Available online at sjuoz.uoz.edu.krd
Science Journal of University of Zakho
Vol. 7, No. 1, pp. 34–36, Mar.-2019
journals.uoz.edu.krd
p-ISSN: 2663-628X
e-ISSN: 2663-6298

A NEW CONJUGATE GRADIENT COEFFICIENT FOR UNCONSTRAINED OPTIMIZATION BASED ON DAI-LIAO

Alaa Luqman Ibrahim [a,*], Muhammad Amin Sadiq [b], Salah Gazi Shareef [c]

[a] Department of Mathematics, College of Science, University of Duhok, Duhok, Kurdistan Region, Iraq (alaa.ibrahim@uod.ac)
[b] College of Administration and Economics, Cihan University Duhok, Kurdistan Region, Iraq (muhammad.math@uod.ac)
[c] Department of Mathematics, Faculty of Science, University of Zakho, Kurdistan Region, Iraq (salah.shareef@uoz.edu.krd)

Received: Dec. 20-18 / Accepted: Feb., 2019 / Published: Mar., 2019 https://doi.org/10.25271/sjuoz.2019.7.1.525

ABSTRACT:
Conjugate gradient method plays an enormous role in resolving unconstrained optimization problem, particularly for large scale. In this paper, a new conjugate gradient method for unconstrained optimization based on Dai-Liao (DL) formula by using Barzilai and Borwein step size. Our new method satisfies both descent and sufficient descent conditions. The numerical results show that the proposed algorithm is potentially efficient and performs better than with Polak and Ribiere (PR) algorithm, depending on number of iterations (NOI) and the number of functions evaluation (NOF).

KEYWORDS*: conjugate gradient, unconstrained optimization, Barzilai and Borwein step size, descent and sufficient descent conditions.*

1. INTRODUCTION

The gradient method has great role in developing optimization techniques, and it has been associated with some well-known methods such steepest method for the unconstrained minimization (Dai and Liao, 2001).
One of the important method which is designed to solve unconstrained optimization problem is called conjugate gradient method (CG).

$$\min f(x) \qquad \forall\, x \in R^n \qquad (1.1)$$

Where $f: R^n \to R$, $f \in C^1$ and its gradient at point x_i denoted by $g_i = \nabla f(x_i)$ or g_i, especially when our dimension is large. Therefore, the nonlinear CG-method formula is given by:

$$x_{i+1} = x_i + \lambda_i d_i, \quad i = 0, 1, 2, \ldots \qquad (1.2)$$

where $\lambda_i > 0$ is a step size, and d_i is a search direction which is determined by:

$$d_i = \begin{cases} -g_0 & \text{if} \quad i = 0 \\ -g_{i+1} + \beta_i d_i & \text{if} \quad i \geq 1 \end{cases} \qquad (1.3)$$

where β_i represents the conjugate gradient coefficient parameter. See ((Hestenes and Stiefel, 1952), (Polak and Ribiere, 1969), (Polyak, 1969), (Fletcher and Reeves, 1964), (Liu, and Storey, 1991), (Fletcher, 1987), (Dai and Yuan, 1996) and (Dai and Liao, 2001) to be familiar with some well-known classical formulas for β_i. The parameter β_i of the classical formula is determined as follows:

$$\beta_i^{HS} = \frac{g_{i+1}^T y_i}{d_i^T y_i} \qquad (1.4)$$

$$\beta_i^{PR} = \frac{g_{i+1}^T y_i}{g_i^T g_i} \qquad (1.5)$$

$$\beta_i^{FR} = \frac{g_{i+1}^T g_{i+1}}{g_i^T g_i} \qquad (1.6)$$

$$\beta_i^{LS} = \frac{g_{i+1}^T y_i}{-g_i^T d_i} \qquad (1.7)$$

$$\beta_i^{CD} = \frac{g_{i+1}^T g_{i+1}}{-g_i^T d_i} \qquad (1.8)$$

$$\beta_i^{DY} = \frac{g_{i+1}^T g_{i+1}}{d_i^T y_i} \qquad (1.9)$$

$$\beta_i^{DL} = \frac{g_{i+1}^T (y_i - t s_i)}{d_i^T y_i}, \text{ where } t > 0 \qquad (1.10)$$

