New Parameter of CG-Method for Unconstrained Optimization

Hamsa Th. Chilmeran — *hamsathrot@uomosul.edu.iq* — *Department of Mathematics, College of Computer Science and Mathematics*

Satia A. Al-Taie — *altaiesatia@uomosul.edu.iq* — *Computer Center*

Enaam Gh. Saeed — *enaamghanim@uomosul.edu.iq* — *Department of Mathematics, College of Computer Science and Mathematics*

University of Mosul, Mosul, Iraq

Received on: 02/10/2018 **Accepted on: 28/03/2019**

ABSTRACT

In this paper, we derived a new parameter by equating the modified QN direction which is suggested by [7] and the standard CG method which satisfied the sufficient descent condition and the global convergence under some assumptions. The numerical experiment of the new algorithm perform better than previous standard algorithm depending on the number of calling the function (NOF) and number of iterations (NOI).

Keywords: CG-method, QN-direction, global convergence.

متغير جديد لخوارزمية التدرج المترافق في الأمثلية غير المقيدة

همسة ثروت سعيد جلميران — *كلية علوم الحاسوب والرياضيات*

ساطعة عبد الله الطائي — *مركز الحاسبة الالكترونية*

انعام غانم سعيد — *كلية علوم الحاسوب والرياضيات*

جامعة الموصل، الموصل، العراق

تاريخ استلام البحث: 2018/10/02 **تاريخ قبول البحث: 2019/03/28**

الملخص

في هذا البحث تم اشتقاق معلمة جديدة وذلك بمساواات متجه شبيه نيوتن المحسن المقترح في المصدر [7] ومتجه التدرج المترافق والتي حققت شرط الانحدار الكافي والتقارب الشامل تحت بعض الفرضيات. النتائج العددية للطريقة الجديدة كانت أفضل من طريقة التدرج المترافق بالاعتماد على عدد التكرارات وعدد حسابات الدالة.

الكلمات المفتاحية: طريقة التدرج المترافق، اتجاه شبيه نيوتن، تقارب شامل.

1 . Introduction

The conjugate gradient algorithm represents an operatively idea to find minimum optimization problem with the form:

Min $f(x), x \in R^n$ (1)

Where $f : R^n \to R$ is continuously differentiable function with available gradient.

CG algorithm needs the information of first derivative, where $g(x)$ denotes the gradient and x_0 be the first approximate solution of (1). A sequence of solutions will be generated when a standard CG algorithm is used to solve (1), in which

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

Such that α_k is the step length, selected to minimize $f(x_k)$ through d_k, its computation sometimes based on the weak Wolfe-conditions:

$$f(x_k + \alpha_k d_k) - f(x_k) \le \tilde{\delta} \alpha_k g_k^T d_k \quad (3)$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k \quad (4)$$

and the strong Wolfe-conditions:[8]

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

) 6($$\left|g(x_k+\alpha_k d_k)^T d_k\right|\le-\sigma g_k^T d_k$$

Where $\tilde{\delta}\in(0,1)$ and $\sigma\in(\tilde{\delta},\frac{1}{2})$.

The search direction (d_k) defined by:

$$d_{k+1}=\begin{cases}-g_{k+1} & \text{If } k=0\\ -g_{k+1}+\beta_k d_k & \text{If } k>0\end{cases} \tag{7}$$

Where β_k is called conjugacy parameter, Some formulas for (β_k) are given by:

$$\beta_k^{HS}=\frac{g_{k+1}^T y_k}{d_k^T y_k}$$ (Hestenes - Stiefel, (1952)), [5] (8)

$$\beta_k^{FR}=\frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$$ (Fletcher - Reeves (FR), (1964)), [3] (9)

$$\beta_k^{CD}=\frac{g_{k+1}^T g_{k+1}}{-d_k^T g_k}$$ (Fletcher (CD) ,(1987)), [4] (10)

$$\beta_k^{LS}=\frac{g_{k+1}^T y_k}{-d_k^T g_k}$$ (Liu-Storey (LS), (1991)), [6] (11)

$$\beta_k^{DY}=\frac{g_{k+1}^T g_{k+1}}{d_k^T y_k}$$ (Dai-Yun (DY) (1999)) ,[3] (12)

$$\beta_k^{N}=\frac{g_{k+1}^T y_k}{d_k^T y_k}-2\frac{\|y_k\|^2}{(d_k^T y_k)^2}g_{k+1}^T d_k$$ (Andrie.N.(2008)), [1] (13)

where $y_k=g_{k+1}-g_k$. (14)

