International Journal of Engineering & Technology, 7 (3.28) (2018) 12-16

International Journal of Engineering & Technology

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

Research paper

An application of conjugate gradient method under strong Wolfe line search for solving unconstrained optimization

Wan Khadijah[1*], Mohd Rivaie[2], Mustafa Mamat[1] Nurul Hajar[1], Nurul 'Aini[1], Zubai'ah Zainal Abidin[1]

[1]*Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA), 21300 Terengganu, Malaysia*
[2]*Department of Sciences and Mathematics, Universiti Teknologi MARA (UiTM), 21080 Terengganu, Malaysia*
**Corresponding author E-mail: wankhadijahws@gmail.com*

Abstract

The conjugate gradient (CG) method is one of the most prominent methods for solving linear and nonlinear problems in optimization. In this paper, we propose a CG method with sufficient descent property under strong Wolfe line search. The proposed CG method is then applied to solve systems of linear equations. The numerical results obtained from the tests are evaluated based on number iteration and CPU time and then analyzed through performance profile. In order to examine its efficiency, the performance of our CG formula is compared to that of other CG methods. The results show that the proposed CG formula has better performance than the other tested CG methods.

Keywords: *Conjugate Gradient Method; Spectral Conjugate Gradient; Strong Wolfe Line Search.*

1. Introduction

Consider the unconstrained optimization problem:

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

where $f:R^n\to R$ is continuously differentiable. The CG methods are generated iteratively by:

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

with search direction, d_k 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 x_k is the current iterate point, $\alpha_k>0$ is a positive stepsize, g_k is a gradient coefficient and β_k is a scalar known as the CG coefficient.
In this paper, the inexact strong-Wolfe line search is used, where the α_k satisfies two conditions as follow:

$$\begin{aligned} &f(x_k+\alpha_k d_k)\leq f(x_k)+\rho\alpha_k d_k^T g_k,\\ &\left|g(x_k+\alpha_k d_k)^T d_k\right|\leq -\sigma d_k^T g_k,\end{aligned} \tag{1.4}$$

with $0<\rho<\sigma<1$. The stepsize is obtained via solving a one-dimensional minimization problem known as the line search method. The value of the α_k determines how far the function is to be minimized along the direction, d_k to get the best x_{k+1}.
Most CG method differs based on their CG coefficient, β_k. Some commonly used and well-known formulas of β_k are presented as follows:

$$\beta_k^{HS}=\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}},\ \beta_k^{FR}=\frac{g_k^T g_k}{\|g_{k-1}\|^2},\ \beta_k^{PRP}=\frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2},$$
$$\beta_k^{CD}=-\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}},\ \beta_k^{LS}=\frac{g_k^T y_{k-1}}{-d_{k-1}^T g_{k-1}},\ \beta_k^{DY}=\frac{g_k^T g_k}{d_{k-1}^T y_{k-1}}.$$

Note that the symbol $\|\cdot\|$ denotes the Euclidean norm of vectors and $y_{k-1}=g_k-g_{k-1}$. The methods above are referred as HS [1], FR [2], PRP [3], LS [4], DY [5] and CD [6].
In [7, 8] published two new CG coefficients which are modified based on the HS method namely the RMIL and RMIL* methods. The methods are both easily defined and simple to implement. They are formulated as:

$$\beta_k^{RMIL}=\frac{g_k^T(g_k-g_{k-1})}{d_{k-1}^T(d_{k-1}-g_k)} \quad \text{and} \quad \beta_k^{RMIL^*}=\frac{g_k^T(g_k-g_{k-1})}{\|d_{k-1}\|^2}.$$

These methods have good numerical results and are globally convergent with exact line search.
Other researchers, like [9] are interested to modify the PRP formula as

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

$$\beta_k^{WYL} = \frac{g_k^T\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{g_{k-1}^T g_{k-1}}$$

Their paper discusses the general convergence results for the proposed formula with some line searches such as the exact line search, the Wolfe–Powell line search and the Grippo–Lucidi line search. Recently, there are many researchers interested in the WYL approach. For further research, please refer to [10-14]. For good comparative studies on newer CG methods, please refer to [15-19]. For newer studies on the optimization method, please refer to [20-21].

