Journal of Engineering and Applied Sciences 12 (18): 4622-4626, 2017
ISSN: 1816-949X

An Alternative Approach for Conjugate Gradient Method

[1]Norhaslinda Zull, [1]Nurul 'Aini, [1]Mustafa Mamat, [1]Nur Hamizah Abdul Ghani,
[1]Nur Syarafina Mohamed, [2]Mohd Rivaie and [3]Sukono
[1]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia
[2]Department of Computer Science and Mathematics, Universiti Teknologi MARA, Terengganu, Malaysia
[3]Faculty of Mathematics and Natural Sciences, Universitas Padjajaran, Bandung, Indonesia

Abstract: Now a days, Conjugate Gradient (CG) has been a subject of extensive research by many researchers. In this study, a new CG method has been proposed. This new CG coefficient, β_k is compared to other four CG methods in their capabilities to solve fifteen selected test functions with different number of variables. The algorithm is established under exact line search and computed via Matlab R2012. This new method should possess the sufficient descent and global convergence properties. The results are interpreted into performance profile by using sigmaplot.

Key words: Optimization, conjugate gradient, exact line search, functions, variables, sigmaplot

INTRODUCTION

The general unconstrained optimization problem is defined by the following rule:

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

Where:
f: $R^n \rightarrow R$ = The continuously differentiable
R^n = An n-dimensional euclidean space

It is commonly solved by iterative method defined as follows:

$$x_{k+1} = x_k + a_k d_k, \ k = 1, 2, ..., \quad (2)$$

Where:
x_k = The current iteration point
$\alpha_k > 0$ = The stepsize

The most common approach to determine the α_k is the exact line search which is written as:

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

Recently, most researchers tend to use the inexact line search as some of them believe that the exact line search is much slower. This problem can be overcome by using new generation of faster computer processors. The basic search direction of Conjugate Gradient (CG) method, d_k is described as:

$$d_k = \begin{cases} -g_k & \text{if} \quad k = 0 \\ -g_k + \beta_k d_{k-1} & \text{if} \quad k \geq 1 \end{cases} \quad (4)$$

for which we have $g_k = \nabla f(x_k)$ at the point x_k while $\beta_k \in R$ is a scalar known as the CG coefficient. The conjugate gradient method can be categorized into classical CG, spectral CG and hybrid CG. The classical CG method is proposed by Hestenes Stiefel (HS) in 1952. Another CG method has been developed by Fletcher Reeves (FR) in 1964. Few years later, a new CG method has been introduced by Polak Ribiere (PR) in 1969. Then, Fletcher again established Conjugate Descent (CD) in 1987:

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

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

All these older versions of CG methods above can be found by Hestenes and Steifel (1952), Fletcher and Reeves (1964), Polak and Ribiere (1969), Fletcher (1987), respectively. Some examples of the CG methods by current researchers are listed:

Corresponding Author: Norhaslinda Zull, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia

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

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

These two methods of CG (Syarafina Mustafa Rivaie (SMR) and Syazni Rivaie Mustafa (SYRM)) can be referred by Mohameda *et al*. (2016) and Shoid *et al*. (2016), respectively. Another recent ones can be referred by Khadijah *et al*. (2016), Rivaie *et al*. (2015), Ghani *et al*. (2016), Hajar *et al*. (2016), Shapiee *et al*. (2016) and Zull *et al*. (2015). Besides, Khadijah *et al*. (2016), Hestenes and Steifel (1952), the other CG methods have been used in this study:

$$\beta_k^{RMIL} = \frac{g_k^T(g_k - g_{k-1})}{\|d_{k-1}\|^2} \quad (7)$$

$$\beta_k^{WYL} = \frac{g_k^T\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{\|g_{k-1}\|^2} \quad (8)$$

The methods of CG, Rivaie Mustafa Ismail Leong (RMIL) can be referred by Rivaie *et al*. (2012) while Wei *et al*. (2006) explains on Wei Yao Liu (WYL).

