2	(19,19), (31,31), (65,65), (95,95)
6	Leon	2	(3,3), (7,7), (10,10), (12,12)
7	Matyas	2	(2,2), (32,32), (62,62), (92,92)
8	Extended wood	4	(21,21, ...,21), (35, 35, ...,35), (51, 51, ...,51), (85, 85, ...,85)
9	Quartic	4	(1, 1, ...,1), (21, 21, ...,21), (41, 41, ...,41), (61, 61, ...,61)
10	Colville	4	(7, 7, ...,7), (17, 17, ...,17), (23, 23, ...,23), (77, 77, ...,77)
11	Powell	4	(14, 14, ...,14), (21, 21, ...,21), (84, 84, ...,84), (94, 94, ...,94)
12	Quadratic QF2	2,4, 10,100,500,1000	(16,16,…,16), (33,33,…,33), (52,52,…,52), (82,82,…,82)
13	Extended White and Holst	2,4, 10,100,500,1000	(3, 3, ...,3), (7, 7, ...,7), (10, 10, ...,10), (12,12,...,12)
14	Rosenbrock	2,4, 10,100,500,1000	(3, 3, ...,3), (39, 39, ...,39), (53, 53, ...,53), (61, 61, ...,61)
15	Extended Denschnb	2,4, 10,100,500,1000	(3, 3, ...,3), (5, 5, ...,5), (12, 12, ...,12), (50, 50, ...,50)
16	Shalow	2,4, 10,100,500,1000	(11, 11, ...,11), (18, 18, ...,18), (60, 60, ...,60), (64, 64, ...,64)
17	Extended Tridiagonal	2,4, 10,100,500,1000	(34, 34, ...,34), (41, 41, ...,41), (72, 72, ...,72), (93, 93, ...,93)
18	Extended Beale	2,4, 10,100,500,1000	(2, 2, ...,2), (6, 6, ...,6), (8, 8, ...,8), (11, 11, ...,11)
19	Diagonal 4	2,4, 10,100,500,1000	(4, 4, ...,4), (21, 21, ...,21), (50, 50, ...,50), (84, 84, ...,84)
20	Quadratic QF1	2,4, 10,100,500,1000	(1,1,…,1), (14,14,…,14), (31,31,…,31), (64,64,…,64)

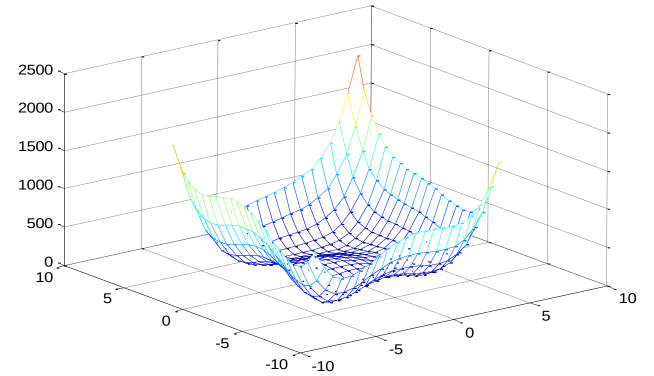

Fig. 2: Himmelblau's function

Table 3: The initial points corresponding the optimal points and number of iteration with the Himmelblau function by using exact line search

Initial Point	Function Value With Optimal Solution	Iteration			
		β_k^{NRMI}	β_k^{FR}	β_k^{HS}	β_k^{RMIL}
(1,1)	$f(3,2)=0$	6	13	6	10
(-1,-1)	$f(-3.7793,-3.2832)=0$	7	-	8	8
	$f(-2.8051,3.1313)=0$	-	25	-	-
	$f(3.5844,-1.8481)=0$	-	-	-	-
(10,10)	$f(3,2)=0$	6	13	6	9
	$f(-3.7793,-3.2832)=0$	-	-	-	-
(-5,-5)	$f(-3.7793,-3.2832)=0$	5	8	5	5
	$f(3,2)=0$	-	-	-	-
	$f(-2.8051,3.1313)=0$	-	-	-	-

Table 4: The initial points corresponding the optimal points and number of iteration with the Himmelblau function by using inexact line search

