Journal of Engineering and Applied Sciences 12 (18): 4632-4636, 2017
ISSN: 1816-949X

Another Descent Conjugate Gradient Method with Strong-Wolfe Line Search

[1]Nur Hamizah Abdul Ghani, [1]Mustafa Mamat, [1]Fatma Susilawati, [1]Norhaslinda Zull,
[1]Nur Syarafina Mohamed and [2]Mohd Rivaie
[1]Faculty of Informatics and Computing, University Sultan Zainal Abidin (UniSZA), Terengganu, Malaysia
[2]Department of Computer Science and Mathematics, University Teknologi Mara (UiTM),
Terengganu, Malaysia

Abstract: This study considers an iterative method which is a nonlinear conjugate gradient method for solving unconstrained optimization problems. We propose a new conjugate gradient method satisfying sufficient descent condition when strong Wolfe line search is used. We then compare the new method against other known conjugate gradient methods. Numerical results show that the proposed method is efficient relative to the number of iteration and computational time.

Key words: Conjugate gradient, strong-wolfe powell line search, sufficient descent, iteration, computational, Malaysia

INTRODUCTION

Due to its simplicity and low memory requirement, Conjugate Gradient (CG) method is recognized as one of best methods in optimization. The CG method is designed to find the optimal solution to the following unconstrained optimization problem:

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

where, f: $R^n \rightarrow R$ is continuously differentiable. The general idea of this method is to minimize any function and approach the optimal point by using the iterative procedure. It starts by an initial guess x_0, followed by improving the solution in following sequence x_k and lastly, ending the calculation by some stopping criteria. The update iterates is given by:

$$x_{k+1} = x_k + a_k d_k \tag{2}$$

Where:

a_k = The step size
d_k = The search direction of function at the current iterates point.

In the implementation of CG technique, the step size is determined by certain line search. The line search requires sufficient accuracy to ensure that the search direction is always in a descent direction. There are two types of line searches can be used either exact or inexact.

The best line search is the exact one. However, the application of exact line search is difficult to find and may fail in most cases of unconstrained optimization problems. Thus, the inexact line search is often considered where it is based on the Strong Wolfe-Powell (SWP) line search. In order to establish the convergence results of CG method the SWP line search requires step size satisfying the following condition:

$$f(x_k) - f(x_k + a_k d_k) \geq -\delta a_k g_k^T d_k \tag{3}$$

And:

$$\left| g(x_k + a_k d_k)^T d_k \right| \leq -\sigma g_k^T d_k \tag{4}$$

where, $0<\delta<\sigma<1$. Meanwhile, the d_k is a search direction defined by:

$$d_k = \begin{cases} -g_k & \text{if} \quad k = 0 \\ -g_k + \beta_k d_{k-1} & \text{if} \quad k \geq 1 \end{cases} \tag{6}$$

Where:

g_k = The gradient of f(x) at point x_k
β_k = A scalar known as the CG parameter

Different formulae for the parameter β_k result in different CG methods and their properties can be significantly different. Some of them are called the FR (Fletcher and Reeves, 1964), PRP (Polyak,1969; Polak and Ribiere,1969) HS (Hestenes and Steifel, 1952), VHS (Shengwei *et al*., 2007) and RMIL (Rivaie *et al*., 2011) which are, respectively given by:

Corresponding Author: Nur Hamizah Abdul Ghani, Faculty of Informatics and Computing, University Sultan Zainal Abidin (UniSZA), Terengganu, Malaysia

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

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

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_k^T (g_k - g_{k-1})} \quad (8)$$

$$\beta_k^{VHS} = \frac{g_k^T \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}\right)}{d_k^T (g_k - g_{k-1})} \quad (9)$$

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

The history of CG method was introduced by Hestenes and Steifel (1952). He presented β_k^{HS} as the CG parameter to solve a linear system of equation with a symmetric positive definite matrix. Then, Fletcher and Reeves (1964) extended the method of CG which is known as β_k^{FR} to solve general unconstrained optimization problems. From the founding of CG parameter until now there is a growing interest in the development of CG methods. Much research has been done in developing new formulae of CG parameter with good numerical performances and satisfying convergence analysis. The recent CG parameters can be found in those study (Abashar *et al*., 2016; Jusoh *et al*., 2013; Rivaie *et al*., 2012; Rivaie *et al*., 2012; Ghani *et al*., 2016; Hajar *et al*., 2016; Mohamed *et al*., 2016; Shapiee *et a l*., 2014; Zull *et al*., 2015; Shoid *et al*., 2016) and (Khadijah *et al*., 2016).

