Jurnal
Teknologi

Full Paper

A NEW COEFFICIENT OF CONJUGATE GRADIENT METHODS FOR NONLINEAR UNCONSTRAINED OPTIMIZATION

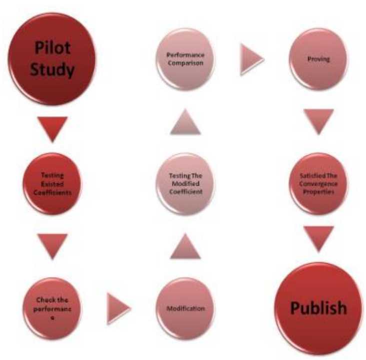
Article history
Received
13 July 2015
Received in revised form
9 December 2015
Accepted
12 January 2016

Nur Syarafina Mohamed[a]*, Mustafa Mamat[a], Fatma Susilawati Mohamad[a], Mohd Rivaie[b]

[a]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (Unisza), Campus Tembila Besut, Terengganu, Malaysia
[b]Department of Computer Sciences and Mathematics, Universiti Teknologi MARA (UiTM) Terengganu, Malaysia

*Corresponding author
sl0633@putra.unisza.edu.my

Graphical abstract



Abstract

Conjugate gradient (CG) methods are widely used in solving nonlinear unconstrained optimization problems such as designs, economics, physics and engineering due to its low computational memory requirement. In this paper, a new modifications of CG coefficient (β_k) which possessed global convergence properties is proposed by using exact line search. Based on the number of iterations and central processing unit (CPU) time, the numerical results show that the new β_k performs better than some other well known CG methods under some standard test functions.

Keywords: Conjugate gradient method, conjugate gradient coefficient, exact line search, global convergence, and unconstrained optimization

1.0 INTRODUCTION

In optimization, the nonlinear CG method is a useful method in finding the minimum value. Considering the form below;

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

where $f: R^n \to R$ is a continuously differentiable function which is bounded from below. Starting from an initial guess at point x_0 , a nonlinear conjugate gradient algorithm generates a sequence of points $\{x_k\}$, according to the following iterative formula:

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

Where x_k is the current iterate point whilst $\alpha_k > 0$ is a step size, which is obtained by one dimensional search known as the search direction, d_k . In this paper, exact line search is used as shown in (3) below

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

The search direction, d_k is defined by

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

(4)

where g_k is the gradient of $f(x)$ at the point x_k . Some of the well known β_k are given as follows:

$$\beta_k^{FR} = \frac{g_k^T g_k}{\|g_{k-1}\|^2} \quad (5)$$

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

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

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

(8)
$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}$$

(9)
$$\beta_k^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{\|d_{k-1}\|^2} \quad (10)$$

where g_k and g_{k-1} denotes the gradients of $f(x)$ at the point x_k and x_{k-1} respectively. For the above corresponding methods, FR is known as Fletcher and Reeves [1], PR is Polak and Ribierre [2], HS is known as Hestenes and Steifel [3], DY is Dai and Yuan [4], CD is Conjugate Descent by Fletcher [5] and lastly RMIL denotes for Rivaie, Mustafa, Ismail and Leong [6]. The symbol of Euclidean norm of vectors is presented by ‖.‖. According to Dai and Yuan [7] and Yuan and Sun [8], if $f(x)$ is strictly convex quadratic function, all these methods are said to be equivalent, but they behave differently for general non quadratic functions.

Based on the history of CG methods which can be seen in [9], in 1952, Hestenes and Stiefel [3] first time proposed a CG method to solve a linear system of equation with a symmetric positive definite matrix, or equivalently, for minimizing a strictly convex quadratic function. After that, in 1964 Fletcher and Reeves [1] applied the CG method to general unconstrained optimization problems. Nowadays CG methods are used as iterative methods for solving large-scale unconstrained optimization problems since the storage of matrices is not needed.

In this decade, many other CG methods have been proposed. Some recent research aims at generating a search direction satisfying the descent condition $g_k^T d_k < 0$ for all k and sufficient descent condition; i.e., there exists a positive constant c such that

$$g_k^T d_k < -c\|g_k\|^2$$

for all k holds to show global convergence. For the global convergence properties, the earliest most well known research is by Zoutendijk [10]. In that paper the global convergence of FR method is proven by using exact line search. A general condition on scalar β_k which ensures the global convergence of nonlinear conjugate gradient method in the case of strong Wolfe inexact line searches could be found in [11].

When the function is a strong convex quadratic, the CG method is said to be identical where the line search is exact. The performances will vary when applied to general nonlinear functions with inexact line search, [12].