Initial Point	Function Value With Optimal Solution	Iteration			
		β_k^{NRMI}	β_k^{FR}	β_k^{HS}	β_k^{RMIL}
(1,1)	$f(3,2)=0$	7	13	7	10
(-1,-1)	$f(-3.7793,-3.2832)=0$	7	-	8	8
	$f(-2.8051,3.1313)=0$	-	-	-	-
	$f(3.5844,-1.8481)=0$	-	17	-	-
(10,10)	$f(3,2)=0$	-	-	-	-
	$f(-3.7793,-3.2832)=0$	6	10	6	6
(-5,-5)	$f(-3.7793,-3.2832)=0$	-	-	-	-
	$f(3,2)=0$	10	-	-	-
	$f(-2.8051,3.1313)=0$	-	2745	9	9

Every function in Table 2 is using to test four CG methods which are FR, HS, RMIL and NRMI to show their performance. The result each CG methods are shown below for exact and inexact line search in terms of iteration and CPU time.

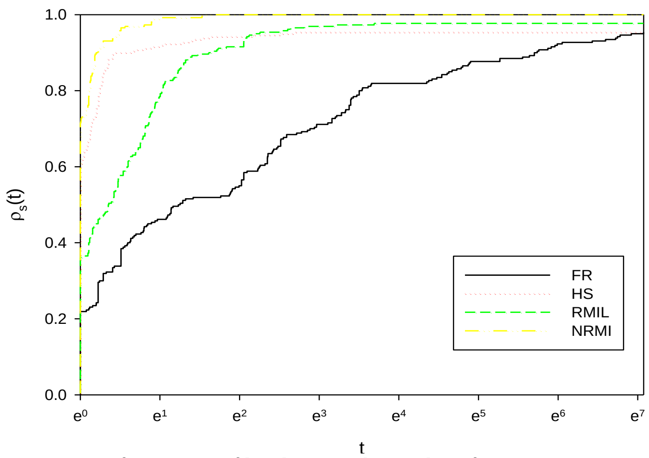

Fig. 3: Performance profile relative to the number of iterations (exact)

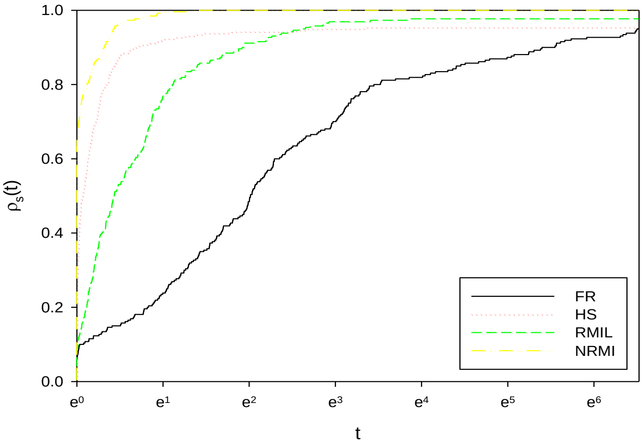

Fig. 4: Performance profile relative to the CPU time (exact)

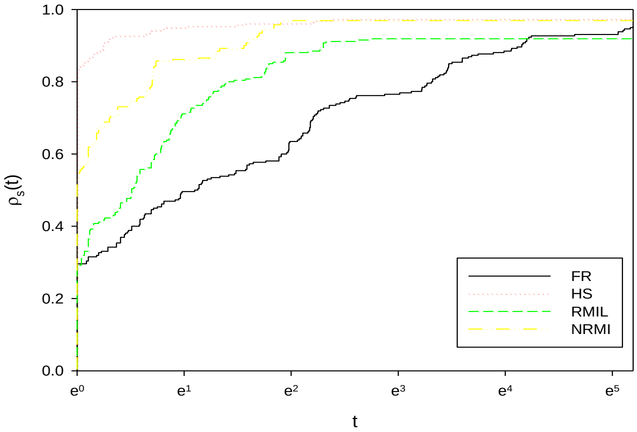

Fig. 5: Performance profile relative to the number of iterations (inexact)

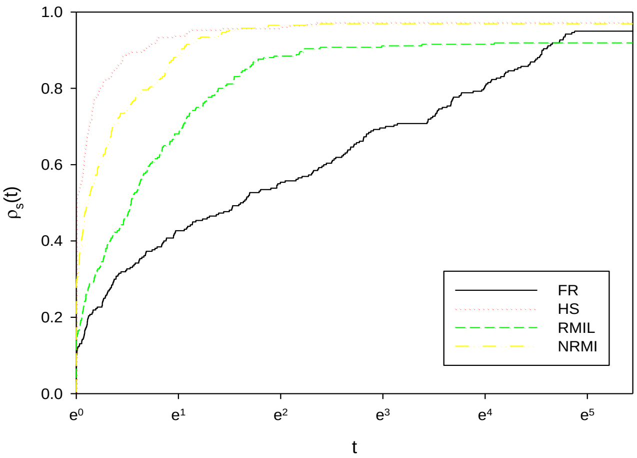

Fig. 6: Performance profile relative to the CPU time (inexact)