MATERIALS AND METHODS

The new conjugate method and its algorithm: Motivated and inspired by the ongoing research in solving unconstrained optimization problems this study proposes a new β_k, denoted as β_k^{HMRF} where HMRF represents Hamizah, Mustafa, Rivaie and Fatma. It derives from a modification of HS method (Hestenes and Steifel, 1952) and VHS method (Shengwei *et al*., 2007). The aim of this study is to improve the previous CG method with good numerical performance in practical computation as well as satisfying the convergence analysis. The new formula for the numerator has been proposed by adding a new scalar of m and the original formula for the denominator as HS and VHS formula have been retained. Powell (1986) has made an analysis of CG parameter and found that it should be restricted to a positive value to avoid it cycles infinitely without approaching a solution point. Thus the condition by Zull *et al*. (2015) helps the new parameter to always be in a positive value. The formula of HMRF method is constructed as follows:

$$\beta_k^{HMRF} = \begin{cases} \dfrac{\|g_k\|^2 - m|g_k^T g_{k-1}|}{d_{k-1}^T (g_k - g_{k-1})}, & \|g_k\|^2 > m|g_k^T g_{k-1}| \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

where: $m = \dfrac{\|g_k\|}{\|g_k - g_{k-1}\|}$ is a scalar.

Next, we present the algorithm of CG method as follows:

- Step 1: set k = 0 and select an initial point $x_0 \in \Re$
- Step 2: compute β_k
- Step 3: compute the d_k based on Eq. 5. If $\|g_k\| = 0$ then stop
- Step 4: compute the α_k based on Eq. 3 and 4
- Step 5: update new point based on iterative Eq. 2
- Step 6: convergent test and stopping criteria if $f(x_{k+1}) < f(x_k)$ and $\|g_{k+1}\| \le \varepsilon$ then stop. Otherwise, go to step 1 with k = k+1

Convergence analysis: In this study, the convergence analysis for the HMRF method with the SWP line search is studied. Every CG method must satisfy the sufficient descent condition to ensure that it is convergent. It is required to guarantee that the search direction is always in a descent direction. The sufficient descent condition is defined as follows:

$$g_k^T d_k \le -C\|g_k\|^2 \text{ for and some constant } C>0 \quad (12)$$

Before that, we need to simplify the HMRF method Eq. 10 for it to be easily implied in further proving. By applying condition $\|g_k\|^2 > m|g_k^T g_{k-1}|$, we have:

$$|\beta_k^{HMRF}| = \frac{\|g_k\|^2 - m|g_k^T g_{k-1}|}{d_{k-1}^T (g_k - g_{k-1})} \le \frac{\|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})} \quad (13)$$

Otherwise, we use β_k^{HMRF} Thus, we obtain:

$$0 \le |\beta_k^{HMRF}| \le \frac{\|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})} \quad (14)$$

Then, the following lemma is needed for the proof that HMRF method possesses the sufficient descent condition under the strong wolfe line search.

Theorem 1: If sequences $\{g_k\}$ and (d_k) are generated by CG algorithm and the stepsize α_k is computed by SWP line search with $\sigma<1/4$ then the sufficient descent condition Eq. 11 holds for some $c\in(0, 1)$.

Proof: The proof is by induction. From Eq. 5 with β_k^{HMRF} multiplying both sides by g^T_k yields:

$$g_k^T d_k = -\|g_k\|^2 + \beta_k^{HMRF} g_k^T d_{k-1} \tag{15}$$

For initial direction k = 0 and $d_0 = -g_0$ it is clear that $g_0^T d_0 \le -\|g_0\|^2 < 0$. Hence, holds for k = 0. Then, we need to show that also holds for $k\ge1$. Therefore, we divide the proof into two following cases. Case (i): If $g_0^T d_{k-1} \le 0$ then from Eq. 14, we obtain:

$$g_k^T d_k = -\|g_k\|^2 + \beta_k^{HMRF} g_k^T d_{k-1} < 0 \tag{16}$$

Case (ii): If, $g_k^T d_{k-1} \ge 0$ then divide Eq. 14 by $\|g_k\|^2$, we obtain:

$$\frac{g_k^T d_k}{\|g_k\|^2} = -1 + \beta_k^{HMRF} \frac{g_k^T d_{k-1}}{\|g_k\|^2} \tag{17}$$

