Indonesian Journal of Electrical Engineering and Computer Science
Vol. 17, No. 1, January 2020, pp. 404~411
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v17.i1.pp404-411

A new kind of parameter conjugate gradient for unconstrained optimization

Basim A. Hassan[1], Hussein O. Dahawi[2], Azzam S. Younus[3]
[1]Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, Iraq
[2,3]Department of Mathematics, College of Education of Pure Sciences, University of Mosul, Iraq

Article Info

Article history:

Received Mar 4, 2019
Revised Jun 2, 2019
Accepted Jul 12, 2019

Keywords:

Conjugate gradient methods
Convergence property
Sufficient descent property

ABSTRACT

The key feature for conjugate gradient methods is a conjugate parameter optimal for solving unrestrained minimization functions. In this paper, a replacement new parameter conjugate gradient for unconstrained optimization. The sufficient descent property cleave to. The global convergence property of the new method is proved under some assumptions. Numerical results explain that the new parameter is superior in practice.

Corresponding Author:

Basim A. Hassan
Department of Mathematics,
College of Computers Sciences and Mathematics,
University of Mosul, Iraq
Email: basimabas39@gmail.com

1. INTRODUCTION

In the literature several optimization strategies may be originate with (theoretically) a much better speed of convergence than the descent gradient methods. Maybe the foremost documented ones area unit the conjugate gradient and quasi-Newton strategies. For details see [1].

Generally, for n number of variables of the problem has the following from:

$$\min \left\{ f(x) \;\middle|\; x \in R^n \right\} \tag{1}$$

where $f : R^n \to R^1$ is a continuously derivable function.

Nonlinear conjugate gradient algorithms are based on the following iterative scheme :

$$x_0 \in R^n, \quad x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

where the search direction d_{k+1} is outlined as a linear combination of the present by product η_{k+1} and also the earlier search direction d_k :

$$d_0 = -\eta_0, \quad d_{k+1} = -\eta_{k+1} + \beta_k d_k \tag{3}$$

where β_k is a parameter conjugate gradient, η_{k+1} denotes gradient of $f(x_{k+1})$ at the point x_{k+1}, $s_k = x_{k+1} - x_k$ and $y_k = \eta_{k+1} - \eta_k$. More details can be found in [2].

The step size α_k is decided in line with the Wolfe line search states as follows :

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k \eta_k^T d_k \quad (4)$$

$$\text{“ } \eta(x_k + \alpha_k d_k)^T d_k \geq \sigma \eta_k^T d_k \text{ ”} \quad (5)$$

where $0 < \delta < \sigma < 1$ and d_k is a descent direction $\eta_k^T d_k < 0$. For details see [3].

It is well known that if the matrix of gradient is positive definite, the most efficient search direction at x_k is the Newton direction:

$$d_{k+1} = -(\nabla \eta_{k+1})^{-1} \eta_{k+1} = -G_{k+1}^{-1} \eta_{k+1} \quad (6)$$

From the secant condition that :

$$(\nabla \eta_{k+1})^T s_k = y_k \quad (7)$$

More details can be found in [4].

The conjugate gradient methods different depend on the calculation of parameters β_k. The idea of variant CG methods had been studied by many researchers for example, see (Hestenes and Stiefel [5]) and (Fletcher and Reeves [6]).

$$\beta_k^{HS} = \frac{\eta_{k+1}^T y_k}{y_k^T d_k}, \quad \beta_k^{FR} = \frac{\eta_{k+1}^T \eta_{k+1}}{\eta_k^T \eta_k} \quad (8)$$

The motivation of this paper is to combine the advantages of conjugate gradient direction d_{k+1}^{CG} and Newton direction d_{k+1}^{N} in order to provide novel parameter with better convergence.

2. A NEW KING OF PARAMETER CONJUGATE GRADIENT

In this section, we derive a new parameter conjugate gradient based on the three order tensor model. Based on the three order tensor model, the information of the second order curvature in the following from :

$$s_k^T G_{k+1} s_k = y_k^T s_k + 6(f_k - f_{k+1}) + 3(\eta_{k+1} + \eta_k)^T s_k \quad (9)$$

For more details can be found in [7].