From both figures, the left side represents the method that is fastest in solving all of the test problems. On the other hand, the right side shows the method that successfully solved all the test problems. For exact line search, Fig. 3 and Fig. 4 show that NRMI has the best performance as it can solve all 100% of the test problems and the NRMI curve appears above FR, HS and RMIL curves. The HS and FR method solve only 95.38% of the problems, while a recently proposed method RMIL method solves 97.69% of the test problems. For inexact line search, Fig. 5 and Fig. 6 show that HS has the best performance as it can solve 97.31%. The NRMI can solve 96.92%, the FR can solve 95.38% and the RMIL can solve 91.92% of the test problems. Therefore, we can say that NRMI is considered superior and more effective compared to FR, HS, and RMIL in exact line search and HS is considered superior and more effective compared to FR, HS and RMIL in inexact line search.

The application of our proposed new conjugate gradient method implemented with the exact and inexact line search rule and different initial points are those chosen to test the efficiency of our method.

Table 5: Numerical results for problem 1

		i.p	Exact		Inexact	
			Iteration	CPU	Iteration	CPU
Index of road deaths (2004-2014)	linear	(5,5)	6	0.1341	2	0.4615
		(8,8)	5	0.0737	2	0.443
		(17,17)	6	0.6766	2	0.4439
		(32,32)	6	0.5683	2	0.4588

Table 6: Numerical results for problem 1 based on exact line search in terms of iteration

i.p	b	ε	Iteration
(5,5)	(4.532545418728772, -0.1675454504572175)	0.3120590949653934	6
(8,8)	4.532545410869674, -0.1675454482049973)	0.3120590954314233	5
(17,17)	(4.532545391532211, -0.1675454456193286)	0.3120590959663863	6
(32,32)	(4.532545459935292, -0.1675454550289031)	0.3120590940195767	6

Table 7: Numerical results for problem 1 based on exact line search in terms of CPU time

i.p	b	ε	CPU
(5,5)	(4.532545418728772, -0.1675454504572175)	0.3120590949653934	0.1341
(8,8)	4.532545410869674, -0.1675454482049973)	0.3120590954314233	0.0737
(17,17)	(4.532545391532211, -0.1675454456193286)	0.3120590959663863	0.6766
(32,32)	(4.532545459935292, -0.1675454550289031)	0.3120590940195767	0.5683

Table 8: Numerical results for problem 1 based on inexact line search in terms of iteration

i.p	b	ε	Iteration
(5,5)	(4.532545454545519, -0.1675454545451770)	0.3120590941196463	2
(8,8)	(4.532545454545442, -0.1675454545455390)	0.3120590941195714	2
(17,17)	(4.532545454545405, -0.1675454545458717)	0.3120590941195026	2
(32,32)	(4.532545454545186, -0.1675454545475912)	0.3120590941191467	2

Table 9: Numerical results for problem 1 based on inexact line search in terms of CPU time

i.p	b	ε	CPU
(5,5)	(4.532545454545519, -0.1675454545451770)	0.3120590941196463	0.4615
(8,8)	(4.532545454545442, -0.1675454545455390)	0.3120590941195714	0.4430
(17,17)	(4.532545454545405, -0.1675454545458717)	0.3120590941195026	0.4439
(32,32)	(4.532545454545186, -0.1675454545475912)	0.3120590941191467	0.4588

where i.p = initial point, a: the approximate solution from the method of linear least square = (4.5325, -0.1675), b: the solution as the program is terminated and rror, $\varepsilon = \|b-a\| / \|a\|$.

The numerical results of Tables show the performance of the algorithm considered in this application, the numerical results demonstrate that our algorithm has made significant performance. The performance from the implementation of problem 1 indicates that regardless of the different initial points, our proposed method was able to solve the problem.

