Int. J. Contemp. Math. Sciences, Vol. 5, 2010, no. 29, 1429 - 1437

A New Conjugate Gradient Coefficient for Unconstrained Optimization

Mustafa Mamat [1], Mohd Rivaie [2], Ismail Mohd [3] and Muhammad Fauzi [4]

[1,3] Department of Mathematics, Faculty of Science and Technology
Universiti Malaysia Terengganu (UMT)
Malaysia

[2,4] Department of Computer Sciences and Mathematics
Universiti Teknologi MARA (UiTM) Terengganu
Campus Kuala Terengganu
Malaysia

[1] mus@umt.edu.my, [2] rivaie75@yahoo.com, [3] ismail@umt.edu.my,
[4] fauziembong@yahoo.com

Abstract

Conjugate gradient method holds an important role in solving unconstrained optimizations, especially for large scale problems. Numerous studies and modifications have been done to improve this method. In this paper, we propose a fundamentally different conjugate gradient method in which the well known coefficient β_k is computed using the eigenvalues generated by using exact Hessian matrix of $f(x)$. This relatively easy formula for β_k has made conjugate gradient method very simple, but still possesses global convergence properties. Numerical results have shown that this formula performs better than the original conjugate gradient methods.

Mathematics Subject Classification: 65K10

Keywords: conjugate gradient methods, conjugate gradient coefficient, eigenvalues, Hessian matrix, global convergence properties

1. Introduction

The conjugate gradient method (CG) plays an important role in solving the unconstrained optimization problem. In general, the method has the following form

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

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

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

where x_k is the current iterate point, $\alpha_k > 0$ is a step size and d_k is the search direction. Basically d_k is defined by

$$d_k = \begin{cases} -g_k & \text{if} \quad k = 0 \\ -g_k + \beta_k d_{k-1} & \text{if} \quad k \geq 1 \end{cases} \tag{1.3}$$

where g_k is the gradient of $f(x)$ at the point x_k. $\beta_k \in R$ is known as conjugate gradient coefficient and different β_k will yield different CG methods. Some well known formulas are given as follows:

$$\beta_k^{FR} = \frac{g_k^T g_k}{\|g_{k-1}\|^2} \tag{1.4}$$

$$\beta_k^{PR} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} \tag{1.5}$$

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} \tag{1.6}$$

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

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

where g_{k-1} are the gradients of $f(x)$ at the point x_{k-1}. In this paper, FR denotes Fletcher and Reeves [9], PR denotes Polak and Ribiere [14], HS denotes Hestenes and Steifel [12], LS denotes Liu and Storey [13] and lastly DY denotes Dai and Yuan [6]. We denotes norm of vectors as $\|.\|$. It also shows that for $f(x)$ that is strictly convex quadratic function, all these method are equivalent, but for general non quadratic, their behavior is quite different. (see Dai and Yuan [5] and Yuan and Sun, [21])

The most studied properties of CG are its global convergence properties. Zoutendijk [22] proved the global convergence of FR method. Al-Baali [1], Touati-Ahmed and Storey [18], Gilbert and Nocedal [10] has further analyzed the global convergence of algorithms related to the FR method with strong Wolfe condition. Powell [15] also proved that FR is a superior method compared to others. 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 iteration (see Hager and Zhang [11]). For further reading and recent finding of CG methods refer to Sun and Zhang [17], Birgin and Matrtinez [4], Dai and Yuan [7], Yuan and Wei [20], Andrei [3] and Shi and Gao [16].

A basic key factor of global convergence is selecting the step size α_k. The most common search is to do the exact line search

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k) \tag{1.9}$$

In this paper, we propose a new CG method with a simple formula for β_k. Our new algorithm is presented in Section 2. The global convergence of the new method is established using the exact line search for non convex function and the proof is presented in Section 3. Some interesting numerical result is presented in Section 4 by comparing our new method with other CG method. Lastly, our discussion and conclusion are presented in Section 5 and Section 6 respectively.

2. New algorithm

In this section we propose our new algorithm known as Eigen Conjugate Gradient (ECG). For twice differentiable two variable functions we can compute the eigen values of the exact Hessian matrix from the current point. The reciprocal of these eigenvalues product is then set as a new β_k that is,

$$\beta_k = \frac{1}{E_1 + E_2} \tag{2.1}$$

where E_1 and E_2 are eigenvalues.

