International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 6 Issue 12, December-2017, Impact Factor: 4.059

Another Modified Conjugate Gradient Coefficient for Optimization

Dr. Basim A. Hassan[1], Alaa A. Mahmood[2]

[1]Assist Professor / College of Computers Sciences and Math.\ University of Mosul / Iraq
[2]Graduate student / College of Computers Sciences and Math.\ University of Mosul / Iraq

ABSTRACT

In this paper, a new modification of conjugate gradient coefficient with global convergence properties are presented. The global convergence result is established using line search Wolfe. Preliminary result shows that the proposed formula is competitive when compared to the other CG coefficients.

Keyword: Unconstrained optimization, Conjugate gradient method, Global convergent property.

INTRODUCTION

The capability of conjugate gradient (CG) methods to solve large-scale unconstrained optimization problems rendering it widely used in mathematical issues. The minimum values of the function for unconstrained optimization are obtained by using the nonlinear CG methods. Consider the unconstrained optimization problem:

$$\min \left\{ f(x) \;\middle|\; x \in R^n \right\} \qquad \text{..........} \quad (1)$$

where $f : R^n \to R^1$ is a continuously differentiable function [12]. The line search method for solving (1) is of the from :

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

where x_1 is a given initial point, d_k is a search direction, and α_k is a step size obtained by a 1-dimensional line search. In the steepest descent method, the search direction is defined as the negative gradient direction,

$$d_{k+1} = -g_{k+1} , \qquad \text{..........} \quad (3)$$

and the step size is chosen to be the 1-dimensional minimizer:

$$\alpha_k = \arg \min_{\alpha > 0} f(x_k + \alpha_k \mathrm{d}_k) \qquad \text{..........} \quad (4)$$

In practical computations, however, the steepest descent method performs poorly, and is badly affected by ill-conditioning. Another class of methods are quasi-Newton (QN) methods, where :

$$d_{k+1} = -B_{k+1} g_{k+1} , \qquad \text{..........} \quad (5)$$

and where B_{k+1} is updated at each iteration to capture the already obtained second derivative information. They are very efficient for medium scale problems, but can not be used to solve large scale problems because of its storage of matrices. The conjugate gradient (CG) method uses the negative gradient direction and the previous search direction to form the current search direction, namely,

$$d_{k+1} = -g_{k+1} + \beta_k d_k , \qquad \text{..........} \quad (6)$$

where $d_1 = -g_1$ and β_k is a scalar. More performance profile, is given in [13].

Since the emergence of the nonlinear conjugate gradient methods, several variants of β_k have been proposed corresponding to different conjugate gradient methods. Few of this parameters has proposed are given in table 1 below.

Table 1: Variants of Conjugate Gradient Updating Parameter

No.	Author(s)	Year	CG Parameter
1	Hestenes and Stiefel [11]	1952	$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{y_k^T d_k}$
2	Fletcher and Reeves [10]	1964	$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$
3	Dai and Yuan [8]	1999	$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{y_k^T d_k}$
4	Al-Bayati and Abbo [2]	2009	$\beta_k^{V1} = \left[1 - \frac{s_k^T y_k}{y_k^T y_k}\right] \frac{g_{k+1}^T y_k}{y_k^T d_k}$
5	Ahmed and ALNaemi [3]	2011	$\beta_k^{N1} = \left[1 - \frac{s_k^T s_k}{y_k^T s_k + 2(f_{k+1} - f_k) + (g_{k+1} + g_k)^T s_k}\right] \frac{g_{k+1}^T y_k}{y_k^T d_k}$
6	Basim and Hameed [4]	2013	$\beta_k^{New} = \left[1 - \frac{y_k^T s_k}{\Vert y_k \Vert^2 + \eta y_k^T s_k}\right] \frac{g_{k+1}^T y_k}{y_k^T d_k}$
7	Basim and Hameed [5]	2014	$\beta_k^{BH} = \left[1 + \frac{2(f_{k+1} - f_k) d_k^T d_k}{(d_k^T g_k)^2}\right] \frac{g_{k+1}^T y_k}{y_k^T d_k}$
8	Basim and Omer [6]	2014	$\beta_k^{BO1} = \left[1 - \frac{d_k^T s_k}{g_k^T g_k}\right] \frac{g_{k+1}^T y_k}{y_k^T d_k}$
9	Basim and Omer [7]	2015	$\beta_k^{BO2} = \left[1 - \frac{s_k^T (s_k - \gamma_k y_k)}{y_k^T (s_k - \gamma_k y_k)}\right] \frac{g_{k+1}^T y_k}{y_k^T d_k}$

In this paper, we show our new conjugate gradient method and algorithm. Next, we prove the global convergence of the new conjugate gradient method. Then, we report the numerical results and discussions. Lastly, the conclusions are presented.