Inspired by the spectral method proposed by [22-23] combined the advantages of spectral CG method and FR CG method and presented the modified FR (MFR) method. When applied with Armijo-type line search, MFR has been shown to possess global convergence properties. Its search direction is defined as follow:

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -\theta_k g_k + \beta_k^{FR} d_{k-1}, & \text{if } k \geq 1. \end{cases} \quad (1.5)$$

where θ_k is written as

$$\theta_k = \frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2}.$$

The MFR method can be rewritten as:

$$d_k = -\left(1 + \beta_k^{FR}\frac{g_k^T d_{k-1}}{\|g_k\|^2}\right)g_k + \beta_k^{FR} d_{k-1}. \quad (1.6)$$

In [24] also applied the DY conjugate gradient method in the MFR method where the β_k^{DY} is substituted in the (1.6) formula.

In [25] proposed a spectral Khadijah-Rivaie-Mustafa-Ismail (SKRMI) CG method under strong Wolfe line search to solve unconstrained optimization problems. The method is presented as follow:

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

Generally, most of the spectral method directly reduces to classical CG method if exact line search is used.

In this paper, in [23] approach as written in equation (1.6) is implemented with the RMIL*classical CG method. This modification of RMIL* method is named as MRMIL* CG method. Under strong Wolfe type line search, MRMIL* method is proven to possess sufficient descent and global convergence properties. Other than that, the new proposed method is utilized for solving artificial nonlinear problems and systems of linear equations in order to test their efficiency and applicability in real-life problems. The new method is only tested on numerical experimental simulations as it is difficult to get actual data for real-life application.

Then, this paper is organized as follows. In section 2, we present the new spectral CG method and its algorithm. In section 3, the proof of the proposed method which includes the convergence analysis of the modified RMIL* method is presented. Then, section 4 reports the numerical results and discuss the findings. In section 5, the proposed method is tested for solving systems of linear equations and the results obtained are discussed. Finally, the work of this paper is summarized in the last section.

2. New conjugate gradient algorithm

In this section, we present a modification of spectral CG method where the β_k^{RMIL*} is applied. This method is motivated by [23]. The equation of this new method is shown as follows:

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -\left(1 + \beta_k^{RMIL*}\frac{g_k^T d_{k-1}}{\|g_k\|^2}\right)g_k + \beta_k^{RMIL*} d_{k-1}, & \text{if } k \geq 1. \end{cases} \quad (2.1)$$

The following is the idea of MRMIL* algorithm.

Algorithm 1: Algorithm of modified RMIL*.

Step 1: Initialization. Given x_0, set $k = 0$.

Step 2: Compute β_k based on β_k^{RMIL*}.

Step 3: Compute d_k based on (2.1). If $g_k = 0$, then stop.

Step 4: Compute α_k based on (1.4).

Step 5: Update the new point based on (1.2).

Step 6: Convergence test and stopping criteria.

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

3. The sufficient descent condition

The spectral CG method reduces to standard RMIL* method if exact line search is applied in the algorithm. This is because under exact line search, $g_k^T d_{k-1} = 0$. Hence, we will have

$$\theta_k = 1 + \beta_k^{RMIL*}\frac{g_k^T d_{k-1}}{\|g_k\|^2} = 1. \quad (3.1)$$

From there, it is obvious that the search direction of spectral CG method will be reduced to the standard CG method. For the sufficient condition to hold,

$$g_k^T d_k \leq -C\|g_k\|^2 \text{ for } k \geq 0 \text{ where } C > 0. \quad (3.2)$$

It is easy to see that for (2.1) the sufficient descent condition holds for any line search since $g_k^T d_k = -\|g_k\|^2$.

Theorem 3.1: Let x_k and d_k be generated by the MRMIL*, then condition (3.2) holds for all $k \geq 0$.