M. Mamat and I. Mohd In [7] derived a Hybrid Modified BFGS Algorithms B_{k+1} defined by

$$d_{k+1}=-B_{k+1}^{-1}g_{k+1}-\lambda g_{k+1} \tag{15}$$

Where λ is a constant, $\lambda\in(0,1)$

And $$B_{k+1}=B_k+\frac{y_k^T y_k}{y_k^T s_k}-\frac{B_k s_k^T s_k B_k^T}{s_k^T B_k s_k},$$

B_{k+1} is the approximation of the Hessian $H_{i,j}=\frac{\partial f}{\partial x_i \partial x_j}$, satisfies

$$B_{k+1}s_k=y_k \tag{16}$$

2. The new parameter (β_k):

In this paper, we derive a new parameter β_k depending on the two directions, the first direction defined by (7) and the second defined by (15)

equating (7) and (15) we get:

$$-g_{k+1}+\beta_k s_k=-B_{k+1}^{-1}g_{k+1}-\lambda g_{k+1} \tag{17}$$

Multiplying (17) with $s_k^T B_{k+1}$

$$-s_k^T B_{k+1} g_{k+1} + s_k^T B_{k+1} \beta_k s_k = -s_k^T g_{k+1} - \lambda s_k^T B_{k+1} g_{k+1} \qquad (18)$$

Substitute (16) in (18), we get

$$\therefore -y_k^T g_{k+1} + y_k^T \beta_k s_k = -s_k^T g_{k+1} - \lambda y_k^T g_{k+1}$$

And the new parameter β_k defined as:

$$\beta_k = \frac{-s_k^T g_{k+1} + (1-\lambda) y_k^T g_{k+1}}{s_k^T y_k},$$

If ELS and λ=0, scalar β_k reduce to the standard scalar β_k^{HS}

$$d_{k+1} = -g_{k+1} + \frac{-s_k^T g_{k+1} + (1-\lambda) y_k^T g_{k+1}}{s_k^T y_k} s_k \qquad (19)$$

3. The Algorithm of the New Modified Parameter s

Step 1: given $x_0 \in R^n$, Set $k = 0$.

Step 2 : let $d_0 = -g_0$

Step 3 : Find the step length $\alpha_k > 0$, satisfying (3) , (4), and set $x_{k+1} = x_k + \alpha_k d_k$.

Step 4 : if $\|g_k\| \le 10^{-5}$, then stop

Step 5 : Otherwise, compute the direction using (19)

Step 6 : if $k = n$, or $\frac{|g_k^T g_{k+1}|}{\|g_{k+1}\|^2} \ge 0.2$ [9], then go to step 2 else set $k = k+1$, go to step3.

4. The sufficient descent condition for the new direction

Theorem (1) :The new direction d_k which is given in (19) satisfies the sufficient descent condition $g_{k+1}^T d_{k+1} < 0$ for all k .

For the prove, , it is necessary to assume the following basic assumption:

Assumption 1:The objective function f in equation (1) is differentiable with Lipschitz continuous gradient on level set see [2]

Proof the theorem :

Multiplying (19) by $\frac{g_{k+1}}{\|g_{k+1}\|^2}$, we get:

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \frac{-s_k^T g_{k+1} + (1-\lambda) y_k^T g_{k+1}}{s_k^T y_k} . \frac{s_k^T g_{k+1}}{\|g_{k+1}\|^2} \qquad (20)$$

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} + 1 = \frac{-s_k^T g_{k+1} + (1-\lambda) y_k^T g_{k+1}}{s_k^T y_k} . \frac{s_k^T g_{k+1}}{\|g_{k+1}\|^2} \qquad (21)$$

Since from weak Wolf condition (4) we have:

$$-s_k^T g_{k+1} \le -\sigma d_k^T g_k \tag{22}$$

Substitute (22) in (21) we have:

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} + 1 \le \frac{-\sigma d_k^T g + (1-\lambda) y_k^T g_{k+1}}{s_k^T y_k} \cdot \frac{s_k^T g_{k+1}}{\|g_{k+1}\|^2}$$

From the Norm definition:

$g_{k+1}^T y_k \le \|g_{k+1}\| \|y_k\|$ and since :

$$s_k^T g_{k+1} \le s_k^T y_k, \tag{23}$$

we have

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} + 1 \le \frac{-\sigma d_k^T g_k + (1-\lambda)\|g_{k+1}\|.\|y_k\|}{s_k^T y_k} \cdot \frac{s_k^T y_k}{\|g_{k+1}\|^2}$$

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} + 1 \le \frac{-\sigma d_k^T g_k + (1-\lambda)\|g_{k+1}\|.\|y_k\|}{\|g_{k+1}\|^2}$$

Since $d_k^T g_k \le \dfrac{-d_k^T y_k}{(\sigma+1)}$ (24)

$$\therefore \frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} + 1 \le \frac{\sigma \frac{d_k^T y_k}{\sigma+1} + (1-\lambda)\|g_{k+1}\|.\|y_k\|}{\|g_{k+1}\|^2} \le \frac{\sigma \frac{L\|s_k\|^2}{\sigma+1} + (1-\lambda)\|g_{k+1}\|.\|y_k\|}{\|g_{k+1}\|^2}$$

Let $C = \dfrac{\sigma \frac{L\|s_k\|^2}{\sigma+1} + (1-\lambda)\|g_{k+1}\|.\|y_k\|}{\|g_{k+1}\|^2}$ (25)

$$\therefore d_{k+1} g_{k+1} \le (-1+c)\|g_{k+1}\|^2 \le -(1-c)\|g_{k+1}\|^2$$

$$\therefore d_{k+1} g_{k+1} \le -(1-c)\|g_{k+1}\|^2 \tag{26}$$

The proof is complete □

5. The Global Convergence of the New Proposed Parameter for CG-Method

Theorem (2). Let the assumption (A) hold and the conjugate gradient methods (2) and (7) hold, d_k is a descent direction where λ_k is satisfied (5) and (6) [8]

if $$\sum_{k\ge 1} \frac{1}{\|d_{k+1}\|} = \infty \tag{27}$$

$$\therefore \liminf_{k\to\infty} \|g_k\| = 0 \tag{28}$$

Theorem: let that the assumption (A) holds .consider the algorithm in (2), d_k is defined by (19) , then the new algorithm either stops at stationary point i.e $\|g_k\| = 0$ or

$$\liminf_{k\to\infty} \|g_k\| = 0$$

Proof:

$$d_{k+1} = -g_{k+1} + \frac{-s_k^T g_{k+1} + (1-\lambda) y_k^T g_{k+1}}{s_k^T y_k} s_k$$

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

$$\|d_{k+1}\| \le \|g_{k+1}\| + \left| \frac{-s_k^T g_{k+1} + (1-\lambda) y_k^T g_{k+1}}{s_k^T y_k} s_k \right|$$

$$\|d_{k+1}\| \le \|g_{k+1}\| + \left| \frac{-s_k^T g_{k+1}}{s_k^T y_k} \right| + \left| \frac{(1-\lambda) y_k^T g_{k+1}}{s_k^T y_k} \right| \|s_k\|$$

$$\|d_{k+1}\| \le \|g_{k+1}\| + \left| \frac{-s_k^T g_{k+1}}{s_k^T y_k} \right| + (1-\lambda) \left| \frac{y_k^T g_{k+1}}{s_k^T y_k} \right| . \|s_k\| \qquad (29)$$

Since the new direction is descent and from convex condition and Lipchitz condition we have

$$u\|s\|^2 \le s_k^T y_k \le L\|s\|^2 \qquad (30)$$

$$\|d_{k+1}\| \le \|g_{k+1}\| + \frac{\sigma \|g_k\|^2}{u\|s_k\|^2} + (1-\lambda) \left| \frac{\|g_{k+1}\| . \|y_k\|}{u\|s_k\|^2} \right| . \|s_k\|$$

Let

$$\tilde{\gamma} = \|g_{k+1}\| + \frac{\sigma \|g_k\|^2}{u\|s_k\|^2} + (1-\lambda) \left| \frac{\|g_{k+1}\| . \|y_k\|}{u\|s_k\|^2} \right| . \|s_k\| \ge 0$$

Then

$$\|d_{k+1}\| \le \tilde{\gamma} \qquad (31)$$

$$\sum_{k\ge1} \frac{1}{\|d_{k+1}\|} \ge \sum_{k\ge1} \frac{1}{\gamma} . \sum_{k\ge1} 1 = \infty$$

$$\liminf_{k\to\infty} \|g_k\| = 0$$

The proof is complete □

6. The Numerical Result

In this part we use "Visual Fortran language" to implement the new algorithm defined in (19) for presenting the computational performance of (27) unconstraint test functions in [1], All algorithms achieve the strong Wolf line search conditions and $\|g_k\| \le 10^{-5}$.