From Eq. 13, we have:

$$\frac{g_k^T d_k}{\|g_k\|^2} \le -1 + \frac{\|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})} \frac{g_k^T d_{k-1}}{\|g_k\|^2} \tag{18}$$

$$\frac{g_k^T d_k}{\|g_k\|^2} \le -1 + \frac{g_k^T d_{k-1}}{d_{k-1}^T (g_k - g_{k-1})} \tag{19}$$

Note that the strong Wolfe condition gives:

$$d_{k-1}^T (g_k - g_{k-1}) \ge -(\sigma - 1) g_{k-1}^T d_{k-1} \tag{20}$$

Combine Eq. 18 and 19 with $\sigma<1/4$ we have:

$$\frac{g_k^T d_k}{\|g_k\|^2} \le -1 - \frac{\sigma g_{k-1}^T d_{k-1}}{(\sigma - 1) g_{k-1}^T d_{k-1}} \tag{21}$$

$$\frac{g_k^T d_k}{\|g_k\|^2} \le -1 + \frac{\sigma}{(1+\sigma)} < 0 \tag{22}$$

Let $c = 1-\sigma/(1+\sigma)$ we get:

$$g_k^T d_k \le -c\|g_k\|^2 \tag{23}$$

This shows that the result holds for $k\ge1$. Therefore, the sufficient descent condition holds. The proof is completed.

RESULTS AND DISCUSSION

In this study, we present the numerical result of this study. It will be compared based on iteration numbers and CPU times. The best convergence result yields from the less iteration number and less time needed to obtain the solution point. The 236 tests problems which we use are specified full in Table 1. They are including a set of 16 nonlinear unconstrained problems with different initial points and dimensions range from small scale to large scale. These test problems are taken from the list of test functions by Andrei (2008). In addition, we used four initial points with different ranges as suggested by Hillstrom (1977). The best selected initial point should be

Table 1: A list of test funcitions

Test function	Dimension/s	Initial points
Three hump	2	(3, 3), (10, 10), (20, 20), (44, 44)
Six hump	2	(1, 1), (5, 5), (10, 10), (15, 15)
Booth	2	(1,1), (5,5), (10, 10), (20, 20)
Treccani	2	(1, 1), (4, 4), (8, 8), (15, 15)
Zettl	2	(1, 1), (10, 10), (15, 15), (25, 25)
DIXMAANA	3, 900, 3000, 6000, 9000	(2, 2,…,2), (10, 10,…, 10), (15, 15,…, 15), (30, 30,…,30)
DIXMAANB	3, 900,3000, 6000,9000	(-5,-5,…,-5), (-2,-2,…,-2), (2,2,…, 2), (5,5,…,5)
Hager	2, 4, 10, 100	(2,2,…,2), (5,5,…,5), (10,10,…, 10), (20,20,…,20)
Generalized quartic	2, 500, 1000, 5000, 10000	(2,2,…,2), (10,10,…, 10), (20,20,…,20), (30,30,…,30)
Shallow	2, 500, 1000, 5000, 10000	(2, 2,…,2), (5,5,…,5) (25,25,…,25), (50,50,…,50)
Extended block	2, 500, 1000, 5000, 10000	(2, 2,…,2), (10,10,…,10), (15,15,…,15), diagonal (30,30,…,30)
Extended cliff	2, 500, 1000, 5000, 10000	(2, 2,…,2), (20,20,…,20), (50,50,…,50), (80,80,…,80)
Extended DENSCHNB	2, 500, 1000, 5000, 10000	(1, 1,…,1), (5,5,…,5) (10,10,…,10), (15,15,…,15)
Extended and Roth	2, 500, 1000, 5000, 10000	(2,2,…,2), Freudenstein (5,5,…,5), (30,30,…,30), (84,84,…,84)
Extended rosenbrock	2, 500, 1000, 5000, 10000	(2, 2,…,2), (5,5,…,5) ,(25,25,…,25), (50,50,…,50)
Extended white and	2, 500, 1000, 5000, 10000	(2, 2,…,2), (5,5,…,5) holst (25,25,…,25), (50,50,…,50)

based on the random number generator (Rivaie *et al.*, 2015). The stopping criterion is set to $f(x_{k+1})<f(x_k)$ and $\|g_k\|\le 10^{-6}$. All tests are implemented using Matlab R2011b on a PC with Intel (R), Core(TM), i7-4712MQ 2.30 GHz and 8GB RAM memory.