In this paper a new CG coefficient β_k is proposed based on the already proven β_k with the exact line search. Section two will discuss the motivation and the new β_k together with the new algorithm of CG method. Section three will continue with convergence analysis of this new β_k together with its proofs. This paper used small-scale data which consists of ten problem functions with various dimension. In result section, it shows a graphical comparison of the new β_k with other β_k namely the RMIL, PR, HS, FR, DY and CD. The paper ended with a conclusion section.

2.0 THE NEW CG COEFFICIENT

In this section, the new β_k proposed is known as β_k^{SMR}. The motivation of this β_k^{SMR} comes from β_k^{RMIL} where the denominator of β_k^{SMR} is retained as same as β_k^{RMIL}, which is $\|d_{k-1}\|^2$. From (10), the numerator of β_k^{RMIL} is given as $g_k^T(g_k - g_{k-1})$. This numerator is also same as the numerator used in (7) and (8). The numerator acts as a restart properties to avoid problems associated with jamming, [13]. By expanding this expression, we get $g_k^T g_k - g_k^T g_{k-1}$ which implies $\|g_k\|^2 - (g_k^T g_{k-1})$. In preventing any negative value of β_k, some modifications has been proposed. Hence;

$$\beta_k^{SMR} = \max\left\{0, \frac{\|g_k\|^2 - |g_k^T g_{k-1}|}{\|d_{k-1}\|^2}\right\} \quad (11)$$

Based on this β_k^{SMR}, a complete algorithm of CG method could be generated as follows:

Step 1: Initialization. Set $k = 0$ and select x_0

Step 2: Compute β_k^{SMR} based on (11)

Step 3: Compute search directions d_k based on (4). If $\|g_k\| = 0$, then stop.

Step 4: Solve α_k using the exact line search, α_k based on (3).

Step 5: Updating new initial point using (2)

Step 6: Convergent test and stopping criteria. If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| \le \varepsilon$ then stop. Otherwise go to Step 2 with $k = k+1$.

3.0 CONVERGENCE ANALYSIS

This section will discuss about the convergence properties of β_k^{SMR} where the sufficient descent condition and the global convergence properties must hold in order for an algorithm to converge. Some of the proof is almost similar to the proof of β_k^{RMIL}, see [14].

3.1 Sufficient Descent Condition

For the sufficient condition to hold, then

$$g_k^T d_k \le -C\|g_k\|^2 \text{ for } k \ge 0 \text{ and } C > 0 \tag{12}$$

Theorem 1
Consider a CG method with search direction (4) and β_k^{SMR} defined as (11), then, condition (12) will holds for all $k \ge 0$

Proof:
From (4), if, $k = 0$ then $g_0^T d_0 = -C\|g_0\|^2$. Hence, condition (12) hold. In order to show condition (12) also hold for $k \ge 1$, multiply (4) by g_k^T . Then,

$$g_k^T d_k = -g_k^T g_k + \beta_k^{SMR} g_k^T d_{k-1}$$
$$= -\|g_k\|^2 + \beta_k^{SMR} g_k^T d_{k-1} \tag{13}$$

Since the line search is exact, it implies $g_k^T d_{k-1} = 0$. Thus,

$$g_k^T d_k = -\|g_k\|^2 .$$

Hence, the descent condition holds, $g_k^T d_k \le -C\|g_k\|^2$. Proof completed. ■

3.2 Global Convergence Properties

Next, a new coefficient of CG methods with β_k^{SMR} must converges globally to fulfil the convergence properties. Before any step is proceeding, β_k^{SMR} need to be simplified to make proving step much easier. From (11),

$$\beta_{k+1}^{SMR} = \max\left\{0, \frac{\|g_{k+1}\|^2 - |g_{k+1}^T g_k|}{\|d_k\|^2}\right\}$$

$$\beta_k^{SMR} = \begin{cases} 0 & \text{for } \|g_{k+1}\|^2 < |g_{k+1}^T g_{k-1}| \\ \dfrac{\|g_{k+1}\|^2 - |g_{k+1}^T g_k|}{\|d_k\|^2} \le \dfrac{\|g_{k+1}\|^2}{\|d_k\|^2} & \end{cases}$$

Hence,

$$0 \le \beta_{k+1}^{SMR} \le \frac{\|g_{k+1}\|^2}{\|d_k\|^2} \tag{14}$$

In order to prove (14), the following assumptions are needed in the analysis of CG methods global convergence properties.