Consider $\|.\|$ denotes the Euclidean norm and $y_i = g_{i+1} - g_i$. The global convergence of above conjugate gradient methods is studied by many researchers see (Hager, and Zhang, 2006), (Liu, and Storey, 1991) and (Wolfe, 1969). To prove the convergence condition of these methods, we need the step size λ_i satisfies the following strong Wolfe conditions:

$$f(x_i + \lambda_i d_i) \leq f(x_i) + \delta \lambda_i g_i^T d_i \qquad (1.11)$$

$$|g(x_i + \lambda_i d_i)^T d_i| \leq -\rho g_i^T d_i \qquad (1.12)$$

Where $0 < \delta < \rho < 1$.
However, the standard Wolfe conditions are employed to prove the convergence of numerical methods such as Qusi-Newton method (Wolfe, 1969):

$$f(x_i + \lambda_i d_i) \leq f(x_i) + \delta \lambda_i g_i^T d_i \qquad (1.13)$$

$$g(x_i + \lambda_i d_i)^T d_i \geq \rho g_i^T d_i. \qquad (1.14)$$

The majority studied characteristics of CG are its global convergence properties. The global convergence of Fletcher and Reeves (FR)is proved by (Zoutendijk, 1970). Analyzed the global convergence of algorithm regarding the Fletcher and Reeves method by Wolfe condition (Al-Baali, 1985). Further, FR as a superior method was proved by (Powell, 1986). Furthermore, Neculai Andrei presented a collection of unconstrained optimization functions in 2008.
This paper organized as follows: In section two, the new conjugate gradient (CG) algorithm is presented with its algorithm. In section three, the new conjugate gradient satisfies the descent condition and sufficient descent condition. In section four the numerical results, percentages, graphics and discussion are presented. Finally, we conclude in section five.

* Corresponding author

2. NEW CONJUGATE GRADIENT COEFFICIENT

The aim of this section is to derive a new conjugate gradient coefficient known as β_i^{New} for unconstrained optimization from the Dai-Liao (DL) formula.

2.1 The new CG coefficient ($\boldsymbol{\beta_i^{New}}$)

Barzilai and Borwein (1988), proposed a new value of λ_i defined by

$$\boldsymbol{\lambda_i} = \frac{s_i^T s_i}{s_i^T y_i} \qquad \textbf{(2.1)}$$

Consider $\boldsymbol{s_i = x_{i+1} - x_i = \lambda_i d_i = -\lambda_i g_i}$ **(2.2)**

By putting (2.2) in (2.1), we get

$$\boldsymbol{\lambda_i^* = \lambda_i \frac{g_i^T g_i}{d_i^T y_i}}$$

Let, $\boldsymbol{\beta_i^{New} = \frac{\lambda_i}{\lambda_i} \beta_i^{DL}}$ **(2.3)**

Now, we replace λ_i by λ_i^* in denominator of equation (2.3), and get new parameter as follows:

$$\boldsymbol{\beta_i^{New}} = \frac{\lambda_i g_{i+1}^T (y_i - ts_i)}{\lambda_i \frac{g_i^T g_i}{d_i^T y_i} d_i^T y_i}$$

After some algebraic operations, we obtain

$$\boldsymbol{\beta_i^{New}} = \frac{g_{i+1}^T y_i}{g_i^T g_i} - t\frac{g_{i+1}^T s_i}{g_i^T g_i}$$

$$\boldsymbol{\beta_i^{New} = \beta_i^{PR} - t\frac{g_{i+1}^T s_i}{g_i^T g_i}}, \textbf{ where } \boldsymbol{t > 0} \qquad \textbf{(2.4)}$$

2.2 Outline of new CG algorithm ($\boldsymbol{\beta_i^{New}}$)

Step 1: Given $x_0 \in R^n$.
Step 2: $i = 0$, $g_0 = \nabla f(x_0)$, $d_0 = -g_0$, if $g_0 = 0$, stop.
Step 3: Compute λ_i by using cubic line search to minimize $f(x_i + \lambda_i d_i)$,
(i.e.) $f_{i+1} \le f_i$.
Step 4: Updating new point based on Eq. (1.2)
Step 5: Compute g_{i+1}, if $\|g_{i+1}\| \le 10^{-5}$ stop.
else determine $s_i = x_{i+1} - x_i$ and $y_i = g_{i+1} - g_i$
Step 6: Evaluate d_{i+1} by Eq. (1.3), where β_i is computed by (2.4).
Step 7: If $\|g_{i+1}\|^2 \le \frac{|g_i^T g_{i+1}|}{0.2}$ is satisfied go to step 2,
else
$i = i + 1$ and go to step 3

3. THE PROOF OF THE DESCENT AND THE SUFFICIENT DESCENT CONDITION OF THE NEW CG ALGORITHM

Clearly, any conjugate gradient coefficient should satisfy both the decent and sufficient decent conditions. Therefore, in this section, we are going to show that new conjugate gradient β_i^{New} satisfies the descent and the sufficient descent conditions.

