Journal of University of Duhok, Vol. 24, No.2 (Pure and Eng. Sciences), Pp 55-61, 2021

A MODIFIED CONJUGATE GRADIENT METHOD WITH GLOBAL-CONVERGENCE FOR UNCONSTRIANED OPTIMIZATION PROBEMS

ASMAA M. ABDULRAHMAN* and BAYDA G. FATHI**

*Dept. of mathematics, College of Science, University of Duhok, Kurdistan Region-Iraq

**Dept. of mathematics, Faculty of Science, University of Zakho, Kurdistan Region-Iraq

(Received: July 10, 2021; Accepted for Publication: October 19, 2021)

ABSTRACT

In this paper a new conjugate gradient method for unconstrained optimization is suggested, the new method is based on Conjugate Descent (CD) formula which is a modified of the CD formula and which is also sufficiently descent and globally convergent. Numerical evidence shows that this new conjugate gradient algorithm is considered as one of the competitive conjugate gradient methods.

1. INTRODUCTION

As it is shown in the following unconstrained optimization problem

$$\min\{f(x): x \in R^n\} \qquad (1)$$

where f is a function of continuously differentiability of n real variables with gradient $g = \nabla f$. Methods of conjugate gradient are a class of significant ways for solving (1), particularly for problems with large scale, which have the following form:

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

in the sense that α_k is the step size, d_k is the conjugate search direction. The direction of the search is usually defined by:

$$d_k = \begin{cases} -g_k & k = 1 \\ -g_k + \beta_k d_{k-1} & k > 1 \end{cases} \qquad (2)$$

where $g(x) = \nabla f(x)$ and β_k is scalar parameter. Certain choices of the parameter β_k are given which are in chronological order.

One of the formulae for β_k is Conjugate Descent (CD) of Fletcher [5, 6]:

$$\beta_k^{CD} = \frac{-\|g_k\|^2}{d_{k-1}^T g_{k-1}} \qquad (3)$$

where $y_k = g_{k+1} - g_k$ and $\|.\|$ denoted the Euclidean norm. The step size α_k is obtained through the exact linear search or inexact linear search α_k should satisfy the conditions of strong Wolfe [1]:

$$f(x_k + \alpha_k d_k) \le f(x_k) + \rho \alpha_k g_k^T d_k \qquad (4)$$

and

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k \qquad (5)$$

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

Conjugate Descent ensures a descent direction for general functions if the line search satisfies the strong Wolfe conditions (4), (5) with $\sigma < 1$. But the global convergence of the CD method is proved, see [2].

In 1999, Dai and Yuan [1] proposed the DY conjugate gradient method using β_k defined by (3). In 2001 [3], they proposed an updated form of β_k with three parameters, which may be regarded as a convex combination of several earlier choices of β_k listed in [7]; but the three parameters are restricted in small intervals.

In [4], Dai and Yuan proposed a family of globally convergent conjugate methods, in which β_k introduced by

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}.$$

$$\beta_k = \frac{\|g_k\|^2}{\lambda |g_{k-1}|^2 + (1-\lambda) d_{k-1}^T y_{k-1}},$$

where $\lambda \in [0\,,1]$ is a parameter.

In this note, we present a new formula which is a modified of the CD method defined by

$$\beta_k^{New} \frac{\|g_k\|^{\,2}}{\mu |d_{k-1}^T g_k| - d_{k-1}^T g_{k-1}}. \qquad (6)$$

Note that if we use the exact line search, our new algorithm reduces to the standard CD algorithm. In this paper, however, we consider general nonlinear functions and an inexact line search.

The rest of the paper is organized as follows. In Section 2, the new suggested algorithm is presented. The sufficient descent and Global convergence properties of the proposed methods are analysed in Section 3. Some numerical results are reported in Section 4. Finally, we draw a conclusion in Section 5.

1. ALGORITHM FOR NEW METHOD

In this section, we state the steps of our algorithm.

step 1: Set $k = 1$, select initial point $x_1 \in R^n$, a very small positive $\varepsilon > 0$

step 2: $g_k = \nabla f(x_k)$, if $g_k = 1$ then stop

else set $d_1 = -g_1$

step 3: Compute step length α_k to minimize $f(x_k)$

step 4: $x_{k+1} = x_k + \alpha_k d_k$

step 5: $g_{k+1} = \nabla f(x_{k+1})$
step 6: Compute the parameter β_k by (6).
step 7: $d_{k+1} = -g_{k+1} + \beta_k d_k$,
step 8: If $k = i$, or if $\|g_k\| < \varepsilon$ is satisfied then go to step 2
else $k = k + 1$ then go to step 3

We use the same algorithm in [1] which is restated here for the sake of convenience.