MATERIALS AND METHODS

New conjugate gradient coefficient: In this study, we propose a new β_k which is known as β^{LAMR}_k where LAMR denotes Linda, Aini, Mustafa and Rivaie as:

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

The algorithm is given as follows:

- Step 1: Initialization. Given x_0, set k = 0
- Step 2: Compute β_k based on Eq. 5-9
- Step 3: Compute d_k based on Eq. 4. If $\|g_k\|$, then stop
- Step 4: Compute α_k based on Eq. 3
- Step 5: Updating a new point based on Eq. 2
- Step 6: Convergence 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.

Convergence analysis: The convergent properties of β^{LAMR}_k will be studied in this study. A good algorithm should fulfill the sufficient descent and global convergence properties.

Sufficient descent condition: For the sufficient descent condition to hold:

$$g_k^T d_k \leq -c\|g_k\|^2 \quad \text{for } k \geq 0 \text{ and } c > 0 \quad (10)$$

The following theorem shows that the LAMR holds the sufficient descent condition.

Theorem 1: Suppose that assumption 1 holds true. Consider a CG method of β^{LAMR}_k Eq. 9 and the search direction given as Eq. 4 and the stepsize α_k is determined by using exact line search, then the sufficient descent condition (10) holds for all $k\geq 0$.

Proof: If k = 0, then it is clear that $g^T_0\ d_0$ hence, Eq. 10 holds true. The following shows that for $k\geq 1$, Eq. 10 will also hold true. Equation 4 by g_k:

$$\begin{aligned} g_k^T d_k &= g_k^T(-g_k + \beta_k d_{k-1}) \\ g_k^T d_k &= -\|g_k\|^2 + \beta_k g_k^T d_{k-1} \end{aligned} \quad (11)$$

For an exact line search, we know that $g^T_k\ d_{k-1} = 0$. Thus:

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

From there, we can see that $g^T_k\ d_k \leq -c\|g_k\|^2$ also holds true for $k\geq 1$. Thus, the proof is completed.

Global convergence properties: The CG algorithm with the new β_k is shown to be globally convergent under exact line search. Firstly, this new β_k will be simplified. From Eq. 9, we know that:

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

$$\beta_k^{LAMR} \leq \frac{\frac{\|d_{k-1}\|}{\|d_{k-1} - g_k\|}\|g_k\|^2}{\frac{\|d_{k-1}\|}{\|d_{k-1} - g_k\|}\|d_{k-1}\|^2}$$

Hence, we obtain:

$$\beta_k^{LAMR} \leq \frac{\|g_k\|^2}{\|d_{k-1}\|^2} \quad (12)$$

Assumption 1: f(x) is bounded below on the level set $l = \{x|f(x)\leq f(x_0)\}$ where x_0 is the initial point. In some neigbourhood, N of the level set l, f(x) is continuously differentiable and its gradient is Lipschitz continuous, then, there exists a constant L>0 such that $\|g(x)-g(y)\|\leq L\|x-y\|$ for $\forall x, y\in N$.

The level set L is convex. Positive constants c_1 and c_2 exists satisfying $c_1\|z\|^2\leq z^T F(x)z\leq c_2\|z\|^2$ for all $z\in R^n$ and $x\in L$ where F(x) is the Hessian matrix of f.

Lemma 1: Suppose that Assumption 1 holds true, consider any d_k of the form Eq. 4 and α_k obtained under exact minimization rule for all $k\geq 0$. Then, the following Zoutendijk condition holds where:

$$\sum_{k=1} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty \tag{13}$$

Proof: The proof can be referred by Zoutendijk (1970). By using Lemma 1, the following theorem is established to prove the global convergence of the new method with exact line search.

Theorem 2: Suppose that Assumption 1 and Theorem 1 hold true. Consider the CG method in the form of Eq. 2 and 4 with Eq. 9 as the β_k in addition to the stepsize, α_k obtained by exact line search. Then:

$$\lim_{k\to\infty}\|g_k\| = 0 \text{ or } \sum_{k\geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty \tag{14}$$

Proof: To prove this theorem, the method of contradiction is used. Firstly, suppose that there exists a positive constant $\delta>0$ such that:

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

Squaring both sides of the equation, we obtain:

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

Since, $g^T_k d_{k-1} = 0$ for exact line search, we have:

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

Dividing both sides by $\|g_k\|^4$ gives us:

$$\frac{\|d_k\|^2}{\|g_k\|^4} = \frac{\|g_k\|^2}{\|g_k\|^4} + \frac{\beta_k^2 \|d_{k-1}\|^2}{\|g_k\|^4}$$

Substituting Eq. 12 into the equation, we get:

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq \frac{1}{\|g_k\|^2} + \left(\frac{\|g_k\|^4}{\|d_{k-1}\|^4}\right)\left(\frac{\|d_{k-1}\|^2}{\|g_k\|^4}\right)$$

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq \frac{1}{\|g_k\|^2} + \frac{1}{\|d_{k-1}\|^2}$$

From Eq. 10 and 15, we can define that $g^T_k d_k = -\|g_k\|^2$. Therefore:

$$\sum_{k=1}^{\infty} \frac{\|d_k\|^2}{\|g_k\|^4} \leq \sum_{k=1}^{\infty} \frac{1}{\|g_k\|^2} + \sum_{k=1}^{\infty} \frac{1}{\|d_{k-1}\|^2}$$

Since, $\|g_k\|\geq\delta$, this implies that $\Sigma 1/\|g_k\|^2\approx 0$. Hence:

$$\sum_{k=1}^{\infty} \frac{\|d_k\|^2}{\|g_k\|^4} \leq \sum_{k=1}^{\infty} \frac{1}{\|d_{k-1}\|^2} \tag{16}$$

From Eq. 10 and 15, we know that $g^T_k d_k = -\|g_k\|^2$. Since, $g_k\to\infty$, consequently, $d_k\to\infty$. Therefore, we can deduce that:

$$\sum_{k=0}^{\infty} \frac{\|d_k\|^2}{\|g_k\|^4} \approx 0$$

This implies that which contradicts the Zoutendjik condition. The proof is thus completed:

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} \geq \infty$$

RESULTS AND DISCUSSION

In this study, the efficiency of the new method is tested by a list of test problems as in Table 1. Besides, the robustness of the methods can be tested using these test problems. These test problems are taken from Andrei (2011). A comparison was made with the other CG methods, CD, HS, RMIL and WYL (5-8) under an exact line search.

Four different initial points have been used starting from a point which is closer to the solution point to the point far away from the solution point. The selections of the initial points are randomized as the best chosen solution points should be based on the random number generator. For additional reading (Hilstrom, 1977).

We consider $\varepsilon = 10^{-6}$ and terminate the calculations when the stopping criteria, $\|g_k\|<10^{-6}$ has been fulfilled. All the problems stated below are computed by Matlab R2012 subroutine programming. The iteration number and CPU time are evaluated in order to compare the performance of these CG methods.

Table 1: A list of test problem

Test problems	Variables	Initial points
BIGGSB1	2	(10,10), (20,20), (30,30), (40,40)
Booth	2	(9,9), (18,18), (27,27), (36,36)
Leon	2	(3,3), (5,5), (8,8), (11,11)
Six hump	2	(2,2), (5,5), (9,9), (12,12)
Trecanni	2	(5,5), (10,10), (15,15), (20,20)
Extended Maratos	2,10	(7,…,7), (16,…,16), (27,…,27), (32,…,32)
Raydan	2,10	(1,…,1), (2,…,2), (3,…,3), (4,…,4)
Diagonal 4	2,10,100, 1000,10000	(11,…,11), (22,…,22), (33,…,33), (44,…,44)
Extended DENSCHNB	2,10,100, 1000,10000	(3,…,3), (7,…,7), (13,…,13), (19,…,19)
Extended Freudenstein and roth	2,10,100, 1000,10000	(2,…,2), (13,…,13), (22,…,22), (33,…,33)
Extended Himmelblau	2,10,100, 1000,10000	(10,…,10), (20,…,20), (30,…,30), (40,…,40)
Extended Rosenbrock	2,10,100, 1000,10000	(2,…,2), (11,…,11), (21,…,21), (32,…,32)
Extended Tridiagonal 1	2,10,100, 1000,10000	(28,…,28), (32,…,32), (48,…,48), (56,…,56)
FLETCHCR	2,10,100, 1000,10000	(3,…,3), (6,…,6), (9,…,9), (14,…,14)
Shallow	2,10,100, 1000,10000	(2,…,2), (6,…,6), (9,…,9), (11,…,11)