For functions with more than two variables $(n \geq 2)$ these new β_k could be further elaborated as;

$$\beta_k = \frac{1}{E_1 + E_2 + \ldots + E_{n-1} + E_n} \tag{2.2}$$

where E_1, E_2, E_{n-1}, E_n are eigenvalues.

New point is then computed by using (1.2) and (1.3). The complete algorithm is shown as follows:

Algorithm 1: The ECG method.

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

Step 2: Computing conjugate gradient coefficient
Compute β_k base from (2.1) or (2.2)

Step 3: Computing search direction
$d_k = -g_k + \beta d_{k-1}$. If $g_k = 0$, then stop.

Step 4: Computing step size
Solve $\alpha_k = \min_{\alpha > 0} f(x_k + \alpha d_k)$,

Step 5: Updating new point
$x_{k+1} = x_k + \alpha_k d_k$

Step 6: Convergent 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. Convergent analysis

The convergence properties which we present in this section follow from Dai, et. al. [8]. In the paper, they have proven the global convergence of FR and PR methods. Here, we only showed the result of convergence for the general CG methods.

For this proof, we assume that every d_k satisfies the descent condition

$$g_k^T d_k < 0 \tag{3.1}$$

for all $k \geq 1$.

We make the following basic assumptions on the objective function.

Assumption 3.1

(i) $f(x)$ is bounded below on the level set $\ell = \{x | f(x) \leq f(x_1)\}$ where x_1 is the starting point.

(ii) 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

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \text{for all } x, y \in N \tag{3.2}$$

The step size α_k in (1.2) is computed by carrying out a line search. In this case we consider the Wolfe line search which consists of finding a positive step size ($\alpha_k > 0$) such that

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k g_k^T d_k , \tag{3.3}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma \alpha_k g_k^T d_k , \tag{3.4}$$

where $0 < \rho < \sigma < 1$.

In order to prove global convergence for the FR method, we used the strong Wolfe line search which requires α_k to satisfy (3.3) and

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq -\sigma (g_k)^T d_k \tag{3.5}$$

The following important result was obtained by Zoutendijk [22] and Wolfe [19]

Lemma 3.2

Consider that Assumption 3.1 is true. Consider any iteration method of the form (1.2) to (1.3), where d_k satisfies (3.1) and α_k is obtained by the Wolfe line search. Then

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty \tag{3.6}$$

The following theorem is a general and positive result for CG methods with the strong Wolfe line search.

Theorem 3.3
Consider that Assumption 3.1 is true. Any CG method of the form (1.2) to (1.3), with d_k satisfying (3.1) and with strong Wolfe line search (3.3) and (3.4), then either:

$$\lim_{k\to\infty}\|g_k\| = 0 \tag{3.7}$$

or

$$\sum_{k=1}^{\infty}\frac{(g_k)^4}{\|d_k\|^2} < +\infty \tag{3.8}$$

Proof:
To proof Theorem 3.3, we refer to (1.3) for $k \geq 1$

$$d_k + g_k = \beta_k d_{k-1} \tag{3.9}$$

Squaring both sides of (3.9), yields,

$$\|d_k\|^2 = -\|g_k\|^2 - 2g_k^T d_k + \beta_k^2\|d_{k-1}\|^2 \tag{3.10}$$

It follows from this relation and (3.10) that

$$\|d_k\|^2 \geq \beta_k^2\|d_{k-1}\|^2 - \|g_k\|^2 \tag{3.11}$$

Definition (1.3) implies the following relation

$$g_k^T d_k - \beta_k g_k^T d_{k-1} = -\|g_k\|^2 \tag{3.12}$$

Which, with the line search (3.5), shows that

$$\left|g_k^T d_k\right| + \sigma|\beta_k\|g_{k-1}d_{k-1}| \geq \|g_k\|^2 \tag{3.13}$$

The above inequality and the Cauchy-Schwartz inequality yield

$$\left(g_k^T d_k\right)^2 + \beta_k^2\left(g_{k-1}^T d_{k-1}\right)^2 \geq c_1\|g_k\|^4 \tag{3.14}$$

where $c_1 = \left(1+\sigma^2\right)^{-1}$ is a positive constant.
Therefore, it follows from (3.11) and (3.12) that