6. Conclusion

In this paper, we propose a new classical CG method known as β_k^{NRMI} and solve real life regression problems. The convergent properties of this β_k^{NRMI} have been studied. For an algorithm to converge, it must fulfill the sufficient descent condition and global convergence properties. Based on the theoretical proof and the numerical result, we can say that this β_k^{NRMI} converges globally and performs better than other standard CG methods. Numerical results in application show successful for the test problems consider and can further be employed for application in regression analysis.

Acknowledgement

We would like to thank The Ministry of Higher Education of Malaysia for the funding of this research under Ministry of Education (Sponsorship MyPhD KPM).

References

[1] A. Abashar, M. Mamat, M. Rivaie, I. Mohd, Global Convergence Properties of a New Class of Conjugate Gradient Methods for Unconstrained Optimization, Applied Math. Sci., 8 (67) (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 Conf. Proc., 1602 (2014), 296-303.
[3] A. U. Moyi, L. W. June, I. Saidu, on the Application of Three-Term Conjugate Gradient Method in Regression Analysis, Int. J. Comput. Appl., (0975 8887) 102 (8) (2014).
[4] E. Dolan, J.J. More, Benchmarking optimization software with performance profile, Math. Prog., 91 (2002) 201–213.
[5] E. Polak, G. Ribiere, Note sur la convergence de directions conjugees, Rev. Francaise Inform. Recherche Operationelle, 3 (1969) 35–43.
[6] G. Zoutendijk, Nonlinear programming computational methods, in: Abadie J. (Ed.) Integer and nonlinear programming, North Holland, Amsterdam, 1970.
[7] I. Jusoh, M. Mustafa, M. Rivaie, A new edition of conjugate gradient methods for large-scale unconstrained optimization, Int. J. of Math. Anal., 8 (2014) 2277–2291.
[8] I. S. Mohammed, M. Mamat, A. Abashar, M. Rivaie, Z. Salleh, A Modified Nonlinear Conjugate Gradient Method for Unconstrained Optimization, Applied Math. Sci., 9 (54) (2015), 2671-2682.
[9] K.E. Hilstrom, A simulation test approach to the evaluation of nonlinear optimization algorithms, ACM. Trans. Math. Softw., 3 (1977) 305–315.
[10] K. U. Kamfa, M. Mamat, A. Abashar, M. Rivaie, P. L. Ghazali, Z Salleh, Another Modified Conjugate Gradient Coefficient with Global Convergence Properties, Applied Math. Sci., 9 (37) (2015), 1833-1844.
[11] M. Hamoda, A. Abashar, M. Mamat, M. Rivaie, A Comparative Study of Two New Conjugate Gradient Methods, AIP Conf. Proc., 1643 (2015), 616-621.
[12] M. Hamoda, M. Rivaie, M. Mamat, Z. Salleh, A Conjugate Gradient Method with Inexact Line Search for Unconstrained Optimization, Applied Math. Sci., 9 (37) (2015), 1823-1832.
[13] M. Mamat, M. Rivaie, I. Mohd, M. Fauzi, A New Conjugate Gradient Methods for Unconstrained Optimization, Int. J. Contemp. Math. Sciences, 5 (29) (2010), 1429-1437.
[14] M.R. Hestenes, E. Steifel, Method of conjugate gradient for solving linear equations, J. Res. Nat. Bur. Stand. 49 (1952) 409–436.
[15] M. Rivaie, A. Abashar, M. Mamat, I. Mohd, The Convergence Properties of a New Type of Conjugate Gradient Methods, Applied Math. Sci., 8 (1) (2014), 33-44.
[16] M. Rivaie, M. Mustafa, A. Abashar, A new class of nonlinear conjugate gradient coefficient with exact and inexact line searches, Appl. Math. Comp., 268 (2015), 1152–1163.
[17] 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.
[18] Rivaie, M., Mamat, M., Mohd, I., Fauzi, M.,Comparative study of conjugate gradient coefficient for unconstrained optimization (2011), Australian Journal of Basic and Applied Sciences, 5 (9), 947-951
[19] N. Andrei, An unconstrained optimization test functions collection, Adv. Modell. Optim., 10 (2008) 147–161.