The step size α_k determine by many algorithms, in exact line search the step length α_k is choose as :

$$\alpha_k = -\frac{\eta_k^T d_k}{d_k^T G d_k} \quad (10)$$

We produce the steps, that lead to a new second order curvature as below :

$$s_k^T G_{k+1} s_k = (f_k - f_{k+1}) + \frac{2}{3} \eta_{k+1}^T s_k - \frac{3}{6} \eta_k^T s_k \quad (11)$$

which implies that :

$$G_{k+1} = \frac{(f_k - f_{k+1}) + 2/3\eta_{k+1}^T s_k - 1/2\eta_k^T s_k}{s_k^T s_k} I_{n\times n} \tag{12}$$

Since $y_k^T s_k = \eta_{k+1}^T s_k - \eta_k^T s_k$, then from the above equation, we have:

$$G_{k+1} = \frac{(f_k - f_{k+1}) + 1/2 y_k^T s_k + 1/6\eta_{k+1}^T s_k}{s_k^T s_k} I_{n\times n} \tag{13}$$

Then Newton direction can be written as :

$$d_{k+1} = -\left(\frac{s_k^T s_k}{(f_k - f_{k+1}) + 1/2 y_k^T s_k + 1/6\eta_{k+1}^T s_k}\right)\eta_{k+1} \tag{14}$$

By combine the advantages of d_{k+1}^{CG} and d_{k+1}^{N} , so, the equation is hold :

$$-(\nabla \eta_{k+1})^{-1}\eta_{k+1} = -\eta_{k+1} + \beta_k d_k \tag{15}$$

Now, we'll realize the parameter β_k .Equation (15) multiplied by y_k^T , then we get :

$$-\left(\frac{s_k^T s_k}{(f_k - f_{k+1}) + 1/2 y_k^T s_k + 1/6\eta_{k+1}^T s_k}\right)\eta_{k+1}^T y_k = -\eta_{k+1}^T y_k + \beta_k d_k^T y_k$$

$$\beta_k d_k^T y_k = -\left(\frac{s_k^T s_k}{(f_k - f_{k+1}) + 1/2 y_k^T s_k + 1/6\eta_{k+1}^T s_{k\,k}}\right)\eta_{k+1}^T y_k + \eta_{k+1}^T y_k \tag{16}$$

from (16) we get :

$$\beta_k d_k^T y_k = \left(1 - \frac{s_k^T s_k}{(f_k - f_{k+1}) + 1/2 y_k^T s_k + 1/6\eta_{k+1}^T s_k}\right)\eta_{k+1}^T y_k$$

then we have :

$$\beta_k = \left(1 - \frac{s_k^T s_k}{(f_k - f_{k+1}) + 1/2 y_k^T s_k + 1/6\eta_{k+1}^T s_k}\right)\frac{\eta_{k+1}^T y_k}{d_k^T y_k} \tag{17}$$

Then the new conjugate gradient directions are :

$$d_{k+1} = -g_{k+1} + \left(1 - \frac{s_k^T s_k}{(f_k - f_{k+1}) + 1/2 y_k^T s_k + 1/6\eta_{k+1}^T s_k}\right)\frac{\eta_{k+1}^T y_k}{d_k^T y_k} d_k \tag{18}$$

For simplicity,we call equation (17) by β_k^{BAH} methods. Also β_k^{BAH} can be write in the manner :

$$\beta_k^{BAH} = \frac{1}{r_k^T y_k}\left(y_k - r \; \frac{\|y_k\|^2}{s_k^T y_k} s_k \right)^T \eta_{k+1} \tag{19}$$

Where:

$$r = \frac{(s_k^T y_k)^2}{\|y_k\|^2}\left[\frac{s_k^T y_k}{s_k^T s_k} * \frac{s_k^T s_k}{(f_k - f_{k+1}) + 2/3\eta_{k+1}^T s_k - 1/2\eta_k^T s_k} \right] \tag{20}$$

Now we are ready to state the steps of the new conjugate gradient methods.
New Algorithms (BAH Algorithms) :

Step 1. Give $x_1 \in R^n$. Set $k=1$ and $d_1 = -\eta_1$.

Step 2. Stop if $\|\eta_1\| \le 10^{-6}$. Otherwise, continue.

Step 3. Find $\alpha_{k+1} > 0$ fulfilling the Wolfe states (4) and (5).

Step 4. Set $x_{k+1} = x_k + \alpha_k d_k$. If $\|\eta_{k+1}\| \le 10^{-6}$, then stop.

Step 5. Compute β_k by the formulaes (19) and d_{k+1} by (3).

Step 6. Put $k = k+1$. Go to step 2.

3. CONVERGENT ANALYSIS

Global convergence of the BAH-algorithm will be proved in this section under the following assumption.