A NEW CONJUGATE GRADIENT METHOD

In this section, we derive a new conjugate gradient method based on quadratic functions. Using the Taylor expansion to second-order terms, f can be written as:

$$f(x_k) = f(x_{k+1}) - g_{k+1}^T s_k + \frac{1}{2} s_k^T G s_k \quad \text{..........} \quad (7)$$

By using exact line search $g_{k+1}^T d_k = 0$, the above equation we get:

$$2(f_k - f_{k+1}) = s_k^T G s_k \quad \text{..........} \quad (8)$$

we get Hessian is the identity matrix scaled by G :

$$G = \frac{2(f_k - f_{k+1})}{s_k^T s_k} I_{n*n} \quad \text{..........} \quad (9)$$

The best direction to be following in the current point is the Newton direction:

$$\begin{aligned} d_{k+1} &= - G^{-1} g_{k+1} \\ &= - \frac{s_k^T s_k}{2(f_k - f_{k+1})} g_{k+1} \end{aligned} \quad \text{..........} \quad (10)$$

Since Newton direction are conjugate gradient with exact line searches we get:

$$-\left(\frac{s_k^T s_k}{2(f_k - f_{k+1})}\right) g_{k+1}^T y_k = -g_{k+1}^T y_k + \beta_k d_k^T y_k$$

$$\beta_k d_k^T y_k = -\left(\frac{s_k^T s_k}{2(f_k - f_{k+1})}\right) g_{k+1}^T y_k + g_{k+1}^T y_k \quad .... (11)$$

$$\beta_k d_k^T y_k = \left(1 - \frac{s_k^T s_k}{2(f_k - f_{k+1})}\right) g_{k+1}^T y_k$$

then we have :

$$\beta_k^{BA} = \left(1 - \frac{s_k^T s_k}{2(f_k - f_{k+1})}\right) \frac{g_{k+1}^T y_k}{d_k^T y_k} \quad .... (12)$$

$$d_{k+1} = -g_{k+1} + \left(1 - \frac{s_k^T s_k}{2(f_k - f_{k+1})}\right) \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k . \quad .... (13)$$

Also β_k^{BA} can be write in the manner:

$$\beta_k^{BA} = \left[1 - \frac{s_k^T s_k}{2(f_k - f_{k+1})}\right] \frac{g_{k+1}^T y_k}{s_k^T y_k}$$

$$= \left(1 - \frac{s_k^T s_k}{s_k^T y_k} \left[\frac{s_k^T y_k}{s_k^T s_k} * \frac{s_k^T s_k}{2(f_k - f_{k+1})}\right]\right) \frac{g_{k+1}^T y_k}{s_k^T y_k}$$

$$= \frac{1}{s_k^T y_k} \left(y_k - s_k^T y_k \left[\frac{s_k^T y_k}{s_k^T s_k} * \frac{s_k^T s_k}{2(f_k - f_{k+1})}\right] s_k \right)^T g_{k+1} \quad .... (14)$$

$$= \frac{1}{s_k^T y_k} \left(y_k - \frac{\|y_k\|^2}{s_k^T y_k} * \frac{(s_k^T y_k)^2}{\|y_k\|^2} \left[\frac{s_k^T y_k}{s_k^T s_k} * \frac{s_k^T s_k}{2(f_k - f_{k+1})}\right] s_k \right)^T g_{k+1}$$

$$= \frac{1}{s_k^T y_k} \left(y_k - t \frac{\|y_k\|^2}{s_k^T y_k} s_k \right)^T g_{k+1}$$

where $t = \frac{(s_k^T y_k)^2}{\|y_k\|^2} \left[\frac{s_k^T y_k}{s_k^T s_k} * \frac{s_k^T s_k}{2(f_k - f_{k+1})}\right]$. For simplicity, we call equation (14) by β_k^{BA}.

Now we can obtain the new conjugate gradient algorithm, as follows:

New Algorithm (BA Algorithm):

Step1. Give $x_1 \in R^n$, $\varepsilon > 0$ .Set $d_1 = -g_1, k = 1$, and parameters $0 < \delta_1 < \delta_2 < 1$. . If $\|g_1\| \le 10^{-6}$, then stop.