Assumption 1
(i) f is bounded below on the level set R^n and is continuous and differentiable in a neighbourhood N of the level set $\ell = \{x \in R^n \mid f(x) \le f(x_0)\}$ at the initial point x_0
(ii) The gradient $g(x)$ is Lipschitz continuous in N, so there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \le L\|x - y\| \text{ for any } x, y \in N$$

Under this assumptions, the following lemma is obtained, which was proved by Zoutendijk [15]. This lemma also holds for exact minimization rule, Goldstein and Wolfe rule shown in [16].

Lemma 1
Suppose the Assumption 1 holds. Consider any CG methods of the form (4) where d_k is a descent search direction and α_k satisfy the exact minimization rules. Then the following conditions know as Zoutendijk conditions holds

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

By using this lemma, the following convergence theorem of the conjugate gradient method can be obtained by using (14).

Theorem 2
Suppose that Assumption 1 holds. Consider any CG methods in the form of (4) and (2) where α_k is obtained by the exact minimization rules. Suppose that Assumption 1 and the descent condition hold. 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

Theorem 2 is proved by using contradiction. That is, if Theorem 2 is not true then, there exists a constant $c>0$ such that

$$\|g_k\|\geq c \tag{15}$$

Rewriting (4),

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

Squaring both side

$$\|d_{k+1}\|^2=(\beta_{k+1})^2\|d_k\|^2-2g_{k+1}^T d_{k+1}-\|g_{k+1}\|^2 \tag{16}$$

Dividing both side with $\left(g_{k+1}^T d_{k+1}\right)^2$

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

By using completing the square,

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

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

Applying (14) yield

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

$$=\frac{\|g_{k+1}\|^2}{\|d_k\|^2\|d_{k+1}\|^2}+\frac{1}{\|g_{k+1}\|^2}$$

$$\leq\frac{1}{\|d_k\|^2}+\frac{1}{\|g_{k+1}\|^2} \tag{17}$$

By noting that $\frac{1}{\|d_0\|^2}=\frac{1}{\|g_0\|^2}$, then from (17),

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

Hence,

$$\frac{\|d_k\|^2}{(g_k d_k)^2}\leq\sum_{i=0}^{k}\frac{1}{\|g_i\|^2}$$

$$\frac{(g_k d_k)^2}{\|d_k\|^2}\geq\frac{c^2}{k} \tag{18}$$

Therefore from (18) and (15), it follows that

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

This contradicts the Zoutendijk condition in Lemma 2. Therefore the proof is completed. ■

Corollary 1

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

Proof

By using contradiction, assume that $\|g_k\|\geq c$ and $\sum_{k=0}^{\infty}\|d_k\|^2=\infty$. For $\|g_k\|\to\infty$, yield $\frac{1}{\|g_k\|}\to 0$. From (17),

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2}\leq\frac{1}{\|d_k\|^2}$$

$$\|d_k\|^2\leq\frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2}$$

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

This contradicts the Zoutendijk conditions. Hence the corollary holds. ■

4.0 NUMERICAL RESULTS AND DISCUSSIONS

In this section, the efficiency of the new algorithm is analysed. Considering the test functions proposed by Andrei [2], the performance of SMR is tested comparing with RMIL, PR, FR, HS, DY and CD. Stopping criteria is set to be $\|g_k\|\leq 10^{-6}$ where $\varepsilon=10^{-6}$. testing an algorithm on a relatively large set of test functions is bothersome because it requires the coding of the functions. However, cynical observer can easily obtained if the algorithm is not tested on a large number of functions. It tends to conclude that the algorithm was tuned to particular functions. Even aside from the cynical observer, the algorithm is just not well tested, [17]. According to Hillstrom [18], for each of the test functions, random starting point must be choose from a box surrounding the standard starting point.

