International Journal of Mathematical Archive-4(5), 2013, 280-287
Available online through www.ijma.info ISSN 2229 – 5046

A MODIFIED OF CONJUGATE GRADIENT ALGORITHM FOR UNCONSTRAINED OPTIMIZATION

Mardeen Sh. Taher*
Dept. of Mathematics, Faculty of Science, University of Duhok, Kurdistan Region-Iraq

(Received on: 16-04-13; Revised & Accepted on: 08-05-13)

ABSTRACT

The conjugate gradient method is a very useful technique for solving minimization problems and has wide applications in many fields. In this paper we develop a new class of conjugate gradient methods for unconstrained optimization; conjugate gradient methods are widely used for large scale unconstrained optimization problems. In addition to ,The performance of a modified Wolf line search rules related to CG-method type method with the results from standard Wolf line search rules are compared.

Keywords*: Unconstrained optimizations, line search, Wolfe conditions, conjugate gradient method modified secant condition.*

1. INTRODUCTION

In this study we consider the unconstrained minimization problem

$$\min f\left(x\right) \tag{1.1}$$

and the conjugate gradient method of the form:

$$x_{k+1} = x_k + \alpha_k d_k \tag{1.2}$$

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

where $x_k \in R^n$ is the current iterative, d_k is a decent direction of $f(x)$ at x_k , $g_k = \nabla f(x_k), \alpha_k$ is step size obtained by a line search and β_k is a scalar. The scalar so chosen that the method (1.2) ,(1.3) reduces to the linear conjugate gradient method when f is a strictly convex quadratic and when $\alpha_k$is the exact one – dimensional minimizer. Various conjugate gradient methods have been proposed ,and they are mainly differ in the choice of the parameter β_k . Some well-known formulas for β_k are called the Fletcher-Reeves (FR), Polak-Ribiere-polyak (PRP),and Hestenes-Stiefel (HS) ([6], [11], [12], and [7] respectively), are given below:

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

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

$$\beta_k^{HS} = g_k^T(g_k - g_{k-1})/d_k^T\,(g_k - g_{k-1}) \tag{1.6}$$

where $\|.\|$ denotes the Euclidean norm. The Conjugate gradient method is a very efficient line search method for solving large unconstrained problems, due to its lower storage and simple computation. The conjugate gradient method is still the best choice for solving (1.1).

Corresponding author: *Mardeen Sh. Taher**
Dept. of Mathematics, Faculty of Science, University of Duhok, Kurdistan Region-Iraq

In practical computations, it is generally believed that the conjugate gradient method is preferred to the relatively exact line searches. As a result, in the already-existing convergence analyses and implementations of the conjugate gradient method, the strong Wolfe conditions, namely,

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta\alpha_k g_k^T d_k \tag{1.7}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \le -\sigma g_k^T d_k \tag{1.8}$$

where $0 < \delta < \sigma < 1$, are often imposed on the line search. However, recent studies show that one can analyze the conjugate gradient method under several practical line searches other than the strong Wolfe line search, and good numerical results can be obtained. For example, the nonlinear conjugate gradient method in [5] converges globally provided that the step size satisfies the standard Wolfe conditions, namely, (1.7) and

$$g(x_k + \alpha_k d_k)^T d_k \le \sigma g_k^T d_k \tag{1.9}$$

2. DERIATION OF THE MODIFIED COJUGATE GRIDIENT AlGORITHM

Considerv$_x = x_{k+1} - x_k$,$y_x = g_{k+1} - g_k$, when $g_k = \nabla f(x_k)$ and the unconstrained nonlinear problem is minmizef(x) , $x \in R^n$. Conjugate directionswhich introduce in (1.3) have the property:

$$d_{k+1}^T G_k d_k = 0 \quad \text{for } k \ge 1 \tag{2.1}$$

Where G_k is the Hessian of $f(x_k)$,from (2.1) , we have

$$\begin{aligned} d_{k+1}^T G_k d_k &= \frac{1}{\alpha_k} d_{k+1}^T G_k (x_{k+1} - x_k) \\ &= \frac{1}{\alpha_k} d_{k+1}^T G_k (x_{k+1} - x_k) \\ &= \frac{1}{\alpha_k} d_{k+1}^T (g_{k+1} - xg_k) = \frac{1}{\alpha_k} d_{k+1}^T y_k \end{aligned} \tag{2.2}$$