The convergence properties of the new method are stated in the following theorem.

2. CONVERGENCE OF THE NEW METHOD

In this section, we give the search direction d_k^T satisfies the sufficient descent condition. We can notice by (5) that

$$d_{k-1}^T(g_k - g_{k-1}) \geq \sigma d_{k-1}^T g_{k-1} - d_{k-1}^T g_{k-1} = (\sigma - 1) d_{k-1}^T g_{k-1} \tag{7}$$

Lemma 1: If $\mu > 1$, then

$$g_k^T d_k < -(1 - \frac{1}{\mu})\|g_k\|^2 < 0 \quad \text{for } k = 1,2, \ldots.$$

Proof: If $k = 1$ then $d_1 = -g_1$ and since $\mu > 1$ we get

$$g_1^T d_1 = -\|g_1\|^2 < -(1 - \frac{1}{\mu})\|g_1\|^2 < 0.$$

Assume by induction that $g_{k-1}^T d_{k-1} = -\|g_{k-1}\|^2 < -(1 - \frac{1}{\mu})\|g_{k-1}\|^2 < 0$. By (2), (5), (6) and (7), we have

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

$$= -\|g_k\|^2 + \frac{\|g_k\|^2}{\mu|d_{k-1}^T g_k| - d_{k-1}^T g_{k-1}} g_k^T d_{k-1}$$

$$\leq \|g_k\|^2 + \frac{\|g_k\|^2}{\left|\mu|d_{k-1}^T g_k| - d_{k-1}^T g_{k-1}\right|} |g_k^T d_{k-1}|$$

$$\leq \|g_k\|^2 + \frac{\|g_k\|^2}{\mu|d_{k-1}^T g_k|} |g_k^T d_{k-1}|$$

$$= -(1 - \frac{1}{\mu})\|g_k\|^2$$

we remark that our Lemma 1 implies that d_k is a sufficient descent direction in (6).

Lemma 2: Once the sequence $\{x^k\}$ is clearly generated by (1) and (2), the step size α_k satisfies (5) and (6), and d_k is a direction that descents, f is bounded and $g(x)$ satisfies Lipschitz condition in the level set, then:

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{8}$$

The proof of this Lemma is given in [1].

Global convergence of our method will be established and numerical evidence will be listed to support.

Theorem 1 (Global convergence): If $\mu > 1$ in (6), f is bounded and $g(x)$ is Lipschitz in the level set, thus our algorithm either terminates at a stationary point or $\lim_{k\to\infty} \|g_k\| = 0$.

Proof: in case the conclusion never holds, then then a real number does exist $\varepsilon > 0$ such that

$\|g_k\| > \varepsilon$, for all $k = 1,2,3, \ldots$ since $d_k + g_k = \beta_k d_{k-1}$, we have :

$$\|d_k\|^2 = \beta_k^2 \|d_{k-1}\|^2 - \|g_k\|^2 - 2g_k^T d_k \tag{9}$$

Now

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

$$= -\|g_k\|^2 + \frac{\|g_k\|^2}{\mu|d_{k-1}^T g_k| - d_{k-1}^T g_{k-1}} g_k^T d_{k-1}$$

$$\leq \frac{d_{k-1}^T g_{k-1} + g_k^T d_{k-1}}{\mu|d_{k-1}^T g_k| - d_{k-1}^T g_{k-1}} \|g_k\|^2$$

There exist σ such that:

$$\sigma g_{k-1}^T d_{k-1} \leq g_{k-1}^T d_{k-1} \leq -\sigma g_{k-1}^T d_{k-1}$$

Then

$$\sigma g_{k-1}^T d_{k-1} \leq \frac{d_{k-1}^T g_{k-1} - \sigma g_{k-1}^T d_{k-1}}{\mu|d_{k-1}^T g_k|} \|g_k\|^2$$

$$= \frac{(1-\sigma) d_{k-1}^T g_{k-1}}{\mu|d_{k-1}^T g_k|} \|g_k\|^2$$

Since $d_{k-1}^T g_{k-1} < 0$ and $d_k^T g_k < 0$ we see that

$$\|g_k\|^2 \leq \frac{\mu|d_{k-1}^T g_k|}{(1-\sigma)|d_{k-1}^T g_{k-1}|} d_k^T g_k$$

That is,

$$\beta_k^{New} = \frac{\|g_k\|^2}{\mu|d_{k-1}^T g_k| - d_{k-1}^T g_{k-1}}$$

$$\leq \frac{\|g_k\|^2}{\mu|d_{k-1}^T g_k|} \leq \frac{|d_k^T g_k|}{(1-\sigma)|d_{k-1}^T g_{k-1}|}$$

