Modified Conjugate Gradient Method for Unconstrained Optimization

Thamera . K . Alkhashab
Department of Computer Systems,
Technical institute of Mosul, Mosul, IRAQ

ABSTRACT
Conjugate gradient method holds an important role in solving unconstrained Optimizations , especially for large scale problems. Numerous studies and modific ations have been done to improve this method . In this paper , we propose a new conjugate gradient meth od which is computed by modifying Dai and Yuan formula . This new β_k formula for the denominator is introduced and the numerator of Dai and Yuan for mula is retrained , but still possesses global converge nce properties. Numerical results based on number of iterations and number of function evaluations by usin g exact line search have shown that the new formul a is an efficient when we comparative it with the oth er conjugate gradient methods.

Keywords
Conjugate gradient methods , global convergence , unconstrained optimization , exact line search

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

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

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

$$x_{k+1} = x_k + \alpha_k d_k \quad k=0,1,2,--- \qquad (1.2)$$

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:

$$d_{k+1} = \begin{cases} -g_{k+1} & , \ if\ k=0 \\ -g_{k+1} + \beta_k d_{k,} & if\ k \ge 1 \end{cases} \qquad (1.3)$$

where g_{k+1} is the gradient of *f(x)* at the point x_{k+1} . $\beta_k \in R$ is known as conjugate gradient coefficient and different β_k will yield different CG methods . Some well-known formulas are given as follows :

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \qquad (1.4)$$

$$\beta_k^{HS} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{d_k^T (g_{k+1} - g_k)} \qquad (1.5)$$

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)} \qquad (1.6)$$

$$\beta_k^{PR} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{\|g_k\|^2} \qquad (1.7)$$

Where g_k is the gradient of f(x) at the point x_k . In this paper , FR denotes Fletcher and Reeves [9] , HS denotes Hestenes and Steifel [12] , DY denotes Dai andYuan [6] and lastly PR denotes Polak and Ribi ere [15] .We denotes norm of vectors as $\|.\|$. It also shows that for f (x) that is strictly convex quadratic function ,all these methods are equivalent , but for ge neral non quadratic, their behavior is quite different. [5] , [23] . The most studied properties of CG are its global convergence properties . Zoutendijk [24] pro ved the global convergence of FR method. Al-Baali

[1] , Touati - Ahmed and Storey [20] , Gilbert and Nocedal [10] has further analyzed the global conver

gence of algorithms related to the FR method with strong Wolfe condition. Powell [16] also proved that FR is a superior method compared to others . For further reading and recent finding of CG methods refer to Sun and Zhang [19] , Birgin and Matrtinez [4] , Dai and Yuan [7] ,Yuan and Wei [22] , Andrei [3] and Shi and Gao [18] . A basic key factor of global convergence is selecting the step size α_k . The most common search is to do the exact line.

$$f(x_k + \alpha_k d_k) = \min_{\alpha \ge 0} f(x_k + \alpha d_k) \qquad (1.8)$$

In this paper, the proposed method is solved by using the exact line search. The remaining of study is organized as follows. Section 2. Presented the new algorith

m. The global convergence of the new method is pro ved by using the exact line search for non convex fun nction in Section 3. Some interesting numerical resu lts we get it by comparing the new method with oth er CG methods presented in Section 4 . Finally ,Secti on 5. Presented conclusions .

2. NEW DAI AND YUAN ALGORITHM
In this section we propose new algorithm based on the original DY algorithm , we named it MDY . The new formula for the denominator has been proposed , the formula for numerator as the Dai and Yuan formu la has been retained [17] . then

$$\beta_k^{MDY} = \frac{\|g_{k+1}\|^2}{d_k^T (d_k - g_{k+1})} \qquad (2.1)$$

Here the algorithm of MDY is shown as follows :

MDY Algorithm

Step 1 : Initialize, select $x_0 \in R^n$, $\varepsilon > 0$, k=0

Step 2 : Compute β_k from Eq. (2.1)

Step 3 : Compute search direction

$$d_{k+1} = \begin{cases} -g_{k+1} & , \ if \ k = 0 \\ -g_{k+1} + \beta_k d_k, & if \ k \geq 1 \end{cases}$$

If $\|g_{k+1}\| = 0$, then terminate, else continue

Step 4 : Compute step size

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

Step 5 : Update new point

Step 6 : Convergence test and stopping criteria If

$f(x_{k+1}) < f(x_k)$ and If $\|g_{k+1}\| < \varepsilon$,

then terminate else 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 of convergence for the general CG methods.

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

$$g_k^T d_k < 0 \qquad \forall k \geq 1 \qquad (3.1)$$

We now make the following basic assumption on the objective function .