From Quasi-Newton, the search direction can be calculated in the form

$$d_{k+1} = -H_{k+1} g_{k+1} \tag{2.3}$$

By $H_{k+1} y_k = v_k$ and equation (2.3) , we get

$$d_{k+1}^T y_k = -(H_{k+1} g_{k+1})^T y_k = -g_{k+1}^T (H_{k+1} y_k) = -g_{k+1}^T v_k \tag{2.4}$$

Perry replaced the conjugacy condition $d_{k+1}^T y_k = 0$ by condition $d_{k+1}^T y_k = -g_{k+1}^T v_k$.

Recently Dai and Liao proposed the condition $d_{k+1}^T y_k = -\tau g_{k+1}^T v_k$ where $\tau \ge 0$ is scalar.

In new modified taken $\tau = \frac{1}{\alpha_k}$ where $\alpha_k = \frac{d_k^T y_k}{\|d_k\|^2} > 0$.

So, the conjugacy condition $d_{k+1}^T y_k = -\tau g_{k+1}^T v_k$ become

$$d_{k+1}^T y_k = -\frac{1}{\alpha_k} g_{k+1}^T v_k \tag{2.5}$$

Now, multiply the conjugate gradient direction in (1.3) by y_x , we get

$$d_{k+1}^T y_k = -g_{k+1}^T y_k + \beta_k d_k^T y_k \tag{2.6}$$

Therefore,

$$\beta_k = \frac{d_{k+1}^T y_k + g_{k+1}^T y_k}{d_k^T y_k} \tag{2.7}$$

Now, taking the conjugate condition in (2.5) and putting in (2.7), we obtain

$$\beta_k = \frac{-\frac{1}{\alpha_k} g_{k+1}^T v_k + g_{k+1}^T y_k}{d_k^T y_k}$$

Implies

$$\beta_k = \frac{g_{k+1}^T(\frac{1}{-\alpha_k} v_k + y_k)}{d_k^T y_k} \tag{2.8}$$

By this way, we get a modified formula of the conjugate gradient direction, and it is possesses the property of a decent direction and it is proved in theorem1.

ALGORITHM OF THE MODIFIED COJUGATE GRIDIENT

Step 0: choose an initial point$x_0 \in R^n$, $\varepsilon \in (0,1)$, and set $d_0 = -g_0 = \nabla f(x_0)$, $k = 0$

Step 1: If $\|g_k\| \leq \varepsilon$ then stop; otherwise go to the next step.

Step 2: Compute step size α_k by some line search rules.

Step 3: Let$x_{k+1} = x_k + \alpha_k d_k$ if $\|g_{k+1}\| \leq \varepsilon$ then stop.

Step 4: Calculate the search direction

$$d_{k+1} = -g_{k+1} + \beta_k d_k \tag{2.9}$$

where $\beta_k = \frac{g_{k+1}^T(-\frac{1}{\alpha_k} v_k + y_k)}{d_k^T y_k}$

Step5: Set k=k+1 and go to step 2

Theorem 1: Assume that the sequence $\{x_k\}$ is generated by the algorithm (1), then the modified of CG-method in (2.9) is satisfied the sufficient descent condition in to two cases: exact and inexact line search.

Proof: we will get this theorem by mathematical induction:

It is clear when k= 0, then $d_0 = -g_0$ implies $d_0^T g_0 \leq -\|g_0\|^2$

suppose that the current search direction is descent direction at the iteration (k) , $k > 0$ that is mean this inequity $d_k^T g_k \leq c\|g_k\|^2$ is satisfy .

Now, we prove the current search direction is descent direction at the iteration (k+1), we have

$d_{k+1} = -g_{k+1} + \beta_k d_k$, where $\beta_k = \frac{g_{k+1}^T(\frac{-1}{\alpha_k} v_k + y_k)}{d_k^T y_k}$, then

$$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T(\frac{-1}{\alpha_k} v_k + y_k)}{d_k^T y_k} d_k \tag{2.10}$$

multiplying (2.10) by g_{k+1} , and get

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T(\frac{-1}{\alpha_k} v_k + y_k)}{d_k^T y_k} d_k^T g_{k+1} \tag{2.11}$$

It is easy to show that the produce a descent search direction if the step-length α_k is chosen by an exact line search which requires $d_k^T g_{k+1} = 0$.