In this study, we apply the performance profile by Dolan and More (2002) to compare the numerical performance of CG methods. The performance profile provides the means to evaluate and compare the performance of the set solvers S on a test problem P. Assume that n_s solvers and n_p problems exist for each problem p and solver s. Let $t_{p,s}$ denote the performance measure (e.g., number of iterations and CPU time) required by solvers s to solve problem p. Then, the performance ratio is given by:

$$r_{p,s}=\frac{t_{p,s}}{\min\{t_{p,s}:s\in S\}} \tag{24}$$

Assume that a parameter $r_M\ge r_{p,s}$ is chosen for all p, s and $r_{p,s}=r_M$ if and only if solver s does not solve problem p. Define that for all $t\in\Re$:

$$p_s(t)=\frac{1}{n_p}\text{size}\{p\in P:r_{p,s}\le t\} \tag{25}$$

where, $p_s(t)$ is the probability for a solver $s\in S$ that a performance ratio $r_{p,s}$ was within a factor of the best possible ratio. The function p_s is the cumulative distribution function of the performance ratio. The performance profile for a solver is non-decreasing, piecewise and continuous from the right. The value $p_s(1)$ is the probability that the solver will win over the rest of solvers. Dolan and More (2002) for more details about performance profile.

Figure 1 and 2 show the performance profile of all of the tested methods-HMRF, VHS, RMIL, HS, PRP and FR method, relative to the number of iterations and CPU times, respectively. From the graph of performance profile, the right side gives the percentage of the test problems that are successfully solved by each of the methods. Meanwhile, the left side gives the percentage of the test problems with the fastest method. Based on the curve at the left side of both Fig. 1 and 2 the HMRF method shows the fastest method as its curve is above the other entire curves. Furthermore, the HMRF method can solve almost 99% test problems which is the highest as compared to the other methods. Meanwhile, the other methods which are VHS, FR, HS, RMIL and PRP, only solve 95, 88, 75, 69 and 64% of the test problems, respectively. In conclusion, the new method requires less iteration and less time needed to obtain the solution point based on numerical results of iteration numbers and CPU times. Therefore, we consider our new method; HMRF method is a comparable and efficient method in solving unconstrained optimization problems.

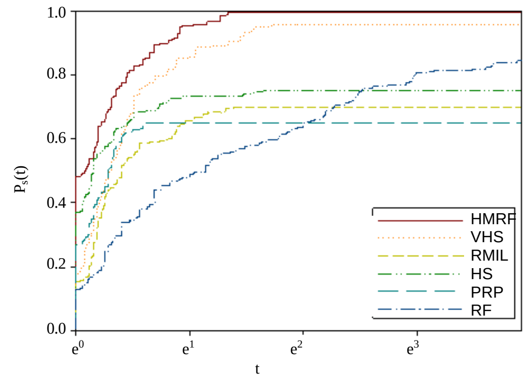

Fig. 1: Performance profile with respect to iteration number

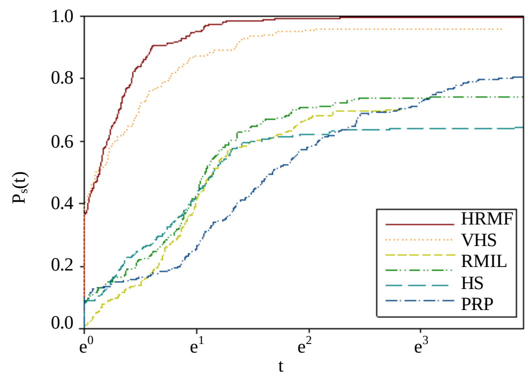

Fig. 2: Performance profile with respect to CPU times in seconds

CONCLUSION

This study provides the new conjugate gradient method for unconstrained optimization problems. We have theoretically shown that the HMRF method guarantee sufficient descent condition which is $g^T_k\, d_k\le -C\|g_k\|^2$ under the strong-Wolfe Powell line search. The computational results show that HMRF method is effective and performs better than the FR, PRP, HS, RMIL and VHS method.

ACKNOWLEDGEMENTS

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

REFERENCES

Abashar, A., M. Mamat, M. Rivaie and I. Mohd, 2016. Global convergence properties of a new class of conjugate gradient method for unconstrained optimization. Appl. Math. Sci., 8: 3307-3319.

Andrei, N., 2008. An unconstrained optimization test functions collection. JAMO., 10: 147-161.

Dolan, E.D. and J.J. More, 2002. Benchmarking optimization software with performance profiles. Math. Programming, 91: 201-213.

Fletcher, R. and C.M. Reeves, 1964. Function minimization by conjugate gradients. Comput. J., 7: 149-154.