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} + \frac{(g_{k-1}^T d_{k-1})^2}{\|d_{k-1}\|^2} \geq \frac{1}{\|d_k\|^2}\left[c_1\|g_k\|^4 - \frac{(g_{k-1}^T d_{k-1})^2}{\|d_{k-1}\|^2}\|g_k\|^2\right] \tag{3.15}$$

If (3.7) is not true, relations (3.15) and (3.6) will imply the following inequality

$$\frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} + \frac{\left(g_{k-1}^T d_{k-1}\right)^2}{\|d_{k-1}\|^2} \geq \frac{c_1}{2}\frac{\|g_k\|^4}{\|d_k\|^2} \tag{3.16}$$

holds for all sufficiently large *k*. Now inequality (3.8) follows from (3.16) and (3.8).
The following is a direct corollary of Theorem 3.3.

Corollary 3.4
Consider that Assumption 3.1 is true. Any CG method of the form (1.2) to (1.3), with d_k satisfying (3.1) and with strong Wolfe line search (3.3) and (3.4). If

$$\sum_{k=1}^{\infty}\frac{(g_k)^t}{\|d_k\|^2} = +\infty \tag{3.17}$$

for any $t \in [0,4]$, the method converges in the sense that (3.7) is true.

Proof:

To proof Corrollary 3.4, we use contradiction.

If (3.7) is not true, it follows from Theorem 3.3 that,

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty \tag{3.18}$$

Because $\|g_k\|$ is bounded away from zero and $t \in [0,4]$, it is easy to see that (3.18) contradicts (3.17). This shows that the Corollary is true.
Hence, if a conjugate gradient method fails to converge, one can easily see that the length of the search direction will converge to infinity.

4. Numerical result

In this section we use the test problems considered in Andrei [2] to analyze the improvement of ECG compared to FR, PR and HS. We considered $\varepsilon = 10^{-6}$ and stopping criteria is set to $\|g_k\| < 10^{-6}$. For each of the test problem, four initial starting points are used starting from a point that is further away from the solution point to the point close to it. All the problems are solved by *Maple 11* subroutine program by using the exact line search. Numerical result is presented in Table1.

Table 1: Performance comparison of different CG methods based on number of iterations.

No.	Function	Initial point	ECG	FR	PR	HS
1	Rosenbrock (*n*=2)	(13,13)	174	563	23	23
		(50,50)	156	2994	29	29
		(100,100)	170	12231	30	30
		(200,200)	367	>12231	41	41
2	Rosenbrock (*n*=4)	(13,13,13,13)	468	585	23	23
		(50,50,50,50)	366	4642	29	29
		(100,100,100,100)	314	>1000	30	30
		(200,200,200,200)	885	>1000	41	41
3	Cube (*n*=2)	(3,-6)	266	145	33	33
		(10,-10)	3367	240	31	31
		(-10,-10)	598	235	31	31
		(-15,15)	3344	328	43	43
4	Shalow (*n*=2)	(10,10)	70	163	Fail	Fail
		(50,50)	104	866	Fail	Fail
		(100,100)	124	5100	Fail	Fail
		(200,200)	180	>1000	Fail	Fail
5	Shalow (*n*=4)	(10,10,10,10)	41	172	Fail	Fail
		(50,50,50,50)	584	888	Fail	Fail
		(100,100,100,100)	775	>1000	Fail	Fail
		(200,200,200,200)	449	>1000	Fail	Fail
6	Wood (*n*=4)	(2,2,2,2)	1498	26	Fail	Fail
		(5,5,5,5)	1116	30	Fail	Fail
		(10,10,10,10)	1807	33	Fail	Fail
		(50,50,50,50)	2077	4364	Fail	Fail

5. Discussion

It is shown that for all the given problems, ECG successfully reaches the solution point. It is also proven that ECG outperform FR in Problems 1,2,4, and 5. For Problems 6, it is shown that ECG outperforms FR for points that are further away from the solution point.

Even though it seems obvious that PR and HS are better than ECG, but their convergence is proved to be local (Powell [15]). Moreover they tend to fail in solving the given problem using the exact line search as shown for Problems 4,5, and 6. It is also shown that for Problem 3 and 6, the higher number of iteration may be due to the Hessian matrix that is not positive definite hence the eigenvalues is not guaranteed to be positive and leads to insufficient descent condition. Rate of convergence also prove to be superior when compared to FR method. Above all, this simple formula of β_k is relatively easy to use compared to the other CG methods.