Now, if the step-length α_k is chosen by an inexact line search which requires $d_k^T g_{k+1} \neq 0$. Dividing both sides of (2.11) by$\|g_{k+1}\|^2$, and obtain

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} + 1 = \frac{g_{k+1}^T(\frac{-1}{\alpha_k} v_k + y_k)}{d_k^T y_k} \frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2} \tag{2.12}$$

From Wolf condition $g(x_k + \alpha_k d_k)^T d_k \leq \sigma g_k^T d_k$, we get the following inequality

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} + 1 \leq (\frac{\sigma g_k^T d_k + g_{k+1}^T y_k}{d_k^T y_k}) \frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2} \tag{2.13}$$

We should note that the Wolf condition guarantees $d_k^T y_k > 0$ and that

$$d_k^T y_k = d_k^T g_{k+1} - d_k^T g_k > d_k^T g_{k+1} \tag{2.14}$$

By using (2.14) in (2.12), we get

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} + 1 \leq (\frac{-g_{k+1}^T d_k + g_{k+1}^T y_k}{d_k^T g_{k+1}}) \frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2} \tag{2.15}$$

Implies, $\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} + 1 \leq (\frac{-g_{k+1}^T d_k + g_{k+1}^T y_k}{\|g_{k+1}\|^2})$ (2.16)

From descent direction $d_k = -g_k$, (2.16) become

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} + 1 \leq (\frac{g_{k+1}^T g_k + \|g_{k+1}\|^2 - g_{k+1}^T g_k}{\|g_{k+1}\|^2}) \tag{2.17}$$

implies $d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 \leq 0$ (2.18)

Therefore descent direction is satisfied when inexact line search.

3. LINE SEARCH METHODS WITH MODIFIED WOLF CONDITION

The line search method proceed as follows ,each iteration computes a search direction d_k, the iterations given by (1.2), most line search algorithms require d_k to be a descent direction, i.e.,$d_k^T g_k < 0$, A popular inexact line search condition stipulates that α_k should first of all give sufficient decrease in the objective function f, as measured by the inequality in (1.7). The sufficient decrease condition (1.7) is not enough to ensure convergence since as we have just seen, this condition is satisfied for all small enough α_k. To rule out unacceptably small steps the second requirement called a curvature condition is introduced in (1.8).

3.1 MODIFIED WOLF CONDITIONS:

The two modified Wolf conditions become as follows:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k \mu \tag{3.1}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k \tag{3.2}$$

where $\delta = c_1 + 3.5. c_1(1 - c_1)$, $c_1 \in (0,1)$

$\mu = \mu + \mu(1 - \mu)$ and $\mu = g_k^T d_k$

3.2 ALGORITHM LINE SEARCH TECHNIQUE WITH MODIFIED WOLF CONDITIONS:

Step 0: set an initial iterate $x_k$by educated guess, set k =0.

Step 1: Until x_k has converged.

Step 2: Calculate a search direction d_k from x_k, ensuring that the decent direction ($d_k^T g_k < 0$) is satisfied, d_k is defined in (2.9)

Step 3: Calculate a suitable size step α_k so that the (3.1) and (3.2) are satisfies.

Step 4: set $x_{k+1} = x_k + \alpha_k d_k$.

Step 5: set k=k+1.

Theorem 2.Consider any iteration of the form $x_{k+1} = x_k + \alpha_k d_k$, where d_k is a descent direction in (2.9), and α_k satisfies the Wolfe conditions (3.1), (3,2). Suppose $f(x)$ is bounded from below in R^n and $f(x)$ is continuously differentiable in an open set D containing the sublevel set