Ghani, N.H.A., M. Rivaie and M. Mamat, 2016. A Modified Form of Conjugate Gradient Method for Unconstrained Optimization Problems. In: AIP Conference Proceedings, Chen, C.Y., S.N.I. Ibrahim, L.S. Lee, F.P. Lim and M.Z.A.M. Jaffar et al., (Eds.). AIP Publisher, Melville, New York, USA., pp: 020076-1-020076-8.

Hajar, N., M. Mamat, M. Rivaie and I. Jusoh, 2016. A New Type of Descent Conjugate Gradient Method with Exact Line Search. In: AIP Conference Proceedings, Chen, C.Y., S.N.I. Ibrahim, L.S. Lee, F.P. Lim and M.Z.A.M. Jaffar et al., (Eds.). AIP Publisher, Melville, New York, USA., pp: 020089-1-020089-8.

Hestenes, M.R. and E. Stiefel, 1952. Methods of conjugate gradients for solving linear systems. J. Res. Nat. Bureau Stand., 49: 409-435.

Hilstrom, K.E., 1977. A simulation test approach to the evaluation of nonlinear optimization algorithms. ACM Trans. Mathe. Software, 3: 305-315.

Jusoh, I., M. Mamat and M. Rivaie, 2013. A New Family of Conjugate Gradient Methods for Small-Scale Unconstrained Optimization. In: AIP Conference Proceedings, Ishak, A., I. Hashim, E.S. Ismail and R. Nazar (Eds.). AIP Publisher, Melville, New York, USA., pp: 1360-1365.

Khadijah, W., M. Rivaie, M. Mamat and I. Jusoh, 2016. A Spectral KRMI Conjugate Gradient Method Under the Strong-Wolfe Line Search. In: AIP Conference Proceedings, Chen, C.Y., S.N.I. Ibrahim, L.S. Lee, F.P. Lim and M.Z.A.M. Jaffar et al., (Eds.). AIP Publisher, Melville, New York, USA., pp: 020072-1-020072-8.

Mohamed, N.S., M. Mamat, F.S. Mohamad and M. Rivaie, 2016. A new coefficient of conjugate gradient methods for nonlinear unconstrained optimization. J. Sci. Technol. Eng., 78: 131-136.

Polak, E. and G. Ribiere, 1969. Note on the convergence of methods of conjugate directions. French J. Comput. Sci. Oper. Res., Ser. Rouge, 3: 35-43.

Polyak, B.T., 1969. The conjugate gradient method in extremal problems. USSR. Comput. Math. Phys., 9: 94-112.

Powell, M.J., 1986. Convergence properties of algorithms for nonlinear optimization. Siam Rev., 28: 487-500.

Rivaie, M., M. Mamat and A. Abashar, 2015. A new class of nonlinear conjugate gradient coefficients with exact and inexact line searches. Appl. Math. Comput., 268: 1152-1163.

Rivaie, M., M. Mamat, L.W. June and I. Mohd, 2012. A new class of nonlinear conjugate gradient coefficients with global convergence properties. Appl. Math. Comput., 218: 11323-11332.

Rivaie, M., M. Mustafa, M. Ismail and M. Fauzi, 2011. A comparative study of conjugate gradient coefficient for unconstrained optimization. Aus. J. Bas. Appl. Sci., 5: 947-951.

Shapiee, N., M. Rivaie, M. Mamat and I. Mohd, 2014. A New Modification of Hestenes-Stiefel Method with Descent Properties. In: AIP Conference Proceedings, Zin, W.Z.W., S.C. Dzul-Kifli, F.A. Razak and A. Ishak (Eds.). AIP Publisher, Melville, New York, USA., pp: 520-526.

Shengwei, Y., Z. Wei and H. Huang, 2007. A note about Wyl's conjugate gradient method and its applications. Appl. Math. Comput., 191: 381-388.

Shoid, S., M. Rivaie and M. Mamat, 2016. A Modification of Classical Conjugate Gradient Method Using Strong Wolfe Line Search. In: AIP Conference Proceedings, Chen, C.Y., S.N.I. Ibrahim, L.S. Lee, F.P. Lim and M. Z.A.M. Jaffar et al., (Eds.). AIP Publisher, Melville, New York, pp: 020071-1-020071-8.

Zull, N., M. Rivaie, M. Mamat, Z. Salleh and Z. Amani, 2015. Global convergence properties of a new spectral conjugate gradient by using strong wolfe line search. Appl. Math. Sci., 9: 3105-3117.