3.1 Assumption

(1) *f(x)* is bounded below on the level set

$$l = \{x | f(x) \leq f(x_1)\}$$

where x_1 is the starting point.

(2) In some neighborhood N of l , $f(x)$ is continuously differentiable and its gradient is Lipschitz continuous ; then , there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\| \qquad (3.2)$$

$\forall x, y \in N$ .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 po sitive step size ($\alpha_k > 0$) such that

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k \qquad (3.3)$$

$$g(x_k + \alpha_k d_k)^T d_k \leq -\sigma g_k^T d_k \qquad (3.4)$$

Where $0 < \delta < \sigma < 1$ .To prove global convergen ce for the FR method , we used the strong Wolfe line search which requires α_k to satisfy (3.3) and

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

The following important result was obtained by Zoute ndijk [24] and Wolfe [21] .

3.2 Lemma

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

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

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

3.3 Theorem

Consider that the Assumption is true for any CG method of the form (1.2),(1.3), with d_k satisfying (3.1) and with strong Wolfe line search (3.3) and (3.4) , hence either

$$\lim_{k \to \infty} \|g_k\| = 0 \qquad (3.7)$$

Or

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

The following corollary is based on the Theorem .

3.4 Corollary

Consider that the Assumption is true for any CG met hod 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

$$\sum_{k=1}^{\infty} \frac{(g_k)^t}{\|d_k\|^2} = +\infty \qquad (3.9)$$

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

Proof

In order to proof Corollary, we use contradiction . If (3.7) is not true , it follows from the theorem that

$$\sum_{k=1}^{\infty} \frac{(g_k)^4}{\|d_k\|^2} < +\infty \qquad (3.10)$$

Because $\|g_k\|$ is bounded away from zero and $t \in [0,4]$, it is easy to see that (3.10) Contradicts (3.9).This shows that the Corollary is true .Finally , if a conjugate gradient method fails to converge , one can easily see that the length of the search direction will converge to infinity [14] .

4. NUMERICAL RESULTS

This section presents the performance of FORTRAN implementation to the algorithm MDY by using a set of well-known unconstrained optimization test functio ns , for each function we have considered numerical experiments with the number of variables n = 100 , 1000 , 10000 and 100000 .We compared the perform ance of the algorithm MDY with four famous formule s FR , HS , DY and PR ,which they defined in (1.4) – (1.7) . All these algorithms are implemented with the standard Wolfe line search conditions with $\delta = 0.001$ and $\sigma = 0.5$, the stopping condition defined by $\|g_{k+1}\| \le 1 \times 10^{-6}$. The comparison includes the following:

NOI : number of iterations .

NOF : number of function evaluations .

From table (I) we see that for more problems the new algorithm is really much better than other CG algorith ms especially for high dimensions . The comparison is based on number of iterations and number of functi on evaluations , for solving (13) problems .

Note that the symbol * in table (I) means that the algorithm is fail to converge.

Table(I) Comparison of algorithms with respect to (NOI , NOF) for different dimensions (n=100,1000,10000,100000)

Test functions	N	FR NOI-NOF	HS NOI-NOF	DY NOI-NOF	PR NOI-NOF	MDY NOI-NOF
NOND	100	30-78	28-67	28-68	30-78	26-64
	1000	30-78	28-67	28-68	30-78	26-64
	10000	30-78	28-67	28-68	30-78	26-64
	100000	30 -78	28-67	28-68	30-78	26-64
POWELL	100	31-92	41-109	56-169	50-136	43-125
	1000	36-110	41-109	56-169	62-203	44-127
	10000	36-110	41-109	63-208	68-242	48-139
	100000	39-131	47-133	63-184	72-279	53-159
WOOD	100	29-66	33-73	28-68	29-67	26-63
	1000	29-66	33-73	2868	29-67	26-63
	10000	30-68	34-75	28-68	30-69	26-63
	100000	30-68	35-77	28-68	30-69	26-63
ROSEN	100	30-76	30-76	30-76	30-76	30-76
	1000	30-76	30-76	30-76	30-76	30-76
	10000	30-76	30-76	30-76	30-76	30-76
	100000	30-76	30-76	30-76	30-76	30-76
DIAGONAL-2	100	54-207	62-225	54-205	62-225	54–210
	1000	157-647	180-685	157-647	180-725	150-620
	10000	*	*	462-1927	*	467-1910
	100000	*	*	*	*	*
WOLFE	100	51-103	57-115	51-103	57-115	51–103
	1000	59-119	79-159	59-119	57-115	59-119
	10000	120-244	105-214	126-254	140-282	120-240
	100000	134-276	114-232	128-260	119 - 243	128–265

