Applied Mathematical Sciences, Vol. 9, 2015, no. 37, 1813 - 1822
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ams.2015.411994

A New Nonlinear Conjugate Gradient Coefficient for Unconstrained Optimization

[1*]Mohamed Hamoda, [2]Mohd Rivaie, [3]Mustafa Mamat and [1]Zabidin Salleh

[1]School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu (UMT), 21030 Kuala Terengganu, Malaysia

[2]Department of Computer Science and Mathematics, Univesiti Teknologi MARA (UITM) 23000 Terengganu, Malaysia

[3]Department of Computer Science and Mathematics, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, 22200 Terengganu, Malaysia

Abstract

In this paper, we suggest a new nonlinear conjugate gradient method for solving large scale unconstrained optimization problems. We prove that the new conjugate gradient coefficient β_k with exact line search is globally convergent. Preliminary numerical results with a set of 116 unconstrained optimization problems show that β_k is very promising and efficient when compared to the other conjugate gradient coefficients Fletcher - Reeves (FR) and Polak -Ribiere – Polyak (PRP).

Keywords: Conjugate gradient coefficient, exact line search, global convergence, large scale, unconstrained optimization

1. Introduction

In this paper, we focus our attention on the unconstrained optimization problem,

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

Where $f : R^n \to R$ is continuously differentiable function and R^n denotes an *n*-dimensional Euclidean space. We denote by $g(x)$, the gradient of f at x. The conjugate gradient (CG) method is the best methods for solving (1.1), especially when the dimension is large. The iterates of the CG method for solving (1.1) are obtained by

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

Where x_k is current iterate point and the α_k is step size. The step size is computed by carrying out some line search, for example, the exact line search where,

$$\alpha_k = \arg\min_{\alpha \geq 0} f(x_k + \alpha d_k) \tag{1.3}$$

The d_k is the search direction defined by

$$d_k = \begin{cases} -g_k & \text{if} \quad k = 0, \\ -g_k + \beta_k d_{k-1} & \text{if} \quad k \geq 1, \end{cases} \tag{1.4}$$

Where $g_k = g(x_k)$ and β_k is a scalar. The most well-known classical formula for β_k are the Hestenes-Stiefe (HS) method [11]. The Fletcher – Reeves (FR) method [7]. The Polak-Ribiere -Polyak (PRP) method [15, 16]. The conjugate descent (CD) method[6]. The Liu – Storey (LS) method [14] and the Dai – Yuan (DY) method [2]. The parameters of these β_k as follows

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} \tag{1.5}$$

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \tag{1.6}$$

$$\beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}} \tag{1.7}$$

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}} \tag{1.8}$$

$$\beta_k^{LS} = -\frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T g_{k-1}} \tag{1.9}$$

$$\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}} \tag{1.10}$$

The most studied properties of CG methods are its global convergence properties. Zoutendijk [22] and Powell [17] proved that FR method with exact line search is globally convergent. Zhang et al [13], Proposed a modified FR method MFR which is globally convergent under inexact line search. Polyak [16] and Powell [18] showed that PRP has a good numerical performance, but does not have such good convergence property. Touati-Ahmed and Storey [20], Gilbert and Nocedal [8] gave another way to discuss the global convergence of the PRP method with the weak Wolfe – Powell line search, where the parameter β_k in (1.6) is not allowed to be negative, $\beta_k = \max\{\beta_k^{PRP}, 0\}$, therefore, during the past few years, many authors has been investigated to create new formula for β_k ,[3, 4, 9, 10, 19, 21].

In this paper, we will show a new β_k in section 2. In section 3, we will study the sufficient descent condition and the global convergence proof of the new β_k. In section 4, we present the numerical results and discussion. Finally, we present the conclusions in section 5.

2. New β_k parameter and algorithm

In this section, we present a modified of PRP method which is known as β_k^{MRM}, where MRM denotes Mohamed, Rivaie and Mustafa, β_k^{MRM} is defined by,

$$\beta_k^{MRM} = \frac{g_k^T\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}\right)}{\|g_{k-1}\|^2 + \left|g_k^T d_{k-1}\right|} \tag{2.1}$$

The following algorithm is a general algorithm for solving optimization by CG methods.

Algorithm (2.1)

Step 1: Given $x_0 \in R^n, \varepsilon \geq 0$, set $d_0 = -g_0$ if $\|g_0\| \leq \varepsilon$ then stop.

Step 2: Compute α_k by exact line search (1.3).