In the early conjugate gradient algorithms, the restart strategy was usually restarted whenever $k = n$ or $k = n+1$. When n is very large, another efficient restart is called Wolfe restart, which is defined by :

$\left| g_{k+1}^T g_k \right| \ge 0.2 \|g_{k+1}\|^2$ [9], using this strategy, the new algorithm in the table below performs better than the standard algorithm for most of the problems, depending on the

Number of iteration(NOI), and depending on the Number of functions (NOf) with different dimensions (N).

The Numerical Results for the new Algorithm

NO	Test Fun.	Dim.	Standard Histen\Steven algorithm		Modified algorithm	
			NOI	NOF	NOI	NOF
1	Arwhead (Cute)	1000	17	122	13	66
2	Broyden Tridiagonal	1000	33	58	35	60
3	Diagonal 2	1000	177	300	263	439
4	DIXMAAN A (Cute)	1000	6	12	6	12
5	DIXMAAN E (Cute)	1000	259	413	210	324
6	DIXMAAN F (Cute)	1000	232	447	204	366
7	DIXMAAN G (Cute)	1000	228	371	213	348
8	DIXMAAN H (Cute)	1000	244	445	224	387
9	DIXMAAN J (Cute)	1000	227	444	262	458
10	EDENSCH (Cute)	1000	39	353	47	515
11	Extended Cliff	1000	18	19	18	19
12	Extended Hiebert	1000	79	171	78	188
13	Extended Himmelblou	1000	10	19	17	28
14	Extended Powel BD1	1000	52	83	14	27
15	Extended psc1	1000	6	13	7	15
16	Extended Three Expo Terms	1000	12	19	11	19
17	Extended Tridiagonal 1	1000	34	53	41	62
18	Generalized psc1	1000	1141	1362	457	793
19	Generalized Tridiagonal 2	1000	64	98	49	74
20	NONDia (Cute)	1000	12	24	13	24
21	Penalty	1000	2	5	2	5
22	Raydan2	1000	4	9	4	9
23	Rosen Brock	1000	34	75	35	69
24	Tridiagonal 4	1000	4	8	10	20
25	Tridiagonal 5	1000	4	9	4	9
26	Trigonometric	1000	29	52	27	48
27	White & Holst	1000	26	53	32	64
	Total		**2993**	**5037**	**2296**	**4448**

NO	Test Fun.	Dim.	Standard Histen\Steven algorithm		Modified algorithm	
			NOI	NOF	NOI	NOF
1	Arwhead (Cute)	5000	10	21	8	18
2	Broyden Tridiagonal	5000	77	118	78	121
3	Diagonal 2	5000	509	826	458	768
4	DIXMAAN A (Cute)	5000	6	12	6	12
5	DIXMAAN E (Cute)	5000	484	763	562	887
6	DIXMAAN F (Cute)	5000	534	1055	545	1074
7	DIXMAAN G (Cute)	5000	703	1191	601	1022
8	DIXMAAN H (Cute)	5000	479	793	645	1015
9	DIXMAAN J (Cute)	5000	593	1211	497	987
10	EDENSCH (Cute)	5000	91	2010	62	1200
11	Extended Cliff	5000	10	12	10	12
12	Extended Hiebert	5000	77	163	81	190
13	Extended Himmelblou	5000	19	31	17	28
14	Extended Powel BD1	5000	55	87	55	87
15	Extended psc1	5000	7	15	8	17
16	Extended Three Expo T	5000	14	22	9	16
17	Extended Tridiagonal 1	5000	80	1642	60	957
18	Generalized psc1	5000	230	560	671	992
19	Generalized Tridiagonal 2	5000	60	98	63	95
20	NONDia (Cute)	5000	7	14	8	15
21	Penalty	5000	2	5	2	5
22	Raydan2	5000	4	9	4	9
23	Rosen Brock	5000	35	82	31	62
24	Tridiagonal 4	5000	4	8	10	20
25	Tridiagonal 5	5000	4	9	4	9
26	Trigonometric	5000	19	39	19	39
27	White & Holst	5000	26	53	42	85
	Total		**4139**	**10849**	**4556**	**9742**