Proof: If $k = 0$, then $g_0^T d_0 = -C\|g_0\|^2$. Hence, condition (3.2) holds true. We also need to show that for $k \geq 1$, condition (3.2) will also hold true. From (2.1), multiply with g_k^T,

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

$$= -\|g_k\|^2 - \beta_k^{RMIL}\frac{g_k^T d_{k-1}}{\|g_k\|^2}\|g_k\|^2 + \beta_k^{RMIL} g_k^T d_{k-1}$$

By applying strong Wolfe line search,

$$= -\|g_k\|^2 - \beta_k^{RMIL}\frac{\sigma g_k^T d_k}{\|g_k\|^2}\|g_k\|^2 + \beta_k^{RMIL}\sigma g_k^T d_k$$

$$= -\|g_k\|^2 - \sigma\beta_k^{RMIL} g_k^T d_k + \sigma\beta_k^{RMIL} g_k^T d_k$$
$$= -\|g_k\|^2 .$$
$$\therefore g_k^T d_k \le -C\|g_k\|^2 .$$

Thus, it is clearly shown that the sufficient descent condition is fulfilled. Hence, the proof is completed.

4. Results and discussion

In this section, the numerical results for MRMIL* based on the number of iteration and CPU time are compared with other methods in order to analyze the efficiency of the proposed spectral methods.

We consider $\varepsilon = 10^{-6}$ and norm of gradient value as the stopping criteria. All the problems are solved by MATLAB R2011b and the calculations are performed by using Intel Core i5 processor with RAM 2GB and Windows 8 operation system.

Table 1 presents the list of problem functions selected from [26]. For each of the test function, four initial points are considered starting from a point that is close to the solution point to the point that is far from it, as suggested by [27].

Table 1: List of test problem functions

No.	Function	Dimension	Initial Points
1	Extended White and Holst	2,4, 10, 100, 500, 1000	(-3,...,-3), (3,...,3), (6,...,6), (9,...,9)
2	Extended Rosenbrock	2,4, 10, 100, 500, 1000	(13,...,13), (16,...,16), (20,...,20), (30,...,30)
3	Extended DENSCHNB	2,4, 10, 100, 500, 1000	(10,...,10), (50,...,50), (100,...,100), (200,...,200)
4	Extended Himmelblau	2,4, 10, 100, 500, 1000	(10,...,10), (50,...,50), (100,...,100), (200,...,200)
5	Quadratic QF1	2,4, 10, 100, 500, 1000	(10,...,10),(50,...,50), (100,...,100),(200,...,200)
6	Extended Tridiagonal 1	2,4, 10, 100, 500, 1000	(10,...,10),(50,...,50), (100,...,100),(200,...,200)
7	Extended Maratos	2,4, 10, 100, 500, 1000	(10,...,10),(50,...,50), (100,...,100),(200,...,200)
8	Extended DENSCHF	2,4, 10, 100, 500, 1000	(10,...,10),(50,...,50), (100,...,100),(200,...,200)
9	Diagonal 4	2,4, 10, 100, 500, 1000	(10,...,10),(50,...,50), (100,...,100),(200,...,200)
10	Raydan 2 Function	2,4, 10, 100, 500, 1000	(5,...,5), (10,...,10), (25,...,25), (50,...,50)
11	Extended Freudenstein and Roth	2,4, 10, 100, 500, 1000	(2,...,2), (20,...,20), (75,...,75), (202,...,202)

Then, the numerical results are analyzed via the performance profile which is first introduced by [28]. The performance results are shown in Figure 1 and Figure 2.

This method of performance profile provide the means to assess and compare the performance of the set solvers S on a test set P. Assuming that n_s solvers and n_p problem exist, for each problem p and solver s, they defined:

$t_{p,s}$ = Calculating time (the number of iterations, CPU time or others) required to solve problem p by solver s.