Step 3: Let $x_{k+1} = x_k + \alpha_k d_k$, $g_{k+1} = g(x_{k+1})$ if $\|g_{k+1}\| < \varepsilon$ then stop.

Step 4: Compute β_k by formula (2.1), and generate d_{k+1} by (1.4).

Step 5: Set $k = k+1$ go to Step 2.

The following assumptions are often used in the studies of the conjugate gradient methods.

Assumption A. $f(x)$ is bounded from below on the level set $\Omega = \{x \in R^n, f(x) \leq f(x_0)\}$, where x_0 is the starting point.

Assumption B. In some neighbourhood N of, the objective function is continuously differentiable, and its gradient is Lipschitz continuous, that is there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\| \ \forall x, y \in N . \tag{2.2}$$

3. The Global Convergence properties

In this section, we study the global convergent properties of β_k^{MRM} , first we need to simplify the β_k^{MRM} , so that the proof will be easier. From (2.1) we know that

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

$$\beta_k^{MRM} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{\|g_{k-1}\|^2 + |g_k^T d_{k-1}|} \geq \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} |g_k^T g_{k-1}|}{\|g_{k-1}\|^2 + |g_k^T d_{k-1}|} \geq \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} \|g_k\|\|g_{k-1}\|}{\|g_{k-1}\|^2 + |g_k^T d_{k-1}|} = 0$$

Thus we get, $\beta_k^{MRM} \geq 0$

Also

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

$$\leq \frac{\|g_k\|^2 + \frac{\|g_k\|}{\|g_{k-1}\|} \|g_k\|\|g_{k-1}\|}{\|g_{k-1}\|^2 + |g_k^T d_{k-1}|} \leq \frac{2\|g_k\|^2}{\|g_{k-1}\|^2}$$

Hence we obtain

$$0 \leq \beta_k^{MRM} \leq \frac{2\|g_k\|^2}{\|g_{k-1}\|^2}$$

The following lemmas are very useful in the process of the studies on the conjugate gradient methods

Lemma 3.1.

Suppose that Assumptions A and B hold, let x_k be generated by Algorithm 2.1 where, d_k satisfies $g_k^T d_k < 0$ for all k , and α_k is obtained by (1.3), then,

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \qquad (3.1)$$

This theorem show that B_k^{MRM} has an advancement that the directions will approach to the steepest descent directions while the step length α_k is small.

Theorem 3.1.

Suppose that Assumptions A and B hold, $\{x_k\}$ generated by the Algorithm 2.1, where the step size α_k is determined by the exact line search (1.3). Then (3.1) holds for all $k \geq 0$.

Proof. Suppose that for all k, $g_k \neq 0$. If $k = 0$ then $g_0^T d_0 = g_0^T(-g_0) = -c\|g_0\|^2$. If at a point x_k, d_k is not a descent direction, then by the exact line search, we have $x_{k+1} = x_k$ which implies $g_{k+1} = g_k$ [21]. From (2.1), we have $B_k^{MRM} = 0$. This means that at those points the directions will turn out to be the steepest descent directions. Those points are denoted by $P_1 = \{x_k : B_k^{MRM} = 0\}$ and the other points are denoted by $P_2 = \{x_k : B_k^{MRM} \neq 0\}$.

For all the points in P_1, since the directions are the steepest descent directions, from Lemma 2.1, we have

$$\sum_{x_k \in P_1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \tag{3.2}$$

The same as the above proof, for the points in P_2, we also have

$$\sum_{x_k \in P_2}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \tag{3.3}$$

From (3.2) and (3.3) we have,

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \sum_{x_k \in P_1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} + \sum_{x_k \in P_2}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty$$

The proof is completed.

Theorem 3.2

Suppose that Assumptions A and B hold, the sequence $\{x_k\}$ is generated by Algorithm 2.1, if $\|s_k\| = \|\alpha_k d_k\| \to 0$ while $k \to \infty$, then

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \tag{3.4}$$

Proof. Let θ_k be the angle between $-g_k$ and d_k, where $\cos\theta_k = \dfrac{-g_k^T d_k}{\|g_k\|\|d_k\|}$, then by the exact line search, we have $g_k^T d_{k-1} = 0$, where the search direction defined by (1.4), the following relations hold true:

$$\|d_k\| = \sec\theta_k \cdot \|g_k\|, \quad \beta_{k+1}\|d_k\| = \tan\theta_{k+1} \cdot \|g_{k+1}\|$$