$SL = \{x \in R^n : f(x) \le f(x_1)\}$

where x_1 is the starting point of the iteration, assume that $\nabla f(x) = g(x)$ is Lipschitz-continuous in D, that is mean, there exist constant L>0 such that

$$\|g(x) - g(y)\| \le L\|x - y\|, \quad x, y \in D \tag{3.3}$$

Then

$$\sum_{k\ge1} \cos\theta_k \|g_k\| < \infty \tag{3.4}$$

where θ_k is angle between the search direction d_k and the steepest descent direction $-g_k$ and defined by

$$\cos\theta_k = \frac{-g_k^T d_k}{\|d_k\|\|g_k\|} \tag{3.5}$$

Proof: Subtracting $d_k^T g_k$ from (5.2) and taking into account that $x_{k+1} = x_k + \alpha_k d_k$ we get

$$(g_{k+1} - g_k)^T d_k \le (\sigma - 1)\mu \tag{3.6}$$

while the Lipschitz continuity implies that

$$(g_{k+1} - g_k)^T d_k \le \|g_{k+1} - g_k\|\|d_k\| \le \alpha_k L\|d_k\|^2 \tag{3.7}$$

Combining these two relations we obtain

$$\alpha_k \ge \left(\frac{\sigma-1}{L}\right)\frac{{g_k}^T d_k}{\|d_k\|^2} \tag{3.8}$$

By substituting (5.8) in (5.1) we get

$$f_{k+1} \le f_k + \sigma\left(\frac{\sigma-1}{L}\right)\frac{({g_k}^T d_k)^2}{\|d_k\|^2} \tag{3.9}$$

Now we use (5.5) to write as

$$f_{k+1} \le f_k + \sigma\left(\frac{\sigma-1}{L}\right)\cos^2\theta\|g_k\|^2 \tag{3.10}$$

By summing this expression and recalling that is bounded from below, then we get

$$\sum_{k\ge1} \cos\theta_k \|g_k\| < \infty$$

4. NUMRICAL RESULTS

This section is devoted to test the implementation of the new methods. We compare the modified method with standard the cubic line search ,the comparative tests involve well-known nonlinear problems (standard test function) with different dimension $4 \le n \le 3000$, all programs are written in FORTRAN95 language and for all cases the stopping condition is $\|g_{k+1}\|_\infty \le 10^{-5}$ The results are given in table (1) and table (2) is specifically quote the number of functions NOF and the number of iteration NOI experimental results in table (1) confirm that the new CG method is superior to standard CG method with respect to the NOI and NOF .And the table(2) illustrate effect of modified of Wolf conditions on standard CG method compared with the standard Wolf conditions.

Table: 1
Comparative performance of two algorithms (standard CG method and new CG new method)

Test problem	N	CG(H\S) NOI(NOF)	NEW CG NOI(NOF)
Powell	4	38(68)	28(74)
	100	40(122)	33(86)
	500	41(124)	36(102)
	1000	41(124)	40(119)
Wood	4	30(68)	28(64)
	100	30(68)	28(64)
	500	30(68)	29(66)
	1000	30(68)	29(66)
Rosen	4	29(74)	29(74)
	100	30(76)	30(76)
	500	30(76)	30(76)
	1000	30(76)	30(76)
Cubic	4	16(44)	14(39)
	100	16(44)	15(43)
	500	16(44)	15(43)
	1000	16(44)	15(43)
Milele	4	31(96)	31(94)
	100	34(110)	36(114)
	500	40(138)	37(126)
	1000	74(172)	43(148)
Generalized PSC1	4	37(86)	26(72)
	100	37(86)	28(74)
	500	36(86)	28(74)
	1000	36(86)	28(74
Extended PSC1	4	30(76)	24(68)
	100	30(76)	26(72)
	500	32(76)	26(72)
	1000	32(76)	28(74)
Full Hessian FH1	4	31(98)	28(64)
	100	32(98)	26(62)
	500	32(98)	26(62)
	1000	32(98)	26(62)
Extended Maratos	4	16(44)	15(44)
	100	16(44)	15(45)
	500	16(44)	15(45)
	1000	16(44)	15(45)
FLETCHCR function (CUTE):	4	16(44)	14(38)
	100	16(46)	14(36)
	500	16(46)	14(36)
	1000	16(46)	14(36)
FLETCBV3 function (CUTE):	4	16(46)	15(45)
	100	16(44)	14(44)
	500	16(46)	14(45)
	1000	16(45)	14(45)

Table: 2

Comparative performance of two algorithms (standard CG method under (standard and modified Wolf conditions)