Assumptions

i- $f(x)$ is bounded below on R^n. ii- The gradient $\eta(x)$ is Lipschitz continuous, namely, there exists $L > 0$ such that :

$$\|\eta(x_{k+1}) - \eta(x_k)\| \le L\|x_{k+1} - x_k\|, \ \forall x_{k+1}, \ x_k \in U \tag{21}$$

Under these assumptions on f, then a constant $\Gamma > 0$ exists, such that :

$$\|\eta_{k+1}\| > \Gamma \tag{22}$$

for all $x \in L$. More details can be found in [8].

The sufficient descent condition has a very important property.

3.1. Sufficient descent condition

For the sufficient states to hold, then :

$$d_{k+1}^T \eta_{k+1} \le - c\ \|\eta_{k+1}\|^2 \ , \ c > 0 \tag{23}$$

Theorem 3.1

Let s_k, y_k, $\eta_{k+1} \in R^n$, $\beta_k \in R$ and β_k defined by (19) where $t \in (1/4, \infty)$. If $s_k^T y_k \ne 0$, then $d_{k+1}^T \eta_{k+1} \le -\left[1 - \frac{1}{4r}\right]\|\eta_{k+1}\|^2$

Proof :

Since $d_0 = -\eta_0$, we have $\eta_0^T d_0 = -\|\eta_0\|^2$, which satisfy (23). Multiplying (16) by η_{k+1}, we have :

$$d_{k+1}^T \eta_{k+1} = -\|\eta_{k+1}\|^2 + \left(\frac{\eta_{k+1}^T y_k}{s_k^T y_k} - r \frac{\|y_k\|^2}{(s_k^T y_k)^2} \eta_{k+1}^T s_k \right) s_k^T \eta_{k+1} \tag{24}$$

Yielding :

$$d_{k+1}^T \eta_{k+1} = \frac{(\eta_{k+1}^T y_k)(s_k^T \eta_{k+1})(s_k^T y_k) - \|\eta_{k+1}\|^2 (s_k^T y_k)^2 - r\|y_k\|^2 (\eta_{k+1}^T s_k)^2}{(s_k^T y_k)^2} \tag{25}$$

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

$$(\eta_{k+1}^T y_k)(s_k^T \eta_{k+1})(s_k^T y_k) \le \frac{1}{2}\left[\frac{1}{m^2}(s_k^T y_k)^2 \|\eta_{k+1}\|^2 + m^2 (s_k^T \eta_{k+1})^2 \|y_k\|^2 \right] \tag{26}$$

This yields :

$$d_{k+1}^T \eta_{k+1} \le \frac{\left[\frac{1}{2m^2} - 1 \right](s_k^T y_k)^2 \|\eta_{k+1}\|^2 + \left[\frac{m^2}{2} - r \right](s_k^T \eta_{k+1})^2 \|y_k\|^2}{(s_k^T y_k)^2} \tag{27}$$

from (23) we get :

$$d_{k+1}^T \eta_{k+1} \le \left[\frac{1}{2m^2} - 1 \right] \|\eta_{k+1}\|^2 \le -\left[1 - \frac{1}{2m^2} \right] \|\eta_{k+1}\|^2 \tag{28}$$

Therefore, we get :

$$d_{k+1}^T \eta_{k+1} \le -\left[1 - \frac{1}{4r} \right] \|\eta_{k+1}\|^2 \tag{29}$$

Next we will show that CG methods with BAH converges globally.

3.2. Global convergence property

Dai et al. expressed in [9] that the subsequent result had been basically established Zoutendijk and Wolfe.

Lemma 1.

Let assumptions (i) and (ii) holds. The α_k is take by the Wolfe line search (4) and (5). If :

$$\sum_{k \ge 0} \frac{1}{\|d_{k+1}\|^2} = \infty , \tag{30}$$

then

$$\lim_{k\to\infty} \inf \|\eta_{k+1}\| = 0 \tag{31}$$

Theorem 3.2

Presume that the states in Assumption hold. If $\{d_{k+1}\}$ and $\{\eta_{k+1}\}$ are generated by new technique, then $\lim_{n\to\infty} \inf \|\eta_{k+1}\| = 0$.