So we have

$$\tan\theta_{k+1} = \beta_{k+1}^{MRM} \sec\theta_k \frac{\|g_k\|}{\|g_{k+1}\|} = \sec\theta_k \frac{\|g_k\|}{\|g_{k+1}\|} \frac{g_{k+1}^T\left(g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k\right)}{\|g_k\|^2 + |g_{k+1}^T d_k|} \le \sec\theta_k \frac{\|g_{k+1}\|\|g_k\|\left\|g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k\right\|}{\|g_{k+1}\|\|g_k\|^2}$$

$$= \sec\theta_k \frac{\left\|g_{k+1} - g_k + g_k - \frac{\|g_{k+1}\|}{\|g_k\|} g_k\right\|}{\|g_k\|} \le \sec\theta_k \frac{\|g_{k+1} - g_k\| + \big|\|g_k\| - \|g_{k+1}\|\big|}{\|g_k\|}$$

$$\tan\theta_{k+1} \le \sec\theta_k \frac{2\|g_{k+1} - g_k\|}{\|g_k\|} \tag{3.5}$$

If (3.4) does not hold, then, for all k , there exists $c > 0$ such that

$$\|g_k\| \geq c \tag{3.6}$$

By $\|s_k\| \to 0$ and Lipschitz condition (2.2), there must exist an integer $M \geq 0$ for all $k \geq M$, such that

$$\|g_{k+1} - g_k\| \leq \frac{1}{4}c \tag{3.7}$$

Combining (3.5) and (3.7), we obtain

$$\tan\theta_{k+1} \leq \frac{1}{2}\sec\theta_k \tag{3.8}$$

We know that, for all $\theta \in [0, \frac{\pi}{2})$, the following inequality holds:

$$\sec\theta \leq 1 + \tan\theta \tag{3.9}$$

From (3.8) and (3.9) we get,

$$\tan\theta_{k+1} \leq \frac{1}{2}(1 + \tan\theta_k) \tag{3.10}$$

Utilizing (3.10) induces,

$$\tan\theta_{k+1} \leq \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + ... + \left(\frac{1}{2}\right)^{k+1-m} + \left(\frac{1}{2}\right)^{k+1-m} \tan\theta_m \leq 1 + \tan\theta_m$$

From this result, we note that the angle θ_k must be always less than some angle θ where, $\theta < \frac{\pi}{2}$, but by the Theorem 3.1, we have

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

This implies $\lim_{k\to\infty} \inf \|g_k\| = 0$, which contradicts (3.6). The proof is completed.

4. Numerical results and discussions

In this section, we present the computational performance of a MATLAB program on a set of 116 unconstrained optimization test problems. We selected 24 test functions considered in Andrei [1], each of them is tested in different variables. We performed a comparison with two CG methods Fletcher – Reeves (FR) and Polak-Ribiere-Polyak (PRP), we considered $\varepsilon = 10^{-6}$ and the gradient value as the stopping criteria as Hillstrom [12] suggested that $\|g_k\| \leq \varepsilon$ as the stopping criteria. For each of the test functions problem, we used four initial points, starting from a closer point to the solution and moving on to the one that is furthest from it. A list of problem functions and the initial points used are shown in table1, where the exact line search was used to compute the step size. The CPU processor used was Intel (R) Core^{TM} i3-M350 (2.27GHz), with RAM 4 GB. In some cases, the computation stopped due to the failure of the line search to find the positive step size, and thus it was considered a failure. Numerical results are compared relative on the CPU time and number of iteration. The performance results are shown in Figs.1 and 2 respectively, using a performance profile introduced by Dolan and More [5].