NO	Test Fun.	Dim.	Standard Histen\Steven algorithm		Modified algorithm	
			NOI	NOF	NOI	NOF
1	Arwhead (Cute)	10000	4	10	21	264
2	Broyden Tridiagonal	10000	77	123	87	134
3	Diagonal 2	10000	692	1123	704	1164
4	DIXMAAN A (Cute)	10000	6	12	6	12
5	DIXMAAN E (Cute)	10000	716	1122	614	953
6	DIXMAAN F (Cute)	10000	1095	2008	829	1334
7	DIXMAAN G (Cute)	10000	948	1591	858	1499
8	DIXMAAN H (Cute)	10000	1438	2693	917	740
9	DIXMAAN J (Cute)	10000	1072	2147	958	1573
10	EDENSCH (Cute)	10000	53	866	79	1779
11	Extended Cliff	10000	11	13	11	13
12	Extended Hiebert	10000	77	165	78	186
13	Extended Himmelblou	10000	19	31	17	28
14	Extended Powel BD1	10000	35	59	39	67
15	Extended psc1	10000	7	15	8	19
16	Extended Three Expo Terms	10000	16	26	10	18
17	Extended Tridiagonal 1	10000	36	74	108	2500
18	Generalized psc1	10000	562	1105	637	2010
19	Generalized Tridiagonal 2	10000	54	92	59	93
20	NONDia (Cute)	10000	4	7	4	7
21	Penalty	10000	2	5	2	5
22	Raydan2	10000	4	9	4	9
23	Rosen Brock	10000	35	83	31	62
24	Tridiagonal 4	10000	4	8	10	20
25	Tridiagonal 5	10000	4	9	4	9
26	Trigonometric	10000	17	34	16	32
27	White & Holst	10000	34	74	42	85
	Total		**2270**	**13504**	**6153**	**14615**

The performance's percentage of the modified algorithm depending on the standard H\S algorithm can be summarized in the table below:

	The total of NOI and NOF for the Standard H\S algorithm		The total of NOI and NOF for the modified algorithm		The percentage of equations that get better	
DIM	NOI	NOF	NOI	NOF	NOI	NOF
1000	2993	5037	2296	4448	56%	63%
5000	4139	10849	4556	9742	59%	63%
10000	7022	13504	6153	14615	56%	56%

REFERENCES

[1] Andrie N. (2008), "An Unconstraint Optimization test functions collection", Adv. Model. Optimization. Vol(10), pp:147-161.

[2] Dai,Y. And Yuan, Y. (1999), "A Nonlinear Conjugate Gradient Method With A Strong Global Convergence Property",Siam J. Optim.,Vol(10), pp:177– 182.

[3] Fletcher, R., And Reeves, C., (1964) ," Function Minimization By Conjugate Gradients", Comput. J. Vol(7) ,149–154.

[4] Fletcher, R.: Practical Methods of Optimization, vol. I: Unconstrained Optimization. Wiley, New York (1987).

[5] Hestenes, M.R And Stiefel, E. , (1952),"Method Of Conjugate Gradient For Solving Linear Equations", Journal Of Research Of The National Bureau Of Standards , Vol.(5),No.(49), 409-436.

[6] Liu, Y. And Storye, C., (1991) ," Efficient Generalized Conjugate Gradient Algorithms",Part1:Theory,J.Optimization Theory. Appl. Vol.(69),pp:129–137.

[7] M. Mamat and I. Mohd, (2013), "A New Scaled Hybrid Modified BFGS Algorithms for Unconstrained Optimization", Vol.7,No. 6,263-270.

[8] Nicolai Andrei,(2010), "New accelerated conjugate gradient algorithms as modification of Die-Yuan's computational scheme for unconstrained optimization", journal of computational and applied mathematics 234,3397-3410.

[9] Powel, M.J. D., (1976a),"Restart Procedure For The Conjugate Gradient Method, Mathematical programing", Vol.(1),pp.241-254.