Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 11 (8): 1461-1464, 2011
ISSN 1812-5654 / DOI: 10.3923/jas.2011.1461.1464

Modified Hestenes-Steifel Method for Unconstrained Optimization

[1]M. Rivaie, [1]M. Fauzi, [2]M. Mamat and [2]I. Mohd
[1]Department of Computer Sciences and Mathematics, Universiti Teknologi MARA Terengganu, Kampus Kuala Terengganu, Malaysia
[2]Department of Mathematics, Faculty of Science and Technology, Terengganu, Malaysia

Abstract: Conjugate gradient methods play an important role in unconstrained optimization. Numerous studies and modifications have been devoted recently to improve this method. In this study a new conjugate gradient coefficient (β_k) is proposed by modifying the already proven Hestenes-Steifel formula. In this new β_k, a new formula for the denominator is introduced and the numerator of the original Hestenes-Steifel formula is retained. Numerical results based on the number of iterations by using exact line search have shown that this new formula of β_k performs far better than the original Hestenes-Steifel, and outperforms the other conjugate gradient methods. The numerical results also suggest that this method possesses global convergence properties.

Key words: Conjugate gradient method, conjugate gradient coefficient, exact line search, global convergence

INTRODUCTION

The conjugate gradient methods (CG) are useful in finding the minimum value of function for unconstrained optimization. In general, the method has the following form:

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

where, $f : R^n \to R$ is continuously differentiable. The CG method is an iterative method of the form :

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

where, x_k is the current iterate point, is a stepsize and is the search direction.

The search direction, d_k is defined by:

$$d_k = \begin{cases} -g_k & \text{if} \quad k=0 \\ -g_k + \beta_k d_{k-1} & \text{if} \quad k \geq 1 \end{cases} \tag{3}$$

where, g_k is the gradient of f(x) at the point x_k. $\beta_k \in R$ is known as conjugate gradient coefficient. Some well known formulas are given as follows:

$$\beta_k^{FR} = \frac{g_k^T g_k}{\|g_{k-1}\|^2} \tag{4}$$

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

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

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

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

where, g_k and g_{k-1} are the gradients of f(x) at the point x_k and x_k-1 respectively. The above corresponding methods are known as Fletcher and Reeves (1964), Hestenes and Steifel (1952), Liu and Storey (1991) and Dai and Yuan (2000). We represent norm of vectors as $\|.\|$. For f(x) that is strictly convex quadratic function, all these method are equivalent, but for general non-quadratic functions, their behavior is quite different (Dai and Yuan, 1998; Yuan and Sun, 1999).

The most studied properties of CG methods are its global convergence properties. Zoutendijk (1970) has proven the global convergence of FR method with exact line search. Unfortunately, Powell (1977) has proven a major drawback of FR method. Powell (1984) also showed that PR method can cycle infinitely without reaching the

Corresponding Author: Mohd Rivaie, Department of Computer Sciences and Mathematics, Universiti Teknologi MARA, Terengganu, Kampus Kuala Terengganu, Malaysia

studyminimizer, hence, its convergence is not global. Other researchers such as Al-Baali (1985), Touati-Ahmed and Storey (1990) and Gilbert and Nocedal (1992) have further analyzed the global convergence of algorithms related to the FR method with strong Wolfe condition. Powell (1986) once again proves that FR is a superior method compared to others. Since then, the global convergence of PR, LS, and HS has not been established yet. The main reason is that it cannot guarantee the descent objective function values at each iterate (Hager and Zhang, 2005). For further reading and recent finding of CG methods refer to Sun and Zhang (2001), Birgin and Matrtinez (2001), Dai and Yuan (2002), Yuan and Wei (2009), Andrei (2009) and Shi and Guo (2009).

A key factor of global convergence is selecting the stepsize, α_k. The most common search is to use the exact line search, which is finding the exact value of α_k. Though many researches opt to use the inexact line search, in this study however, our new proposed β_k is solved using the exact line search.

NEW HESTENES STEIFEL METHOD

Here, the new method is proposed based on the original HS method. The new method is named as Modified Hestenes-Steifel method (MHS). A new formula for the denominator has been proposed, and the original formula for the numerator as the Hestenes-Steifel formula has been retained. Hence:

$$\beta_k^{MHS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(d_{k-1} - g_k)} \quad (9)$$

The complete algorithm of MHS is shown as follows:

Step 1: Given X_0, set $k = 0$
Step 2: Compute β_k^{MHS} based on Eq. 9
Step 3: Compute the search direction

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

If $\|gk\| = 0$, then stop

Step 4: Solve $\alpha_k = \min_{\alpha > 0} f(x_k + \alpha d_k)$
Step 5: Updating new pointing using $x_{k+1} = x_k + \alpha_k d_k$
Step 6: 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: The convergence property that is presented here section is based on Dai *et al.* (1999). In this study, they have proven the global convergence of FR and PR method. We also assumed that the property of HS is the same as PR. To prove this convergence, it is assumed that every d_k satisfies the descent condition:

$$g_k^T d_k < 0 \quad (10)$$

For all $k \geq 1$.