No	**Function**	**Dimension**	**Initial points**
1	Three Hump	2	(-10,-10),(10,10),(20,20),(40,40)
2	Six Hump	2	(-10,-10),(-8,8),(8,8),(10,10)
3	Booth	2	(10,10),(25,25),(50,50),(100,100)
4	Treccani	2	(5,5),(10,10),(20,20),(50,50)
5	Zettl	2	(5,5),(10,10),(20,20),(50,50)
6	Diagonal 4	2,4, 10,100,500,1000	(1,..,1),(3,…,3),(6,…,6),(12,…,12)
7	Perturbed Qua.	2,4, 10,100,500,1000	(1,…,1) ,(3,…,3) ,(5,…,5) , (10,…,10)
8	E-Himmelblau	10,100,500,1000,10000	(50,..,50),(70,..,70),(100,..,100) , (125,..,125)
9	E-Rosenbrock	10,100,500,1000,10000	(13,…,13),(25,…,25),(30,…,30),(50,…,50)
10	Shallow	10,100,500,1000,10000	(10,…10),(25,…,25),(50,…,50),(70,…,70)
11	E-Tridiagonal 1	10,100,500,1000,10000	(6,…,6),(12,…,12),(17,…,17),(20,…,20)
12	G-Tridiagonal1	2,4,10,100	(7,…,7),(10,…,10), (13,…,13),(21,…,21)
13	E-white-Holst	2,4,10,100,500,1000,10000	(3,…,3),(5,…,5) ,(7,…,7), (10,…,10)
14	G- Quartic	2,4,10,100,500,1000,10000	(1,…,1),(2,…,2),(5,…,5),(7,…,7)
15	E- Powell	4,20,100,500,1000	(2,…,2),(4,…,4),(6,…,6),(8,…,8)
16	E-Denschnb	2,4,10,100,500,1000,10000	(8,…,8),(13,…,13),(30,…,30),(50,50)
17	Hager	2,4,10,100	(7,…,7) ,(10,…,10), (15,…,15), (23,…,23)
18	E- Penalty	2,4,10,100	(80,..,80),(10,..,100),(111,..,111),(150,..,150)
19	Quadrtic QF2	2,4,10,100,500,1000	(5,…,5) ,(20,…,20) ,(50,…,50) ,(100,..,100)
20	E - QP2	2,4,10,100,500,1000	(10,..,10) ,(20,..,20), (30,..,30) ,(50,..,50)
21	E- Beale	2,4,10,100,500,1000,10000	(-1,…,-1) ,(3,…,3) ,(7,…,7) , (10,…,10)
22	Diagonal 2	2,4,10,100,500,1000	(1,…,1) ,(5,…,5) ,(10,…,10),(15,…,15)
23	Raydan1	2,4,10,100	(1,…,1) ,(3,…,3) ,(7,…,7) , (10,…,10)
24	Sum Squares	2,4,10,100,500,1000	(1,…,1) ,(3,…,3) ,(7,…,7) , (10,…,10)

According to the rules considered in [5], we know that the method whose performance profile plot on top right will be better than the rest of the other methods.

Figures 1-2 show that the performances of these methods are relative to the number of iteration and the CPU time. From figures 1-2, it is easy to see that MRM is the best among the two methods FR and PRP, the performance of PRP seems to be faster than MRM, but it can solve only 93% of the problems and FR solved only 70%, where MRM can solve all the test problems and reach 100%, the performance of MRM lies between FR and PRP. In other words, MRM method is competitive to the other two methods and its notable formula.

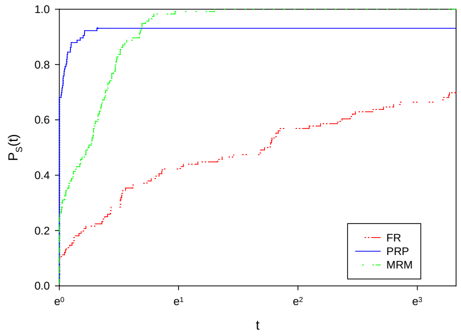

Fig.1. Performance profile relative to the number of iterations.

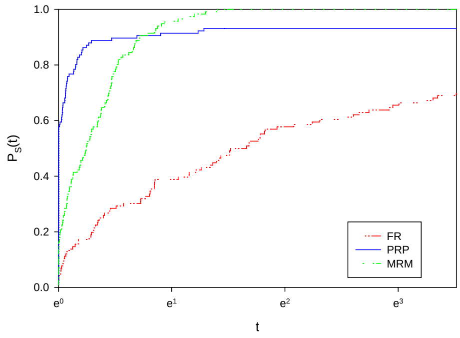

Fig.2. Performance profile relative to the CPU time.

5. Conclusion and further research

This paper gives a new conjugate gradient method for solving unconstrained optimization problems. Under the exact line search, this β_k possesses the global convergence condition. Numerical results show that our method is competitive to other two conjugate gradient methods, Fletcher Reeves (FR) and Polak Ribiere Polyak (PRP), and come out with best numerical results.

For further research, we should study the new method with the strong Wolfe-Powell line search.

Acknowledgements. The authors would like to thank the University of Malaysia Terengganu (Grant no FRGS Vot 59256) and Alasmrya University of Libya.

References

[1] N. Andrei, "An unconstrained optimization test functions collection," *Advanced Modeling and Optimization*, 10 (2008), 147-161.