[20] N. Andrei, 40 conjugate gradients algorithms for unconstrained optimization, Bull. Malays. Math. Sci. Soc., 34 (2011) 319–330.
[21] N. Hajar, M. Mamat, M. Rivaie, Z Salleh, A Combination of Polak-Ribiere and Hestenes-Steifel Coefficient in Conjugate Gradient Method for Unconstrained Optimization, Applied Math. Sci., 9 (63) (2015), 3119-3130.
[22] N. H. M. Yussoff, M. Mamat, M. Rivaie, I. Mohd, A New Conjugate Gradient Methods for Unconstrained Optimization with Sufficient Descent, AIP Conf. Proc., 1602 (2014), 514-519.
[23] N. Shapiee, M. R. M. Ali, M. Mamat, Z Salleh, A New Simple Conjugate Gradient Coefficient for Unconstrained Optimization, Applied Math. Sci., 9 (63) (2015), 3119-3130.
[24] N. Shapiee, M. Rivaie, M. Mamat, I. Mohd, A New Modification of Hestenes-Stiefel with Descent Properties, AIP Conf. Proc., 1602 (2014), 520-526.
[25] N. Shapiee, M. Rivaie, M. Mamat, A new classical conjugate gradient coefficient with exact line search, AIP Conf. Proc., 1739 (2016), 020082.
[26] 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 (63) (2015), 3105-3117.
[27] O. Omer, M. Mamat, A. Abashar, M. Rivaie, The Global Convergence Properties of a Conjugate Gradient Method, AIP Conf. Proc., 1602 (2014), 286-295.
[28] R. B. Yunus, M. Mamat, A. Abashar, M. Rivaie, Z Salleh, Z. A. Zakaria, The Convergence Properties of a New Kind of Conjugate Gradient Methods for Unconstrained Optimization, Applied Math. Sci., 9 (38) (2015), 1845-1856.
[29] R. Fletcher, Practical Method of Optimization, Vol. 1, Unconstrained Optimization, John Wiley and Sons, New York, 1987.
[30] R. Fletcher, C. Reeves, Function minimization by conjugate gradients, Comput., J. 7 (1964) 149-154.
[31] S. Shoid, M. Rivaie, M. Mamat, I. Mohd, Solving Unconstrained Optimization with a New Type of Conjugate Gradient Method, AIP Conf. Proc. 1602 (2014), 574-579.
[32] Y.H. Dai, Y. Yuan, A nonlinear conjugate gradient with a strong global convergence properties, SIAM J. Optim., 10 (1999) 177–182.
[33] Y. H. Dai and Y. Yuan, A note on the nonlinear conjugate gradient method, J. Compt. Appl. Math., 18 (2002), 575-582.
[34] Y. H. Dai and Y. Yuan, Nonlinear conjugate gradient method, Beijing, Shanghai Scientific and Technical Publishers, (1998).
[35] Y. Yuan and W. Sun, Theory and methods of optimization, Beijing, Science Press of China, (1999).
[36] Y. Narushima, H. Yabe and J. A. Ford, A three-term conjugate gradient method with sufficient descent property for unconstrained optimization, SIAM J. Optim., 21 (2011), 212-230.
[37] Y. Liu, C. Storey, Efficient generalized conjugate gradient algorithms part 1: Theory, J. Comput. Appl. Math. 69 (1992) 129–137.
[38] Z. Z. Abidin, M. Mamat, M. Rivaie, A new steepest descent method with global convergence properties, AIP Conf. Proc., 1739 (2016), 020070.
[39] N.H.A Ghani, M. Rivaie, M. Mamat, A Modified form of Conjugate Gradient method for Unconstrained Optimization Problems, AIP Conf. Proc., 1739 (2016), 020076.
[40] Abidin, N.Z., Mamat, M., Dangerfield, B., Zulkepli, J.H., Baten, Md. A., Wibowo, A. Combating obesity through healthy eating behavior: A call for system dynamics optimization (2014) Plos One, 9 (12), e114135.
[41] Mamat, M., Subiyanto, Kartono, A. Mathematical model of cancer treatments using immunotherapy, chemotherapy and biochemotherapy (2013) Applied Mathematical Sciences, 7 (5-8), 247-261.
[42] Mamat, M., Deraman, S.K., Noor, N.M.M., Rokhayati, Y., Diet problem and nutrient requirement using Fuzzy Linear Programming Approach (2012) Asian Journal of Applied Sciences, 5 (1), 52-59.
[43] Mamat, M., Rokhayati, Y., Noor, N.M.M., Mohd, I., Optimizing human diet problem with fuzzy price using fuzzy linear programming approach (2011) Pakistan Journal of Nutrition, 10 (6), 594-559.