Replace β_k in (9) with β_k^{New} , we get

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

$$= \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} - \left(\frac{\|g_k\|}{g_k^T d_k} + \frac{1}{\|g_k\|}\right)^2 + \frac{1}{\|g_k\|^2}$$

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

Since $d_1 = -g_1$, so that

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} < \frac{\|d_1\|^2}{(g_1^T d_1)^2} + \frac{k-1}{\varepsilon^2} = \frac{1}{\|g_1\|^2} + \frac{k-1}{\varepsilon^2} < \frac{1}{\varepsilon^2} + \frac{k-1}{\varepsilon^2} = \frac{k}{\varepsilon^2}$$

Thus

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} > \sum_{k=1}^{\infty} \frac{\varepsilon^2}{k} = +\infty$$

which is contrary to Lemma 2. Hence $\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty$

3. NUMERICAL RESULTS

As it is shown in Table (1), a comparison between the new introduced algorithm and the standard CD method will be done. Such the comparative tests consist of well-known non-linear problems with 8 various functions for n =4, 100, 1000, 5000. Moreover, we wrote the code in Fortran 95 language and we take $\mu > 1,$ for all cases the stopping condition $\|g_{k+1}\| \leq 1 \times 10^{-5}$. On the other hand, the comparative results demonstrated in Table (1) which contains number of iteration (NOI) and number function evaluations (NOF). The experimental results are given in Table (1) verifying that the new conjugate gradient algorithm (β_k^{New}) is superior to standard (CD) with respect to (NOI) and (NOF).

Table (1):- Comparing the two algorithms standard (CD) New algorithm performance

Test functions	n	New Algorithm		Standard CD method	
		NOI	NOF	NOI	NOF
SUM	4	3	11	3	11
	100	14	85	14	85
	500	21	119	22	114
	1000	26	144	26	126
	5000	33	162	38	198
FRED	4	8	23	8	23
	100	8	23	9	25
	500	8	23	9	25
	1000	8	23	9	25
	5000	8	23	9	25
GWood	4	25	60	28	65
	100	25	60	28	65
	500	26	62	29	68
	1000	26	62	29	68
	5000	26	62	29	68
ROSEN	4	27	76	30	85
	100	29	81	30	85
	500	29	81	30	85
	1000	29	81	30	85
	5000	29	81	30	85
NON-DIAGONAL	4	20	57	23	61
	100	27	74	27	73
	500	27	71	27	73
	1000	27	71	27	73
	5000	27	71	27	73
MIELE	4	34	98	51	171
	100	40	130	68	246
	500	46	160	68	246
	1000	46	160	68	246
	5000	52	191	74	284
CUBIC	4	12	37	13	38
	100	13	39	14	40
	500	13	39	15	44
	1000	13	39	15	44
	5000	13	39	15	44
Gcantrel	4	14	88	12	67
	100	18	139	18	142
	500	19	152	23	210
	1000	19	152	23	210
	5000	22	195	28	278
Total		910	3344	1076	4079

In Table (2), the rate of improvement in the new algorithm with the standard CD method is demonstrated.

Table (2): -Comparison of the rate of improvement between the standard CD method and the new algorithm

Tools	Standard CD method	New algorithm
NOI	100%	84.5724%
NOF	100%	81.9808 %

5. CONCLUSION

The numerical performance of the new algorithm is better than the standard algorithm, as it is noticed that (NOI), (NOF) of the standard method are around 100%, that reflects the new algorithm has improved as compared to standard with **(15.43%)** in (NOI) and **(18.02%)** in (NOF).

REFERENCES

Dai, Y. H. and Yuan, Y. (1999). A Nonlinear Conjugate Gradient with a Strong Global Convergence Property, SIAM Journal of Optimization, 10, 177-182.

Dai, Y. H. and Yuan, Y. (1996). Convergence Properties of the Conjugate Descent Method, Mathematical Advances, 25,552-562.

Dai, Y. H. and Yuan, Y. (2001). A Three-Parameter Family of Nonlinear Conjugate Gradient Methods, Math. Comp., 70, 1155-1167.

Dai, Y. H. and Yuan, Y. (1998). A Class of Globally Convergent Conjugate Gradient Methods, Research report ICM-98-030, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences.

Fletcher, R. (1989). Practical Method of Optimization, (2nd ed.), john Wiley , New York.

Fletcher, R. and Reeves (1964). Function Minimization by Conjugate Gradients. published in the computer journal, publisher (Oxford University Press), Colin M, 7, 149-154.

Hager, W. W. and Zhang, H. C. (2006). A Survey of Nonlinear Conjugate Gradient Methods. Pac. J. Optim., 2, 35-58.