Step 2. Find $\alpha_{k+1} > 0$ satisfying the Wolfe conditions:

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k$$

$$d_k^T g(x_k + \alpha_k d_k) \ge \sigma d_k^T g_k$$

Step 3. Let $x_{k+1} = x_k + \alpha_k d_k$ and $g_{k+1} = g(x_{k+1})$. If $\|g_{k+1}\| \le 10^{-6}$, then stop.

Step 4. Compute β_k by the formulae (12), then generate d_{k+1} b y (13) .

Step 5. If the restart criterion of Powell $|g_{k+1}^T g_k| \ge 0.2 \|g_{k+1}\|^2$, is satisfied, then set $d_{k+1} = -g_{k+1}$ otherwise set $k = k + 1$ and continue with step 2.

CONVERGENT ANALYSIS

In this section, the convergence properties of BA will be studied. For an algorithm to converge, it is necessary to show that the sufficient descent condition and the global convergence properties hold.

Sufficient descent condition

For the sufficient condition to hold, then:

$$d_{k+1}^T g_{k+1} \le -c \left\|g_{k+1}\right\|^2 , \quad c > 0 . \qquad \text{..........} \quad (15)$$

Theorem 1.

Let $s_k , y_k , g_{k+1} \in R^n$, $\beta_k \in R$ and $\beta_k^{BA} = \frac{1}{s_k^T y_k}\left(y_k - t \frac{\|y_k\|^2}{s_k^T y_k} s_k \right)^T g_{k+1}$, where $t \in (1/4 , \infty)$. If $s_k^T y_k \ne 0$,

then $d_{k+1}^T g_{k+1} \le -\left[1 - \frac{1}{4t}\right]\|g_{k+1}\|^2$.

Proof :

Since $d_0 = -g_0$, we have $g_0^T d_0 = -\|g_0\|^2$, which satisfy (15) . Multiplying (21) by g_{k+1}, we have

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \left(\frac{g_{k+1}^T y_k}{s_k^T y_k} - t \frac{\|y_k\|^2}{(s_k^T y_k)^2} g_{k+1}^T s_k \right) s_k^T g_{k+1} \qquad \text{..........} \quad (16)$$

yielding

$$d_{k+1}^T g_{k+1} = \frac{(g_{k+1}^T y_k)(s_k^T g_{k+1})(s_k^T y_k) - \|g_{k+1}\|^2 (s_k^T y_k)^2 - t\|y_k\|^2 (g_{k+1}^T s_k)^2}{(s_k^T y_k)^2} \qquad \text{..........} \quad (17)$$

We applying the inequality $w^T v \le \frac{1}{2}(\|w\|^2 + \|v\|^2)$ with $w = \frac{1}{m}(s_k^T y_k) g_{k+1}$ and $v = m(g_{k+1}^T s_k) y_k$ where $m \in (\frac{1}{\sqrt{2}}, \sqrt{2t}\ ]$, to the first term of the above equality, we get :

$$(g_{k+1}^T y_k)(s_k^T g_{k+1})(s_k^T y_k) \le \frac{1}{2}\left[\frac{1}{m^2}(s_k^T y_k)^2 \|g_{k+1}\|^2 + m^2 (s_k^T g_{k+1})^2 \|y_k\|^2 \right] \qquad \text{..........} \quad (18)$$

This yields :

$$(g_{k+1}^T y_k)(s_k^T g_{k+1})(s_k^T y_k) \le \frac{1}{2}\left[\frac{1}{m^2}(s_k^T y_k)^2 \|g_{k+1}\|^2 + m^2 (s_k^T g_{k+1})^2 \|y_k\|^2 \right] \qquad \text{..........} \quad (19)$$

from (19) we get :

$$d_{k+1}^T g_{k+1} \le \left[\frac{1}{2m^2} - 1 \right] \|g_{k+1}\|^2 \le -\left[1 - \frac{1}{2m^2} \right] \|g_{k+1}\|^2 \qquad \text{..........} \quad (20)$$

Therefore, we get :

$$d_{k+1}^T g_{k+1} \le -\left[1 - \frac{1}{4t} \right] \|g_{k+1}\|^2 . \qquad \text{..........} \quad (21)$$

GLOBAL CONVERGENCE PROPERTIES

Next we will show that CG methods with BA converges globally. The following basic assumptions are needed in the analysis of global convergence properties of CG methods.