[2] Y.-H. Dai and Y. Yuan, "A nonlinear conjugate gradient method with a strong global convergence property," *SIAM Journal on Optimization*, 10 (1999), 177-182. http://dx.doi.org/10.1137/s1052623497318992

[3] Y. H. Dai, "Convergence of nonlinear conjugate gradient methods," *Journal of Computational Mathematics*, 19 (2001), 539-548.

[4] Y. H. Dai and Y. Yuan, "Further studies on the Polak-Ribiere-Polyak method," Research report ICM-95-040, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences1995.

[5] E. D. Dolan and J. J. Mor, "Benchmarking optimization software with performance profiles," *Mathematical Programming,* 91 (2002), 201-213. http://dx.doi.org/10.1007/s101070100263

[6] R. Fletcher, *Practical Method of Optimization*, 2 ed. I. New York, 1987. http://dx.doi.org/10.1002/9781118723203

[7] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *The computer journal,* 7 (1964), 149-154. http://dx.doi.org/10.1093/comjnl/7.2.149

[8] J. C. Gilbert and J. Nocedal, "Global convergence properties of conjugate gradient methods for optimization," *SIAM J. Optim.,* 2 (1992), 21-42. http://dx.doi.org/10.1137/0802003

[9] L. Grippo and S. Lucidi, "A globally convergent version of the Polak-Ribière conjugate gradient method," *Math. Program.,* 78 (1997), 375–391. http://dx.doi.org/10.1007/bf02614362

[10] H. Hai, Y. Shengwei, and L. Suihua, "A new conjugate gradient method combined HS and DY formulas," M. a. I. Science, Ed., ed: Guangxi University, 2006.

[11] M. R. Hestenes and E. Stiefel, "Method of conjugate gradient for solving linear systems," *J. Res. Nat. Bur. Stand.,* 49 (1952), 409–436. http://dx.doi.org/10.6028/jres.049.044

[12] K. E. Hillstrom, "A simulation test approach to the evaluation of nonlinear optimization algorithms," *ACM Transactions on Mathematical Software* 3 (1977), 305-315. http://dx.doi.org/10.1145/355759.355760

[13] W. J. Z. L. Zhang, D.H. Li, "Global convergence of a modified Fletcher–Revves conjugate gradient method with Armijo-type line search," *Numer. Math.* , 104 (2006), 561–572. http://dx.doi.org/10.1007/s00211-006-0028-z

[14] Y. Liu and C. Storey, "Efficient generalized conjugate gradient algorithms, Part 1: Theory," *Journal of Optimization Theory and Applications,* 69 (1991), 129-137. http://dx.doi.org/10.1007/bf00940464

[15] E. Polak and G. Ribiere, "Note Sur la convergence de directions conjuge`es," *Rev. Francaise Informat Recherche Operationelle,* 3E (1969), 35–43.

[16] B. T. Polyak, "The conjugate gradient method in extreme problems," *USSR Computational Mathematics and Mathematical Physics,* 9 (1969), 94–112. http://dx.doi.org/10.1016/0041-5553(69)90035-4

[17] M. J. Powell, "Nonconvex minimization calculations and the conjugate gradient method," in *Numerical analysis*, ed: Springer, 122-141, 1984. http://dx.doi.org/10.1007/bfb0099521

[18] M. J. D. Powell, "Restart procedures for the conjugate gradient method," *Mathematical Programming* 12 (1977), 241–254. http://dx.doi.org/10.1007/bf01593790

[19] M. Rivaie, M. Mamat, L. W. June, and I. Mohd, "A new class of nonlinear conjugate gradient coefficients with global convergence properties," *Applied Mathematics and Computation*, 218 (2012), 11323-11332. http://dx.doi.org/10.1016/j.amc.2012.05.030

[20] D. Touati-Ahmed and C. Storey, "Globally convergent hybrid conjugate gradient methods," *J. Optim. Theory Appl.*, 64 (1990), 379–397.

[21] Z. X. Wei, S. W. Yao, and L. Y. Liu, "The convergence properties of some new conjugate gradient methods," *Applied Mathematics and Computation*, 183 (2006), 1341-1350. http://dx.doi.org/10.1016/j.amc.2006.05.150

[22] G. Zoutendijk, "Nonlinear programming, computational methods," in *Integer and nonlinear programming*, ed North-Holland, Amsterdam, 37-86, 1970.

Received: December 10, 2014; Published: March 7, 2015