The basic assumptions on the objective function are defined as follows:

Assumption:

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

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \text{for all } x, y \in N \quad (11)$$

To prove global convergence for the FR method, the strong Wolfe line search has been used, for which the $\propto_k$ should satisfy the Wolfe condition (Wolfe, 1969).

Theorem: Consider that Assumption 1 is true. Any CG method of the form Eq. 2 and 3, with d_k satisfying Eq. 10 with strong Wolfe line search, then either:

$$\lim_{k \to \infty} \|g_k\| = 0 \quad (12)$$

or:

$$\sum_{k=1}^{\infty} \frac{(g_k)^4}{\|d_k\|^2} < +\infty \quad (13)$$

The following direct corollary is based on Theorem 2:

Corollary: Consider that Assumption 1 is true for any CG method of the form Eq. 2 and 3, with satisfying Eq.10 with strong Wolfe line search. If:

$$\sum_{k=1}^{\infty} \frac{(g_k)^t}{\|d_k\|^2} = +\infty \quad (14)$$

for any $t \in [0,4]$, the method converges if (12) is true.

Proof: To prove Corollary 3, assume that Theorem 2 is contradicted. Hence, if Eq.12 is not true, then Theorem 3 will yield:

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty \quad (15)$$

Since ‖gk‖ is bounded away from zero and $t \in [0,4]$, it is clear that the Corollary 3 is true. As a conclusion, if a conjugate gradient method fails to converge, then the search direction will converge to infinity.

NUMERICAL RESULTS

Test problems from Andrei (2008) have been used to test and analyze the efficiency of MHS compared to FR, PR, HS, and DY. Stopping criteria is set to $\|g_k\| \le \varepsilon$ where $\varepsilon = 10^{-6}$. As suggested by Hilstrom (1977), for each of the test problem, four initial starting points are used. In doing so, it leads us to test the global convergence properties of our method. Numerical results will be compared based on the number of iterations (Table 2). In this case CPU time is not considered as a comparison, since the average time taken for a single iteration did not show any significant difference. The percentage performance of MHS as compared to the other method is shown in Table 3. The word "Fail" in Table 2 means that the run was stopped due to the line search procedure failed to find the positive stepsize. The iteration will also stop if the iteration number is more than one thousand. All the problems mentioned below are solved by Maple 12 subroutine program using the exact line search (Table 1). Processor used is Intel Core2 Duo T5750 with 2.0 GHz, 667 MHz FSB, 2MBL2 cache.

Table 1: List of problems function

No.	Function
1	Rosenbrock's function for two variables
2	Rosenbrock's function for four variables
3	Cube function for two variables
4	Wood function for four variables
5	Strait function for two variables
6	Six Hump Camel Back function for two variables
7	Three Hump Camel Back function for two variables

Table 2: Performance comparison of different CG methods

No.	Initial point	FR	PR	HS	DY	MHS
1	(13,13)	563	23	23	619	12
	(50,50)	>1000	29	29	>1000	23
	(100,100)	>1000	30	30	>1000	24
	(200,200)	>1000	41	41	>1000	29
2	(13,13,13,13)	585	23	23	587	13
	(50,50,50,50)	>1000	29	29	>1000	23
	(100,100,100,100)	>1000	30	30	>1000	24
	(200,200,200,200)	>1000	41	41	>1000	32
3	(3,-6)	145	33	33	145	39
	(10,-10)	240	31	31	240	33
	(-10,-10)	235	31	31	234	33
	(-15,15)	328	43	43	491	21
4	(2,2,2,2)	26	Fail	Fail	Fail	170
	(5,5,5,5)	30	Fail	Fail	Fail	131
	(10,10,10,10)	33	Fail	Fail	Fail	204
	(50,50,50,50)	>1000	Fail	Fail	Fail	259
5	(10,10)	9	5	5	9	18
	(50,50)	37	7	7	37	14
	(100,100)	117	8	8	117	28
	(200,200)	284	9	9	284	14
6	(10,-10)	105	6	6	105	6
	(50,-50)	8	6	6	8	6
	(100,-100)	8	6	6	8	6
	(200,-200)	8	6	6	8	6
7	(10,-10)	6	4	4	6	6
	(50,-50)	5	4	4	5	4
	(100,-100)	3	3	3	3	3
	(200,-200)	5	3	3	5	3

Table 3: Percentage performance of MHS

Comparison	Successful	Equivalent	Unsuccessful
FR(%)	75.00	7.14	17.86
PR(%)	46.43	25.00	28.57
HS(%)	46.43	25.00	28.57
DY(%)	89.29	7.14	3.57

DISCUSSION

From Table 2, it is shown that for all the given problems, MHS successfully reaches the solution point. It is also proven that MHS outperformed FR in test problems 1, 2, 3, 5, 6, and 7. MHS also outperformed DY and CD in almost all the problems. For PR, HS and DY, the MHS is superior or equal to the problem 1, 2, 4, 6 and 7.