Theorem 3.1 Suppose that the sequence $\{x_i\}$ is generated by (1.2), then the search direction (1.3) with β_i^{New} given as (2.4), satisfies the descent condition. i.e.

$$\boldsymbol{d_{i+1}^T g_{i+1} \le 0} \qquad \textbf{(3.1)}$$

Proof: By combining the equations (1.3) and (2.4), we obtain

$$\boldsymbol{d_{i+1} = -g_{i+1} + (\beta_i^{PR} - t\frac{g_{i+1}^T s_i}{g_i^T g_i}) d_i} \qquad \textbf{(3.2)}$$

After multiplying both sides of the equation (3.2) by g_{i+1}^T, the following formula is obtained

$$\boldsymbol{g_{i+1}^T d_{i+1} = -\|g_{i+1}\|^2 + \beta_i^{PR} g_{i+1}^T d_k - t\frac{g_{i+1}^T s_i}{g_i^T g_i} g_{i+1}^T d_i}$$

Since $s_i = \lambda_i d_i$

$$\boldsymbol{g_{i+1}^T d_{i+1} = -\|g_{i+1}\|^2 + \beta_i^{PR} g_{i+1}^T d_k - t\lambda_i \frac{(g_{i+1}^T d_i)^2}{g_i^T g_i}} \qquad \textbf{(3.3)}$$

If $d_i^T g_{i+1} = 0$, then the equation (3.3) gives $g_{i+1}^T d_{i+1} = -\|g_{i+1}\|^2 \le 0$. Then the proof is completed.
On the other hand, if $d_i^T g_{i+1} \ne 0$. Hence, the first two terms of equation (3.3) are equal or less than zero because the PR algorithm satisfies the descent condition, i.e.

$$\boldsymbol{-\|g_{i+1}\|^2 + \frac{(g_{i+1}^T y_i)(g_{i+1}^T d_i)}{g_i^T g_i} \le 0,}$$

Clearly t, λ_i, $g_i^T g_i$ and $(g_{i+1}^T d_i)^2$ are positive. Here, we obtain the third term of equation (3.3) less than or equal to zero. Hence

$$\boldsymbol{g_{i+1}^T d_{i+1} = -\|g_{i+1}\|^2 + \beta_i^{PR} g_{i+1}^T d_i - t\lambda_i \frac{(g_{i+1}^T d_i)^2}{g_i^T g_i} \le 0} \ .\blacksquare$$

Theorem 3.2 Assume that d_{i+1} given by (1.3) and (2.4) and the step size λ_i is obtained by (1.11) and (1.12) then, the sufficient descent condition is satisfied, i.e.

$$\boldsymbol{d_{i+1}^T g_{i+1} \le -c\|g_{i+1}\|^2} \qquad \textbf{(3.4)}$$

Proof: It is clear that the first two terms of equation (3.3) are equal or less than to zero. Therefore, the equation (3.3) can be written as follows:

$$\boldsymbol{g_{i+1}^T d_{i+1} \le -(t\lambda_i \frac{(g_{i+1}^T d_i)^2}{g_i^T g_i \|g_{i+1}\|^2}) \ \|g_{i+1}\|^2} \qquad \textbf{(3.5)}$$

Which means (3.4) holds where $c = t\lambda_i \frac{(g_{i+1}^T d_i)^2}{g_i^T g_i \|g_{i+1}\|^2}$.■

4. NUMERICAL RESULTS

In this section, we examine the implementation of the new CG method. The tests include well-known nonlinear problems standard test functions (Andrei, 2008), with different dimensions $4 \le n \le 5000$. The FORTRAN 95 language is the program which is used to write the numerical algorithm.