Proof :

From (6) and definition of β_k by (19) we get :

$$\|d_{k+1}\| = \|-\eta_{k+1} + \beta_k d_k\| \le \|\eta_{k+1}\| + |\beta_k| \, \|d_k\|$$

$$\|d_{k+1}\| \le \|\eta_{k+1}\| + \left\| \left(y_k - r \frac{\|y_k\|^2}{s_k^T y_k} s_k \right) \right\| \frac{\|\eta_{k+1}\|}{\|s_k\|\|y_k\|} \|s_k\|$$

$$\le \|\eta_{k+1}\| + \frac{\|y_k\|\|\eta_{k+1}\| + r \dfrac{\|\eta_{k+1}\|\|y_k\|^2\|s_k\|}{\|s_k\|\|y_k\|}}{\|s_k\|\|y_k\|} \|s_k\| \tag{32}$$

$$\le \|\eta_{k+1}\| + \frac{\|y_k\|\|\eta_{k+1}\| + r\|\eta_{k+1}\|\|y_k\|}{\|s_k\|\|y_k\|} \|s_k\|$$

$$\le [2 + r] \, \|\eta_{k+1}\|$$

This relation explain to facilitate :

$$\sum_{k\ge1} \frac{1}{\|d_{k+1}\|^2} \ge \left(\frac{1}{2+r} \right) \frac{1}{\Gamma} \sum_{k\ge1} 1 = \infty \tag{33}$$

Consequently, from Lemma 1 we have $\lim_{k\to\infty} \inf \|\eta_{k+1}\| = 0$, which for target perform is uniformly, then equivalent to $\lim_{k\to\infty} \|\eta_{k+1}\| = 0$.

4. NUMERICAL RESULTS

We tested BAH-algorithm. The test functions and their primary values are wan from [10]. Furthermore, Optimization problems used in many papers for example, see [11-18]. In addition to these functions, there are various other functions that havebeen used for testing in the followwing research [19-25]. The numerical results are reported in Table 1 : "the first column and the second one represent the problem name and its dimension in [10], respectively. NI, NR and NF in the table denote the number of iterations, function evaluations and the number of restart calls, respectively".

All the algorithm area unit enforced in Fortran 90 ninety. All told cases, double preciseness arithmetic were used. The parameters in Wolfe states area unit set as $\delta_1 = 0.001$ and $\delta_2 = 0.9$. BAH-algorithm is efficient we see from Table 1.

Table 1. The Numerical Results of the FR and BAH Methods

P. No.	n	FR algorithm			BAH algorithm		
		NI	NR	NF	NI	NR	NF
1	100	47	18	93	42	21	95
	1000	78	45	131	39	18	86
2	100	32	15	52	22	10	41
	1000	22	10	42	23	12	44
3	100	25	11	43	24	9	45
	1000	46	28	741	25	5	54
4	100	32	13	64	35	16	68
	1000	77	46	129	28	11	55
5	100	15	6	25	17	9	29
	1000	F	F	F	26	20	281
6	100	37	8	67	43	18	69
	1000	73	27	115	58	20	91
7	100	89	32	174	72	45	163
	1000	107	40	211	88	52	226
8	100	32	12	65	24	14	54
	1000	53	22	116	36	21	89
9	100	9	4	18	10	6	18
	1000	12	7	82	9	7	55
10	100	74	21	123	93	30	140
	1000	370	88	616	341	77	567
11	100	69	50	1202	54	37	653
	1000	98	82	1967	46	33	502
12	100	23	11	45	21	12	40
	1000	27	11	55	17	9	41
13	100	49	22	66	18	12	33
	1000	129	67	166	13	9	26
14	100	122	62	156	14	9	25
	1000	130	66	166	16	10	29
15	100	112	55	147	44	19	65
	1000	110	54	145	61	34	82
Total		1999	933	7022	1333	585	3485

F : The algorithm fail to converge.

Problems numbers indicant for : "1. is the Extended Rosenbrock, 2. is the Extended Beale, 3. is the Generalized Tridiagonal 1, 4. is the Extended Tridiagonal 1, 5. is the Extended Three Expo Terms, 6. is the Generalized Tridiagonal 2, 7. is the Extended Maratos, 8. is the Extended Quadratic Penalty QP2, 9. is the ARWHEAD (CUTE), 10. is the Partial Perturbed Quadratic, 11. is the EDENSCH (CUTE), 12. is the LIARWHD (CUTE), 13. is the DENSCHNC (CUTE), 14. is the Extended Block-Diagonal BD2, 15. is the Generalized quartic GQ2".

Can summarize our numerical results in Table 2 based on the percentage performance for all Tools used in these comparisons.

Table 2. Percentage Performance of the Methods

	NI	NR	NF
FR	100 %	100 %	100 %
BAH	66.68 %	62.70 %	49.62 %

It is clear from Table 2 that taking, over all the tools as a 100% for FR method the BAH method has an improvement, in about 33% NI ; 37% NR and 50% NF, these results indicate that New method is in general is the best .