From Table 3, it is shown that MHS is superior compared to the other methods. The highest percentage of successful comparison is with DY with a combined rate of successful and equivalent rate of 96.43%. Combined rate for FR exceeds 80%. Though, the successful rate comparisons for HS and PR are quite low at 46.43%, their combined rate of successful rate and equivalent rate exceeds 70%. Above all, all the comparisons showed that the combined rate of successful rate and equivalent rate exceeds 70%. Therefore, we consider the MHS is superior, compared to the other methods.

For problem 4, though MHS is inferior when compared to FR, their success is undeniable when compared to the other methods. Note that the other methods fail in finding the positive stepsize by using the exact line search, but we have yet to run and compared all these methods using the inexact line search. Hence, MHS provides good alternative in finding the solution using the exact line search.

CONCLUSION

In this study, a new β_k based on the already proven Hestenes-Steifel method has been proposed. Our numerical results have shown that, our new method is superior, compared to the other standard conjugate gradient methods. Our numerical result also suggested that this new method converge globally. Further numerical testing should be done for large scale problems so that this method may become a new conjugate gradient family.

REFERENCES

Al-Baali, M., 1985. Descent property and global convergence of Fletcher-Reeves method with inexact line search. IMA J. Numer. Anal., 5: 121-124.

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

Andrei, N., 2009. Accelerated conjugate gradient algorithm with finite difference Hessian/vector product approximation for unconstrained optimization. J.Comput.Applied Mathe., 230: 570-582.

Birgin, E.G. and J.M. Martinez, 2001. A spectral conjugate gradient method for unconstrained optimization. J. Applied Mathe. Optimizat., 43: 117-128.

Dai, Y. and Y. Yuan, 1998. Nonlinear Conjugate Gradient Method. Shanghai Scientific and Technical Publishers, Beijing.

Dai, Y.H., J.Y. Han, G.H. Liu, D.F. Sun, X. Yin and Y. Yuan, 1999. Convergence properties of nonlinear conjugate gradient method. SIAM J. Optimization, 10: 348-358.

Dai, Y. and Y. Yuan, 2000. A nonlinear conjugate gradient with a strong global convergence properties. SIAM J. Optimizat., 10: 177-182.

Dai, Y.H. and Y. Yuan, 2002. A note on the nonlinear conjugate gradient method. J. Comput. Applied Mathe., 18: 575-582.

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

Gilbert, J.C. and J. Nocedal, 1992. Global convergence properties of conjugate gradient methods for optimization. SIAM J. Optimizat., 2: 21-42.

Hager, W.W and H.C. Zhang, 2005. A new conjugate gradient method with guaranteed descent and efficient line search. SIAM J. Optimizat., 16: 170-192.

Hestenes, M.R. and E. Steifel, 1952. Method of conjugate gradient for solving linear equations. J. Res. Nat. Bureau Standards, 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.

Liu, Y. and C. Storey, 1991. Efficient generalized conjugate gradient algorithms part 1: Theory.J.Comput. Applied Mathe., 69: 129-137.

Powell, M.J.D., 1977. Restart procedures for the conjugate gradient method. Mathe. Program., 12: 241-254.

Powell, M.J.D., 1984. Non-Convex Minimization Calculations and the Conjugate Gradient Method. Springer, Berlin.

Powell, M.J.D., 1986. Convergence properties of algorithm for nonlinear optimization. SIAM Rev., 28: 487-500.

Shi, Z.J. and J. Guo, 2009. A new family of conjugate gradient methods. J. Comput. Applied Mathematics, 224: 444-457.

Sun, J. and J. Zhang, 2001. Global convergence of conjugate gradient methods without line search. Ann. Opr. Rch., 103: 161-173.

Touati-Ahmed, D. and C. Storey, 1990. Efficient hybrid conjugate gradient techniques. J. Optimizat. Theory Applied, 64: 379-397.

Wolfe, P., 1969. Convergence conditions for ascent method. SIAM Rev., 11: 226-235.

Yuan, Y. and W. Sun, 1999. Theory and Methods of Optimization. Science Press of China, Beijing.

Yuan, G. and Z. Wei, 2009. New line search methods for unconstrained optimization. J.Korean Stat. Soc., 38: 29-39.

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