For the baseline of comparisons, the performance on problem p by solver s are compared with the best performance by any solver on the same problem. This is referred as the performance ratio, which is defined by:

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s} : s \in S\}}.$$

The parameter $r_M \ge r_{p,s}$ for all p, s, and $r_{p,s} = r_M$ if and only if solver s does not solve problem p. The performance of solver s on any given problem might be of significance, but we would like to acquire a general evaluation of the performance of the solver, then the following is defined:

$$\rho_s(\tau) = \frac{1}{n_p}\text{size}\{p \in P : r_{p,s} \le \tau\}.$$

Thus, $\rho_s(\tau)$ is the probability for the solver $s \in S$ that a performance ratio $r_{p,s}$ is within a factor $\tau \in R$ of the best possible ratio. The function ρ_s is the cumulative distribution function for the performance ratio. The performance profile $p_s : R \mapsto [0,1]$ for a solver is non-decreasing, piecewise and continuous from the right. The value $p_s(1)$ is the probability that the solver is successful over the rest of the solvers.

For better understanding about the performance profiles, note that the left side of the figure gives the percentage of the test problems for which method is the fastest; while the right side gives the percentage of the test problems that are successfully solved by each of the methods. The methods that solve the most of the problems in a time that is within a factor τ of the best time can be determined by the top curve of the graph in the figures.

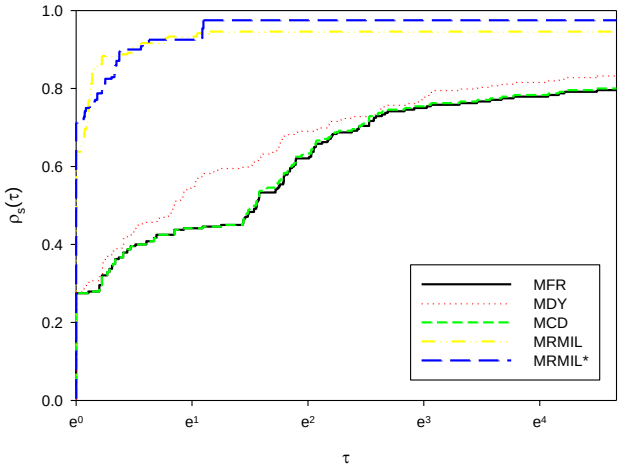


Fig. 1: Performance Profile based on Iteration Number

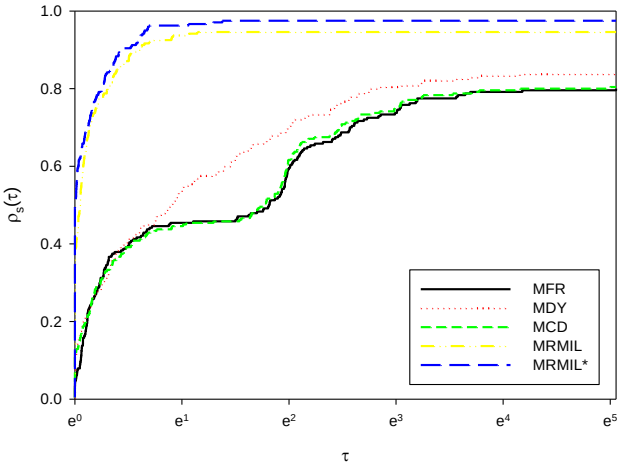


Fig. 2: Performance Profile based on CPU Time

In both figures, MFR, MDY, MCD, MRMIL and MRMIL* represent the spectral CG methods used to solve the test problems under strong Wolfe line search. Fig. 1 represents performance profile based on the number of iterations and Fig. 2 represents performance profile based on CPU time. From both figures, MRMIL solves 94.58%, MRMIL* solves 97.5%, MFR solves 80%, MDY solves 83.75% and MCD solves 80.42%. It is clear in both Fig.1 and Fig. 2 that the performance of MRMIL* CG method is the best with respect to the number of iterations and CPU time. This is followed by MRMIL, MDY, MCD and lastly MFR method.

5. Application of conjugate gradient method

In this section, the CG method is applied for solving the systems of linear equations. Consider a system of linear equations

$$Ax = b,$$

where $A \in R^{mxn}$, $b \in R^m$, $m \geq n$ and rank $A = n$. The matrix $Ax = b$, can be solved directly by $x = A^{-1}b$ if the matrix is a symmetric or square matrix. However, in cases involving non-square matrix problem, the following theorem can be used which is further discussed in this section.