Table 1. Comparison between new CG and PR methods

Test Function	n	CG (PR)		CG (New)	
		NOI	NOF	NOI	NOF
Wood	4	29	67	26	60
	100	30	69	27	62
	500	30	69	27	62
	1000	30	69	27	62
	5000	30	69	28	64
Central	4	22	159	22	158
	100	22	159	22	158
	500	23	171	22	158
	1000	23	171	23	171
	5000	30	270	23	171
Cubic	4	15	45	13	37
	100	16	47	13	37
	500	16	47	13	37
	1000	16	47	13	37
	5000	16	47	14	39

Edger	4	5	14	5	14
	100	5	14	5	14
	500	6	16	5	14
	1000	6	16	5	14
	5000	6	16	5	15
Rosen	4	30	85	27	74
	100	30	85	27	74
	500	30	85	27	74
	1000	30	85	29	74
	5000	30	85	29	74
Mile	4	37	116	35	107
	100	44	148	41	139
	500	44	148	41	139
	1000	50	180	41	139
	5000	50	180	47	170
Powell	4	40	120	31	84
	100	43	135	38	117
	500	46	150	38	117
	1000	46	150	41	134
	5000	50	180	41	134
Total		**976**	**3514**	**871**	**3034**

The numerical results in Table (1) illustrate that, the new conjugate gradient method is more efficient than standard (PR) method with respect to NOI and NOF. Indeed, in central function we have not seen any deference in terms of NOI when $n \leq 500$, yet there is a small deference in terms of NOF. However, for the other functions the result of the new CG is more efficient starting from the very beginning iterations. Finally, it is clear that the new CG is reliable function compare with the standard one.

Table 2. Comparison improvement ratio between the β_i^{New} and PR methods

Tools	CG (PR)	CG (New)
NOI	100%	89.2418%
NOF	100%	86.3404%

Clearly, in Table 2 the NOI and NOF of the PR method are about 100%. That means, the new method has improvement of 10.7582% compared with standard method in NOI and 13.6596% in NOF. Generally, the New (CG) method was improved by 12.2089% compared with PR method.

5. CONCLUSION

Numerous studies of conjugate gradient lead to several of new methods. In this paper, a modified (CG) method for solving nonlinear unconstrained optimization in formula (2.4) based on Dai-Liao (DL) formula by Barzilai and Borwein step size is presented. We found that the new conjugate gradient satisfies both conditions (descent and sufficient descent) and it is reliable for functions up to 4 variables. Limited numerical experiments and comparisons show that, the new algorithm is better than CG (PR) according to the NOI and NOF. In future one can use our new CG in neural network in order to investigate the efficiency of its behavior.

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. Advanced Modeling and Optimization, 10 , 147-161.

Barzilai, J. and Borwein, J.M. (1988), Tow point step size gradient methods, IMA J. Numer. Anal., 8, 141-148.

Dai, Y. H. and Liao, L.Z., (2001), New conjugacy conditions and related nonlinear conjugate gradient methods, Application Mathematical Optimization, 43, 87-101.

Dai, Y. H. and Yuan, Y., (1996), Convergence properties of the Fletcher-Reeves method, IMAJ. Numer. Anal., 2, 155-164.

Fletcher, R. and Reeves, C.M., (1964), Function minimization by conjugate gradients, The Computer Journal. 7, 149-154.

Fletcher, R., (1987), Practical methods of optimization unconstrained optimization, John Wiley & Sons, New York, NY, USA.

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

Hager,W. W. and Zhang, H., (2006),A survey of nonlinear conjugate gradient methods, Pacific Journal of Optimization, 2, 35-58.

Hestenes, M. R. and Stiefel, E., (1952), Methods of conjugate gradients for solving linear systems, Journal of Research of the National Bureau of Standards. 49, 409-436.

Liu, Y. and Storey, C., (1991), Efficient generalized conjugate gradient algorithms, part 1: Theory, Journal of Optimization Theory and Applications, 69, 129-137.

Polak, E. and Ribiere, G., (1969), Note surla convergence des méthodes de directions conjuguées., 3(16), 35-43.

Polyak, B. T., (1969), The conjugate gradient method in extreme problems, USSR Comp. Math. and Math. Phys., 94-112.

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

Wolfe, P. (1969), Convergence conditions for ascent methods, SIAM. Rev. 11, 226-235.