6. Conclusion

Numerous studies of CG methods lead to new variety of CG methods. However the new proposed method seems to be more difficult than the previous one suggested by the early researcher, though they proved to perform better. In this paper we have proposed a new and simple β_k . We have shown numerically that this method proves to be successful and reliable for functions up to four variables. Our numerical result also suggested that this method converge globally. Our further interest is to investigate its behavior for functions with more variable $(n > 4)$. So that it may become a new conjugate gradient family for solving unconstrained optimization problem.

References

[1] Al-Baali, M. (1985). Descent property and global convergence of Fletcher-Reeves method with inexact line search. *IMA. J.Numer.Anal.*, 5, 121-124.

[2] Andrei, N. (2008). An unconstrained optimization test functions colection. *Advanced Modelling and Optimization*, 10(1), 147-161.

[3] Andrei, N. (2009). Accelerated conjugate gradient algorithm with finite difference Hessian/vector product approximation for unconstrained optimization. *J,Comput.Appl.Math.*, 230, 570-582

[4] Birgin, E.G. and Martinez. J.M. (2001). A spectral conjugate gradient method for unconstrained optimization. *J.Appl.Maths.Optim*, 43, 117-128.

[5] Dai, Y. and Yuan, Y. (1998). *Nonlinear conjugate gradient method*. Shanghai Scientific and Technical Publishers, Beijing.

[6] Dai, Y. and Yuan, Y. (2000). A nonlinear conjugate gradient with a strong global convergence properties. *SIAM J. Optim.*, 10, 177-182.

[7] Dai, Y.H. and Yuan, Y. (2002). A note on the nonlinear conjugate gradient method. *J,Comput.Appl.Math.*, 18(6), 575-582.

[8] Dai, Y.H., Han, J.Y., Liu, G.H., Sun, D.F., Yin, X. and Yuan, Y. (1999). Convergence properties of nonlinear conjugate gradient method. *SIAM J. Optim.*, 10, 348-358.

[9] Fletcher, R. and Reeves, C. (1964). Function minimization bu conjugate gradients. *Comput.J.*, 7, 149-154.

[10] Gilbert, J.C. and Nocedal, J. (1992). Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.*, 2(1), 21-42.

[11] Hager, W.W and Zhang, H.C. (2005). A new conjugate gradient method with guaranteed descent and efficient line search. *SIAM J. Optim.*, 16, 170-192.

[12] Hestenes, M. R. and Steifel, E. (1952). Method of conjugate gradient for solving linear equations. *J,Res.Nat.Bur.Stand.*, 49, 409-436.

[13] Liu, Y. and Storey, C. (1992). Efficient generalized conjugate gradient algorithms part 1:theory. *J,Comput.Appl.Math.*, 69, 129-137.

[14] Polak, E. and , Ribiere, G. (1969). Note sur la convergence de directions conjugees. *Rev.Francaise Informat Recherche Operationelle,3E Annee*(16), 35-43.

[15] Powell, M.J.D. (1986). Convergence properties of algorithm for nonlinear optimization. *SIAM Review.*, 28(4), 487-500.

[16] Shi, Z.J. and Gao, J. (2009). A new family of conjugate gradient methods. *J,Comput.Appl.Math.*, 224, 444-457.

[17] Sun, J. and Zhang, J. (2001). Global convergence of conjugate gradient methods without line search. *Annals. Opr. Rch,* 103, 161-173.

[18] Touati-Ahmed, D. and Storey, C . (1990). Efficient hybrid conjugate gradient techniques, *J.Optim.Theory Appl.*, 64, 379-397.

[19] Wolfe, P. (1969). Convergence conditions for ascent method. *SIAM Rev.*, 11, 226-235.

[20] Yuan, G. and Wei, Z. (2009). New line search methods for unconstrained optimization. *J.Korean Stat.Soc.*, 38, 29-39.

[21] Yuan, Y. and Sun, W. (1999). *Theory and methods of optimization*. Science Press Of China, Beijing.

[22] Zoutendijk, G. (1970). Nonlinear programming computational methods. *In:Abadie J.(Ed.) Integer and nonlinear programming*, 37-86.

Received: November, 2009