Theorem 5.1 [29]: The unique vector x^* that minimizes $\|Ax - b\|^2$ is given by the solution of the equation $A^T Ax = A^T b$; that is

$$x = \left(A^T A\right)^{-1} A^T b \, .$$

The proof of this theorem can be seen in [29]. The following problem is cited from [29], where the Theorem 5.1 is applied to solve the problem effectively. The problem can be formulated as

$$A = \begin{bmatrix} 0.3 & 0.1 \\ 0.4 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}, \quad b = \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix},$$

where the decision variable is $x = \left[x_1, x_2\right]^T$. This paper only deals with small scale application of CG method to test the capability of CG method for solving systems of linear equation.

Table 2: Result of Application in No. of Iteration

Initial points	MRMIL	MRMIL*	MFR	MDY	MCD
(10,10)	2	2	2	2	2
(50,50)	2	2	2	2	2
(100,100)	2	2	2	2	2
(200,200)	2	2	2	2	2

Table 3: Result of Application in CPU Time

Initial Points	MRMIL	MRMIL*	MFR	MDY	MCD
(10,10)	0.0042	0.0038	0.0033	0.0043	0.0034
(50,50)	0.0028	0.0027	0.0029	0.0028	0.0034
(100,100)	0.0026	0.0027	0.0027	0.0026	0.0033
(200,200)	0.0026	0.0027	0.0033	0.0026	0.0035

Table 4: CPU Time per Iteration

Method	CPU Time per Iteration
MRMIL	0.001525
MRMIL*	0.001488
MFR	0.001525
MDY	0.001538
MCD	0.001700

From Table 2 and Table 3, all of the tested methods can solve the presented problem. However, MRMIL* method shows the best overall result compared to the other methods as it gives the least number of CPU time per iteration. Thus, based on the results of the numerical test, we may conclude that the MRMIL* method is applicable for real-life application.

6. Conclusion

In this paper, we propose a new version of RMIL* CG method under strong Wolfe line search. The classical RMIL* coefficient is combined with spectral CG method, namely the MRMIL* method. The results show that the modified MRMIL* method performs the best overall compared to the other modified CG methods since MRMIL* converges faster and is able to successfully solve most of the test problems. Based on the promising results, the MRMIL* method should also be applicable in real-life problem.