Assumptions

i- The level set $L = \left\{ x \in R^n \,\middle|\, f(x) \le f(x_0) \right\}$ is bounded.

ii- In some neighborhood U and L, $f(x)$ is continuously differentiable and its gradient id Lipschitz continuous, namely, there exists a constant $L > 0$ such that :

$$\|g(x_{k+1}) - g(x_k)\| \le L\|x_{k+1} - x_k\|, \ \forall\ x_{k+1}, x_k \in U \qquad \text{..........} \quad (22)$$

More details can be found in [8]. Under these assumptions on f, there exists a constant then a constant $\Gamma > 0$ exists, such that :

$$\|g_{k+1}\| > \Gamma \ , \qquad \text{..........} \quad (23)$$

for all $x \in L$.

Dai et al. [9] proved that for any conjugate gradient method with strong Wolfe line search the following general result holds :

Lemma 1.

Suppose that the assumptions (i) and (ii) hold and consider any conjugate gradient method (2) and (6), where d_{k+1} is a descent direction and α_k is obtained by the strong Wolfe line search (3) and (4). If

$$\sum_{k \ge 0} \frac{1}{\|d_{k+1}\|^2} = \infty \ , \qquad \text{..........} \quad (24)$$

then :

$$\lim_{k \Rightarrow \infty} \inf \|g_k\| = 0 \ . \qquad \text{..........} \quad (25)$$

Theorem 2.

Suppose that the assumptions holds and consider the method (2) and (6), where is a descent direction with β_k^{BA} given by (14), and α_k is obtained by the Wolfe line search. If the objective function is uniformly on S, then $\lim_{n \to \infty} \inf \|g_k\| = 0$.

Proof :

From (6) and definition of β_k by (14) we get :

$$\begin{aligned} \|d_{k+1}\| &= \|- g_{k+1} + \beta_k^{BA} s_k\| \\ &\le \|g_{k+1}\| + |\beta_k^{BA}| \|s_k\| \\ &\le \|g_{k+1}\| + \left\| \left(y_k - t \frac{\|y_k\|^2}{s_k^T y_k} s_k \right) \right\| \frac{\|g_{k+1}\|}{\|s_k\|\|y_k\|} \|s_k\| \end{aligned} \qquad \text{..........} \quad (26)$$

from (26) we get :

$$\begin{aligned} \|d_{k+1}\| &\le \|g_{k+1}\| + \frac{\|y_k\|\|g_{k+1}\| + t \dfrac{\|g_{k+1}\|\|y_k\|^2 \|s_k\|}{\|s_k\|\|y_k\|}}{\|s_k\|\|y_k\|} \|s_k\| \\ &\le \|g_{k+1}\| + \frac{\|y_k\|\|g_{k+1}\| + t\|g_{k+1}\|\|y_k\|}{\|s_k\|\|y_k\|} \|s_k\| \\ &\le [1 + 1 + t] \|g_{k+1}\| \le [2 + t] \|g_{k+1}\| \end{aligned} \qquad \text{..........} \quad (27)$$

This relation shows that :

$$\sum_{k \ge 1} \frac{1}{\|d_k\|^2} \ge \left(\frac{1}{2 + t} \right) \frac{1}{\Gamma} \sum_{k \ge 1} 1 = \infty \qquad \text{..........} \quad (28)$$

Therefore, from Lemma 1 we have $\lim_{k \to \infty} \inf \|g_k\| = 0$, which for uniformly convex function equivalent to $\lim_{k \to \infty} \|g_k\| = 0$.

NUMERICAL RESULTS

We test the new Algorithm by solving the ten problems from [1], and compare its performance to that of the FR method [10] under the Wolfe line search conditions. In new Algorithm, we have $\delta = 0.001$ and $\sigma = 0.9$, and the termination condition is $\|g_{k+1}\| \le 10^{-6}$.

The numerical results of our tests are reported in the Table 2. The first column "Problem" represents the name of the tested problem. Dim denotes the dimension of the test problems. NI, NR, and NF denote the number of iterations, the number of restart calls and function evaluations, respectively.

From the numerical results, it is shown that the proposed conjugate gradient method is promising.

DISCUSSION OF RESULT AND CONCLUSION

The results of Table 2 suggest that the proposed algorithm has promising behavior encountering with medium-scale and large-scale unconstrained optimization problems and it is superior to other considered algorithms in the most cases. With the Wolfe line search, the new methods are global convergent.