Table 1 List of test functions used

No	Function	n	Initial Points
1	Three Hump	2	$(1,-1), (-1,1), (-2,2), (2,-2)$
2	Six Hump	2	$(8,8), (-8,-8), (10,10), (-10,-10)$
3	Goldstein-Price	2	$(2,-2), (5,-5), (10,-10), (13,-13)$
4	Himmelblau	2,4,10,100, 500,1000	$(10,10,\ldots,10), (50,50,\ldots,50), (100,100,\ldots,100), (200,200,\ldots,200)$
5	Rosenbrock	2,4,10,100, 500,1000	$(13,13,\ldots,13), (16,16,\ldots,16), (20,20,\ldots,20), (30,30,\ldots,30)$
6	Denschnb	2,4,10,100, 500,1000	$(5,5,\ldots,5), (8,8,\ldots,8), (13,13,\ldots,13), (25,25,\ldots,25)$
7	Beale	2,4,10,100, 500,1000	$(2,2,\ldots,2), (5,5,\ldots,5), (8,8,\ldots,8), (10,10,\ldots,10)$
8	Tridiagonal 1	2,4,10,100, 500,1000	$(10,10,\ldots,10), (12,12,\ldots,12), (17,17,\ldots,17), (20,20,\ldots,20)$
9	Generalized Quartic	2,4,10,100, 500,1000	$(10,10,\ldots,10), (50,50,\ldots,50), (100,100,\ldots,100), (200,200,\ldots,200)$
10	Diagonal 4	2,4,10,100, 500,1000	$(10,10,\ldots,10), (50,50,\ldots,50), (100,100,\ldots,100), (200,200,\ldots,200)$

A list of ten test functions and its initial points involved are shown in Table 1. All test functions mentioned above is solved using MATLAB subroutine program with an Intel Core i7-3470 CPU processor .The comparison of our new algorithm numerical results are based on the number of iterations and CPU time.

The numerical results of SMR, RMIL, PR, FR, HS, DY and CD will be compared based on number of iterations and CPU time. The performances are presented graphically in Figure 1 and Figure 2 by using performance profile initiated by Dolan and More [19]. Performance profile is used to find how well the solvers perform relative to the other solvers. In general, $P_s(t)$ is the fraction of problems with performance ratio t, thus, a solver are said to be preferable when it has the higher values of $P_s(t)$. In a set of problems P and a set S of optimization solvers, performance on problem $p \in P$ is compared by a particular algorithm $s \in S$ with the best performance by any solver. Let $t_{p,s}$ denotes the number of iterations or CPU time required when solving problem $p \in P$ by the method $s \in S$. The performance ratio is defined by $r_{p,s} = \dfrac{t_{p,s}}{\min\{t_{p,s} : s \in S\}}$. It is assumed that $r_{p,s} \in [1, r_M]$ and $r_{p,s} = r_M$ only when problem p is not solved by solver s. Define, $P_s(t) = \dfrac{1}{n_p} \text{size}\{p \in P : r_{p,s} \le t\}$ as uniform fraction of $\dfrac{1}{n_p}$. Then the graph $P_s(t)$ versus $t \in [1, r_M]$ is plotted. Since the smallest performance ratio is 1 and it will be located at the most left of t - axis, hence in a graph of performance profile, the top curve represents the most efficient method.

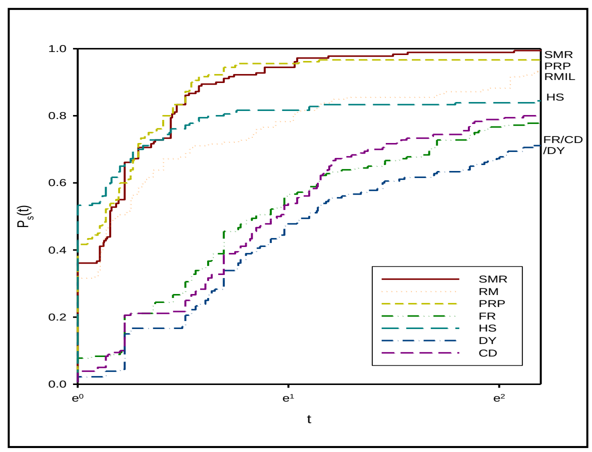


Figure 1 Performance profile based on number of iterations

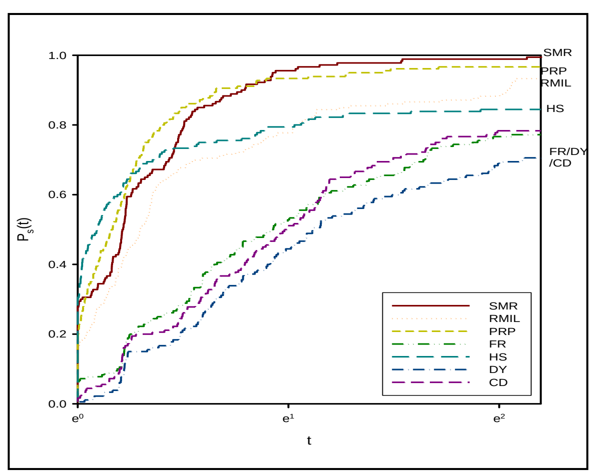


Figure 2 Performance profile based on CPU time

Figure 1 and Figure 2, show clearly that this new algorithm is better than other methods in term of number of iterations and CPU time. Table 2 shows the comparison of the effectiveness for each of the methods based on percentage analysis.