Acknowledgement

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

References

[1] Hestenes MR & Steifel E (1952), Method of Conjugate Gradient for Solving Linear Equations, *J. Res. Nat. Bur. Stand.* 49, 409-436.
[2] Fletcher R & Reeves C (1964), Function Minimization by Conjugate Gradients, *Comput. J.* 7, 149-154.
[3] Polak E and Ribiere G (1969), Note Sur La Convergence de Directions Conjugees, *Rev. Francaise Informat Recherche Operationalle* 3, 35-43.
[4] Liu Y and Storey C (1992), Efficient Generalized Conjugate Gradient Algorithm Part 1: Theory, *J. Comput. Appl. Math.* 69, 129-137.
[5] Dai YH and Yuan Y (1999), A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property, *SIAM J. Optim.* 10, 177-182.
[6] Fletcher R, *Practical method of optimization, Unconstrained Optimization*, 1, John Wiley and Sons, New York, (1987).
[7] Rivaie M, Mustafa M, Mohd I, and Fauzi M (2010), Modified Hestenes-Steifel Conjugate Gradient Coefficient for Unconstrained Optimization, *Journal of Interdisciplinary Maths.* 13(3), 241-251.
[8] Rivaie M, Mustafa M, June LW and Mohd I (2012), A New Class of Nonlinear Conjugate Gradient Coefficient with Global Convergence Properties, *Appl. Math. Comp.* 218, 11323-11332.
[9] Wei Z, Yao S and Liu L (2006), The Convergence Properties of Some New Conjugate Gradient Methods, *Applied Maths. And Comp.* 183, 1341-1350.
[10] Hajar N, Mamat M, Rivaie M and Jusoh I (2016), A New Type of Descent Conjugate Gradient Method with Exact Line Search, *AIP Conf. Proc*. 1739, 020089.
[11] Ghani NHA, Rivaie M and Mamat M (2016), A Modified Form of Conjugate Gradient Method for Unconstrained Optimization Problems, *AIP Conf. Proc.* 1739, 020076.
[12] Ghani NHA, Rivaie M and Mamat M (2017), A New Family of Polak-Ribiere-Polyak Conjugate Gradient Method with the Strong-Wolfe Line Search**,** *AIP Conf. Proc.* 1870, 040060.
[13] Shoid S, Rivaie M and Mamat M (2016), A Modification of Classical Conjugate Gradient Method Using Strong Wolfe Line Search, *AIP Conf. Proc.* 1739, 020071.
[14] Du X, Zhang P and Ma W (2016), Some Modified Conjugate Gradient Methods for Unconstrained Optimization, *Journal of Comp. and Applied Maths.* 305, 92-114.
[15] Mohamed NS, Mamat M and Rivaie M (2016), Solving a Large Scale Nonlinear Unconstrained Optimization with Exact Line Search Direction by Using New Coefficient of Conjugate Gradient Methods, *AIP Conf. Proc.* 1787, 080018.
[16] Mohamed NS, Mamat M and Rivaie M (2017), A New Nonlinear Conjugate Gradient Coefficient under Strong Wolfe-Powell Line Search, *AIP Conf. Proc.* 1870, 040055.
[17] Shapiee N, Rivaie M and Mamat M (2016), A New Classical Conjugate Gradient Coefficient with Exact Line Search, *AIP Conf. Proc.* 1739, 020082.
[18] Shapiee N, Rivaie M and Mamat M (2015), A New Simple Conjugate Gradient Coefficient for Unconstrained Optimization, *Applied Mathematical Sciences* 9 (63), 3119 – 3130.
[19] Khadijah W, Rivaie M and Mamat M (2017), A Three-Term Conjugate Gradient Method under the Strong-Wolfe Line Search, *AIP Conf. Proc.* 1870, 040056.
[20] 'Aini N, Rivaie M and Mamat M (2016), A Modified Conjugate Gradient Coefficient with Inexact Line Search for Unconstrained Optimization, *AIP Conf. Proc.* 1787, 080019.
[21] Abidin ZZ, Mamat M and Rivaie M (2016), A New Steepest Descent Method with Global Convergence Properties, *AIP Conf. Proc.* 1739, 020070.
[22] Birgin EG and Martinez JM (2001), A Spectral Conjugate Gradient Method for Unconstrained Optimization, *Appl. Math Optim*. 43, 117-128.
[23] Zhang L, Zhou W and Li D (2006), Global Convergence of a Modified Fletcher-Reeves Conjugate Gradient Method with Armijo-

Type Line Search, *Numer. Math.* 104, 561-572.
[24] Zhang L. and Zhou W (2008), Two Descent Hybrid Conjugate Gradient Methods for Optimization, *J. Comp. and Appl. Math.* 216, 251-264.
[25] Khadijah W, Rivaie M, Mamat M and Jusoh I (2016), A Spectral KRMI Conjugate Gradient Method under the Strong-Wolfe Line Search, *AIP Conf. Proc.* 1739, 020072.
[26] Andrei N (2008), An Unconstrained Optimization Test Functions Collection, *Adv. Model. Optim.* 10, 147-161.
[27] Hilstrom KE (1997), A Simulation Test Approach to Evaluation of Nonlinear Optimization Algorithms, *A.C.M. Trans. Maths. Softw.* 3, 305-315.
[28] Dolan E and More JJ (2002), Benchmarking Optimization Software with Performance Profile, *Maths. Prog.* 91, 201-213.
[29] Chong EKP and Zak SH, *An Introduction to Optimization*, 3rd Edition, John Wiley and Sons, New Jersey (2008).