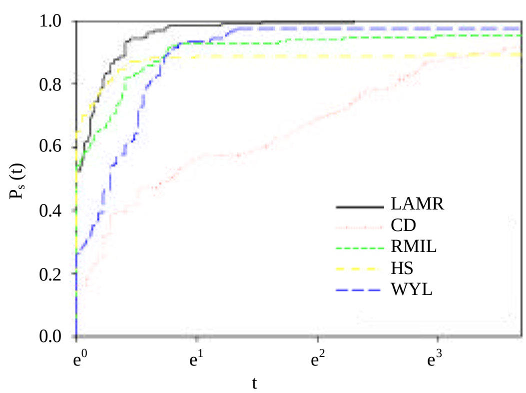

Fig. 1: Performance profile based on the iteration number

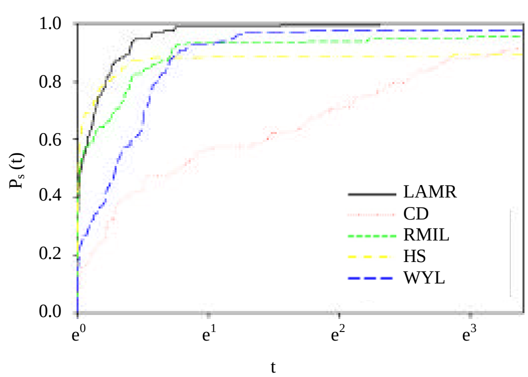

Fig. 2: Performance profile based on the CPU time

The results based on the iteration numbers and CPU time are described as shown in Fig. 1 and 2, respectively via the performance profile introduced by Dolan and More (2002). By this performance profile, the method performance and the best method can be defined clearly. By referring to both figures, the LAMR method has proved to be robust as it solves all the test problems. This new method is also comparable to HS and RMIL methods in term of number iterations and CPU time. Additionally, it is shown that LAMR performs better compared to CD and WYL.

CONCLUSION

Based on the comparative study, it has been proven that the LAMR method fulfills the sufficient descent condition and the global convergence properties. Eventhough the LAMR method is not the best method, overall its results are comparable to that of the best method which is the HS method in terms of iteration number and CPU time. Hence, the LAMR method is encouraging.

ACKNOWLEDGEMENTS

Researhers would like to thank the editors and the referees for their suggestions and comments. Researcher are also thankful to The Ministry of Higher Education of Malaysia (MOHE) for the sponsorship of this research via MyPhd Scheme (Ref. No.: KPT (B) 901206065386).

REFERENCES

Andrei, N., 2011. Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization. Bull. Malaysian Math. Sci. Soc., 34: 319-330.

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.

Fletcher, R., 1987. Practical Method of Optimization, Unconstrained Optimization. 2nd Edn., John Wiley & Sons, New York, USA., ISBN:9780471915478, Pages: 436.

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. Steifel, 1952. Method of conjugate gradient for solving linear equations. J. Res. Nat. Bur. Stand., 49: 409-436.

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

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.

Mohameda, N.S., M. Mamata, F. Susilawati and M.R.A. Mohamada, 2016. A new coefficient of conjugate gradient methods for nonlinear unconstrained optimization. J. Technol., 78: 131-136.

Polak, E. and G. Ribiere, 1969. Note on convergence of conjugate direction methods. French J. Oper. Res., 3: 35-43.

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.

Shapiee, N., M. Rivaie and M. Mamat, 2016. A New Classical Conjugate Gradient Coefficient 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: 020082-1-020082-8.

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.

Wei, Z., S. Yao and L. Liu, 2006. The convergence properties of some new conjugate gradient methods. Appl. Math. Comput., 183: 1341-1350.

Zoutendijk, G., 1970. Nonlinear Programming Computational Methods. In: Integer and Nonlinear Programming, Abadie, J. (Ed.). North-Holland, Amsterdam, pp: 37-86.

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