Test problem	**N**	**CG(H\S)(** standard Wolf conditions) **NOI(NOF)**	**CG (H\S)(**modified Wolf conditions) **NOI(NOF)**
Powell	4	38(68)	32(83)
	100	40(122)	35(98)
	500	41(124)	35(98)
	1000	41(124)	35(98)
	3000	41(124)	35(98)
Wood	4	30(68)	25(59)
	100	30(68)	27(63)
	500	30(68)	27(63)
	1000	30(68)	27(63)
	3000	30(68)	27(63)
Rosen	4	29(74)	28(72)
	100	30(76)	29(74)
	500	30(76)	30(76)
	1000	30(76)	30(76)
	3000	30(76)	30(76)
Cubic	4	16(44)	12(35)
	100	16(44)	12(35)
	500	16(44)	12(35)
	1000	16(44)	14(41)
	3000	16(44)	14(41)
Generalized PSC1	4	37(86)	30(74)
	100	37(86)	30(78)
	500	36(86)	29(74)
	1000	36(86)	29(74
	3000	36(86)	29(74)
Extended PSC1	4	30(76)	30(75)
	100	30(76)	30(75)
	500	32(76)	30(75)
	1000	32(76)	30(74)
	3000	32(76)	30(74)
Full Hessian FH1	4	31(98)	26(62)
	100	32(98)	28(64)
	500	32(98)	28(64)
	1000	32(98)	28(64)
	3000	32(98)	28(64)
Extended Maratos	4	16(44)	14(42)
	100	16(44)	14(42)
	500	16(44)	14(42)
	1000	16(44)	15(44)
	3000	16(44)	16(44)
FLETCHCR function (CUTE):	4	16(44)	16(42)
	100	16(46)	14(36)
	500	16(46)	16(42)
	1000	16(46)	15(38)
	3000	16(46)	16(42)
FLETCBV3 function (CUTE):	4	16(46)	15(44)
	100	16(44)	14(44)
	500	16(46)	14(44)
	1000	16(45)	14(44)
	3000	16(45)	14(44)

5. CONCLUSION

This paper gives a modified conjugate gradient method for solving unconstrained optimization in formula (2.7) and present new technique for modified Wolf conditions. The numerical results show that the given two modified methods are competitive to the Hestenes-Stiefel (HS) conjugate gradient method for the test problems, and it is shown that the search direction satisfied the descent condition.

6. ACKNOWLEDGEMENTS

I would like to express my thanks to Dr. Salah G. Shareef for his continuous advice and many helpful suggestions during out this work.

7. REFERENCES

[1] Al-Baali, M. (1985). “Descent property and global convergence of the Fletcher Reeves method with inexact line search”, IMA Journal of Numerical Analysis 5, pp.121-124

[2] Dai Y.H and Liao, L, Z. (2001), “New Conjugacecodition and related Nonlinear conjugate gradient methods” Applied mathematics and optimazation 43, pp.87-101.

[3] Dai Y.H and Yuan, Y. (1999), “A Nonlinear conjugate gradient method with a strong global convergence property” SIA M Journal on optimization 10, pp.177-182.

[4] Dixon L.C.W9 (1975), “Conjugate Gradient Algorithm Quadratic Termination Without linear searches” Journal of Int. of Math. and its App.

[5] Fletcher. (1987), “practical Method of Optimization”, John Wiley and son chichester.

[6] Fletcher R. and Reeves C., Function minimization by conjugate gradients, Comput. J. 7 (1964), pp. 149-154.

[7] Hestenes; M.R. and Stiefel, E. (1952). “Mehods of conjugate gradient for solving linear system”. Jornal of Research of the Notional Bureau of Standerd, 49, pp.409-436.

[8] Liu, D.C., Nocedal, J. On the Limited Memory Method for Large Scale Optimization. Math. Prog., 1989, 45: 503–528.

[9] Nocedal J.and Wright S.J. (1999), “Numerical Optimization”, Springer Series in Operations Research, ISBN 0-387.

[10] Ping-Qi P. (1984), “on advaced of Second order Quasi-Newton Method”, journal of computational mathematica, Vol. 2 No.1, January.

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

[12] Polyak B. T., The conjugate gradient method in extremem problems, USSR Comp. Math. And Math. Phys. 9 (1969), pp. 94-112.

[13] Roger F. (2011),”On the Barzilai.

Source of support: Nil, Conflict of interest: None Declared