Table 2 Percentage analysis

Method	Percentage (%)
RMIL	93.28%
PRP	97.09%
FR	78.29%
HS	84.76%
DY	71.52%
CD	79.76%
SMR	99.75%

5.0 CONCLUSIONS

In this paper a new β_k is proposed and have been have proved that it is globally convergence under descent condition. Numerical results have shown that the new β_k performs better than other methods. For future study, numerical testing should be done for large-scale problems so that the new CG algorithm will become a new conjugate gradient family.

Acknowledgement

This work was supported by MyBrain(MyPhD) of Ministry of Higher Education (MOHE), Malaysia.

References

[1] Fletcher, R. and Reeves, C. 1964. Function Minimization By Conjugate Gradients. *Computational Journal*. 7: 149-154.

[2] Polak, E. and Ribiere, G. 1969. Rev. Francaise Informat Recherche Operationelle. *3E Annee*. 16: 35-43.

[3] Hestenes, M.R. and Steifel, E. 1962. Method Of Conjugate Gradient For Solving Linear Equations. *J. Res. Nat. Bur. Stand.* 49: 409-436.

[4] Dai, Y. H. and Yuan, Y. 2000. A Nonlinear Conjugate Gradient with Strong Global Convergence Properties. *SIAM J. Optim.* 10: 177-182.

[5] Fletcher, R. 1987. *Practical Methods Of Unconstrained Optimization*. New York: J. Wiley and Sons.

[6] Rivaie, Mamat M., June, M. L. W. and Mohd, I. 2012. A New Class Of Nonlinear Conjugate Gradient Coefficients With Global Convergence Properties. *Applied Mathematics and Computation*. 218: 11323-11332.

[7] Dai, Y. H. and Yuan, Y. 1998. *Nonlinear Conjugate Gradient Method*. Beijing: Shanghai Scientific and Technical Publishers.

[8] Yuan, Y. and Sun, W. 1999. *Theory And Methods Of Optimization*. Beijing: Science Press of China.

[9] Narushima, Y. Yabe, H. and Ford, J. A. 2011. Globally Convergent Three-Term Conjugate Gradient Methods that Use Secant Conditions and Generate Descent Search Directions for Unconstrained Optimization. *SIAM J. Optim*. 21: 212-230.

[10] Zoutendijk, G. 1970. Nonlinear programming computational methods. *Abadie J. (Ed.). Integer and Nonlinear Programming*. 37-86.

[11] Dai, Y. H. and Yuan, Y. 2002. Modified Conjugate Gradient Method for Unconstrained Optimization. *J. Compt. Appl. Math*. 18(6): 575-582.

[12] Hamoda, M. Rivaie, M. Mamat, M. Salleh, Z. and Amani, Z. 2015. A New Nonlinear Conjugate Gradient Coefficient for Unconstrained Optimization. *Applied Mathematical Sciences*. 9(37): 1813-1822.

[13] Hajar, N. Mamat, M. Rivaie, M. and Salleh, Z. 2015. A Combination of Polak-Ribiere and Hestenes-Stiefel Coefficient in Conjugate Gradient Method for Unconstrained Optimization. *Applied Mathematical Sciences*. 9(63): 3131-3142.

[14] Rivaie, M. Abashar, A. Mamat, M. Mohd, I. 2014. The Convergence Properties of a New a Type of Conjugate Gradient Methods. *Applied Mathematical Sciences*. 8(1): 33-34.

[15] Zoutendijk, G. 1970. Nonlinear Programming Computational Methods. *Abadie J. (Ed.). Ineteger and Nonlinear Programming*. North Holland, Amsterdam.

[16] Yuan, G. Lu, S. and Wei, Z. 2010. A Line Search Algorithm For Unconstrained Optimization. *J. Soft. Eng. App*. 3: 409-436.

[17] More, J. J. Garbow, B. and Hillstrom, K. E. 1981. Testing Unconstrained Optimization Software. *Journal ACM Transaction on Mathematical Software*. 7(1): 17-41.

[18] Hilstrom, K. E. 1977. A Simulation Test Approach To The Evaluation Of Nonlinear Optimization Algorithms. *A.C.M. Trans. Maths. Softw.* 3(4): 305-315.

[19] Dolan, E. and More, J. J. 2002. Benchmarking Optimization Software With Performance Profile. *Maths.Prog.* 91(2): 201-213.