ENGVAL1	100	21-44	21-44	21-46	22- 46	21–43
	1000	21-86	23-1188	19-42	22-51	22-50
	10000	23-124	26-230	*	32-1959	30-275
	100000	*	*	*	*	*
EX.WOOD	100	27-61	29-65	26-60	29-67	26–57
	1000	27-61	30-67	26-60	29-67	26–57
	10000	29-66	33-73	26-60	29-67	26–57
	100000	29-66	33-73	27-62	29-67	27–60
DIXMAANB	100	5-13	5-13	5-13	5-13	5-13
	1000	5-13	5-13	5-13	5-13	5-13
	10000	6-16	6-16	6-16	6-16	6-16
	100000	6-16	6-16	6–16	6-16	6-16
DIXMAANC	100	5-15	5-15	5-15	5-15	5-15
	1000	5-15	5-15	5-15	5-15	5-15
	10000	5-15	5-15	5-15	5-15	5-15
	100000	5-15	5-15	5–15	5-15	5-15
SHALLOW	100	10-25	10-25	10-25	10-25	10-25
	1000	10-25	10-25	10-25	10-25	10-25
	10000	10-25	10-25	10-25	10-25	10-25
	100000	11-27	11-27	11–27	11-27	11-27
EX.BDI	100	20-40	22-46	19-40	23-48	19–40
	1000	22-46	23-48	21-44	24-50	21–44
	10000	23-48	26-54	22-46	24-50	22–46
	100000	25-52	28-58	24–50	26-54	24–50
DENSCHNF	100	24-51	*	21-44	*	19–42
	1000	24-51	*	21-45	*	21-45
	10000	26-55	*	23-49	*	23-49
	100000	26-55	*	23-49	*	23–49

5. CONCLUSION

In this paper we have proposed a new and simple β_k based on the proven Dai and Yuan method . The comparison results for new method with four famous methods FR,HS,DY and PR for n=100, 1000 , 10000 and 100000 is more effective and efficient than those methods , also numerical results suggested that this new method converge globally.

Further work , we should study this new method for neural network training . Moreover , more numerical experiments for large practical problems should be done.

6. 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 Collection**.** *Advanced Modeling 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.* 230,70-582

[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 Con-jugate 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 Properties of Nonlinear Conjugate Gradient Method. *SIAM J.Optim.,* 10, 348-358.

[9] Fletcher,R. and Reeves, C.(1964**).** Function Minimization by 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] Mustafa Mamat , Mohd Rivaie , Islam Mohdand Muhammad Fauzi.(2010).A New Conjugate Gradient Coefficient for Unconstrained Optimization. *Int.J.Contemp .Math . Sciences ,5(29),1429-1437.*

[15] Polak , E. and , Ribiere, G. (1969**).** Note Sur La Convergence De Directions Conjugees.*Rev. Francaise Informat Recherche Operationelle,3EAnnee*(16), 35-43.

[16] Powell, M.J.D. (1986**).** Convergence Properties of Algorithm for Nonlinear Optimization. *SIAM Review.,* 28(4), 487-500.

[17] Rivaie.M , Fauzi. M, Mamat .M and Mohd .I .(2011). Modified Hestenes-Steifel Method forfor Unconstrained Optimization.*J.Appl Sciences,11(8): DOI:10.3923/jas.2011.1461-1464*

[18] Shi, Z.J. and Gao, J. (2009). A New Family of,Conjugate Gradient Methods**.** *J,Comput.Appl. Math.,* 224, 444-457.

[19] Sun, J. and Zhang, J. (2001**).** Global Convergence of Conjugate Gradient Methods without Line Search. *Annals. Opr. Rch,* 103, 161-173.

[20] Touati-Ahmed, D.and Storey, C .(1990**).** Efficient Hybrid Conjugate Gradient Techniques, *J. Optim.Theory Appl.,* 64, 379-397.

[21] Wolfe, P.(1969) . Convergence Conditions for Ascent Method. *SIAM Rev.,*11,226-235.

[22] Yuan, G and Wei, Z. (2009**).** New Line Search Methods for Unconstrained Optimization . *J.Korean Stat.Soc.,* 38, 29-39.

[23] Yuan, Y.and Sun,W. (1999). Theory and Methods of optimization. Science Press of China,Beijing.

[24] Zoutendijk, G. (1970).Nonlinear ProgrammingComputational Methods. *In : Abadie J.(Ed.)Integer and nonlinear programming,* 37-86.

Received: November, 2009