Table 2: Comparison of different CG-algorithms with different test functions and different dimensions

P. No.	n	FR algorithm NI	FR algorithm NR	FR algorithm NF	BA algorithm NI	BA algorithm NR	BA algorithm NF
1	100	32	13	64	12	7	24
	1000	77	46	129	12	7	25
2	100	37	8	67	44	17	71
	1000	73	27	115	52	23	85
3	100	180	60	313	85	28	162
	1000	F	F	F	102	36	198
4	100	89	32	174	70	44	155
	1000	107	40	211	F	F	F
5	100	F	F	F	9	7	23
	1000	60	31	131	11	9	39
6	100	124	41	231	78	18	140
	1000	445	196	711	149	27	265
7	100	11	6	26	10	7	24
	1000	16	12	125	8	6	21
8	100	130	49	196	102	29	163
	1000	364	119	593	F	F	F
9	100	13	7	25	13	7	26
	1000	15	7	29	11	6	23
10	100	121	65	218	73	24	119
	1000	345	169	634	242	75	378
11	100	74	21	123	85	26	139
	1000	370	88	616	257	67	435
12	100	12	7	25	7	5	15
	1000	11	7	23	7	5	15
13	100	20	11	33	9	6	17
	1000	19	11	35	13	8	24
14	100	122	62	156	15	9	26
	1000	130	66	166	12	9	22
15	100	112	55	147	37	14	56
	1000	110	54	145	37	12	65
Total		2748	1151	4627	1451	495	2525

Fail: The algorithm fail to converge.

Problems numbers indicant for: 1. is the Extended Tridiagonal 1, 2. is the Generalized Tridiagonal 2, 3. is the Extended Powell, 4. is the Extended Maratos, 5. is the Extended Cliff, 6. is the Quadratic Diagonal Perturbed, 7. is the Extended Quadratic Penalty QP1, 8. is the Quadratic QF2, 9. is the NONDIA (CUTE), 10. is the DIXMAANE (CUTE), 11. is the Partial Perturbed Quadratic, 12. is the DENSCHNB (CUTE), 13. is the DENSCHNA (CUTE), 14. is the Extended Block-Diagonal BD2, 15. is the Generalized quartic GQ2.

REFERENCES

[1]. Andrie N., (2008), ' An Unconstrained Optimization Test functions collection ' Advanced Modeling and optimization. 10, pp.147-161.
[2]. Al-Bayati A.Y., Salim A.J. and Abbo K., (2009). Two versions of conjugate gradient algorithms based on conjugacy conditions for unconstrained optimization. American J. of Economics and Business Administration, 1. pp. 97-104.
[3]. Ahmed H. I. and AL-Naemi, (2011). New two types of conjugate gradient algorithms for unconstrained optimization. Canadian J. on Science and Engineering Mathematics. 2, pp. 237-246.
[4]. Basim A. Hassan and Hameed M. Sadiq., (2013). A Nonlinear Conjugate Gradient Method Based on a Modified Secant Condition. Iraqi Journal of Statistical Sciences. 24, pp.1-16.
[5]. Basim A. Hassan and Hameed M. Sadiq., (2014). A Modified Class of Conjugate Gradient Algorithms Based on Quadratic Model for Nonlinear Unconstrained Optimization. Raf. J. of comp & math. 1, pp.25-37.
[6]. Basim A. H. and Omer M. E., (2014),' A New sufficient descent Conjugate Gradient Method for Nonlinearpp. Iraqi Journal of Statistical Sciences, 26, pp.12-24.
[7]. Basim A. H. and Omer M. E., (2015),' A new type of Conjugate Gradient Method with Guaranteed Descent. To be abber.
[8]. Dai Y. H. and Yuan Y., (1999). A nonlinear conjugate gradient method with a strong global convergence property, SIAM J. optimization, pp. 177-182.
[9]. Dai Y.H., Han, J.Y., Liu, G.H., Sun, D.F., Yin, .X. and Yuan, Y., (1999),'Convergence properties of nonlinear conjugate gradient methods. SIAM Journal on Optimization 10 pp. 348-358.
[10]. Fletcher R. and Reeves C. M.,(1964). Funtion minimization by conjagate gradients, Computer Journal 7, pp. 149-154.
[11]. Hestenes, M. R. and Stiefel, E. L. (1952), ' Method of conjugate gradients for Solving linear systems ' Journal National Standards 49, pp. 409-436.
[12]. Wan K., Mohd. R., Mustafa M. and Ibrahim J., (2016),'A spectral KRMI conjugate gradient method under the strong Wolfe line search, AIP Conference Proceedings Published by the American Institute of Physics, pp. 1-8.
[13]. Yu D. and Qin N., (2003),' Testing different conjugate gradient methods for large scale unconstrained optimization, Journal of computational mathematics, 21, pp.311-320.