5. CONCLUSIONS

A new kind of parameter in the conjugate gradient methods for large-scale unconstrained optimization problems is proposed. Reveal Numerical that the new method is superior in practice with competitive FR method. We choose the parameter β_k appropriately, to boost the performance of the conjugate gradient methods.

REFERENCES

[1] Pluim J.P.W., Likar B., and Gerritsen F.A., "A comparison of acceleration techniques for nonrigid medical image registration", University Medical Center Utrecht, Image Sciences Institute, 2006, pp. 151-159.

[2] Saman B.K. and Reza G.," hybridization of the Hestenes–Stiefel and Dai-Yuan conjugate gradient methods based on a least–squares approach", Journal Optimization Methods and Software,2014, pp.1-13.

[3] Wolfe, P.," Convergence conditions for ascent methods". SIAM Rev. 11(2), 1969, pp.226-235.

[4] Rao S. S., " Engineering Optimization Theory and Practice", 4th edition, John Wiley & Sons Inc., New Jersey, Canada,2009.

[5] Hestenes, M. R. and Stiefel, E. L., "Method of conjugate gradients for Solving linear systems" Journal National Standards 49, 1952, pp. 409-436.

[6] Fletcher R. and Reeves C. M., "Funtion minimization by conjagate gradients", Computer Journal 7, 1964, pp. 149-154.

[7] Zhang J.Z., Deng N.Y., and Chen L.H.," New quasi-Newton equation and related methods for unconstrained optimization," Journal of Optimization Theory and Applications, vol. 102,1999, pp. 147-167.

[8] Dai Y. H. and Yuan Y.," A nonlinear conjugate gradient method with a strong global convergence property", SIAM J. optimization, 1999, pp. 177-182.

[9] Dai Y.H., Han, J.Y., Liu, G.H., Sun, D.F., Yin, .X. and Yuan, Y., "Convergence properties of nonlinear conjugate gradient methods". SIAM Journal on Optimization , 10,1999,pp. 348-358.

[10] Andrie N., " An Unconstrained Optimization Test functions collection " Advanced Modeling and optimization,10, 2009, pp.147-161.

[11] Basim A. H. "A new formula for conjugate parameter computation based on the quadratic model", Indonesian Journal of Electrical Engineering and Computer Science (IJEECS),2019; 3: 954-961.

[12] Basim A. H. , Hussein K. K. "A new class of BFGS updating formula based on the new quasi-newton equation", Indonesian Journal of Electrical Engineering and Computer Science (IJEECS),2019; 3: 945-953.

[13] Basim A. H. , Zeyad M. A. and Hawraz N. J. "A descent extension of the Dai - Yuan conjugate gradient technique", Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 2019; 2: 661-668.

[14] Li C. Fang L. and Cao X. "Global convergenceof a kind of conjugate gradient method". TELKOMNIKA (Telecommunication, Computing, Electronics and Control). 2013; 11: 544-549,2013.

[15] Li C., "A modified conjugate gradent method for unconstrained optimization," TELKOMNIKA (Telecommunication, Computing, Electronics and Control)", vol. 11, pp. 6373-6380, 2013.

[16] Dauda1 M. K., Mustafa M., Mohamad A., Nor S.A. " Hybrid conjugate gradient parameter for solving symmetric systems of nonlinear equations", Indonesian Journal of Electrical Engineering and Computer Science (IJEECS),2019; 16: 539-543.

[17] Gilbert J.C. and Nocedal J., "Global convergence properties of conjugate gradient methods for optimization," SIAM J. Optim. 2 (1992), pp. 21–42.

[18] Hager W.W. and Zhang H., "A survey of nonlinear conjugate gradient methods," Pac. J. Optim. 2 (2006), pp. 35–58.

[19] Li G., Tang C., and Wei Z., "New conjugacy condition and related new conjugate gradient methods for unconstrained optimization," J. Comput. Appl. Math. 202 (2007), pp. 523–539.

[20] Babaie S. K., "An eigen value study on the sufficient descent property of a modified Polak-Ribi- Polak conjugate gradient method," *Bull. Iranian Math. Soc*., vol. 40, pp. 235-242, 2014.

[21] Guangming Z., "A descent algorithm without line search for unconstrained optimization," *Applied Mathematics and Computation*, vol. 215, pp. 2528-2533, 2009.

[22] Ghani N., Kamaruddin S., Ramli M., Musirin I. And Hashim H. "Enhanced BFGS quasi-Newton backpropagation models on MCCI data". Indonesian Journal of Electrical Engineering and Computer Science. 2017;8:101-106.

[23] Ahmad A. and Zabidin S., "Modification of Nonlinear Conjugate Gradient Method with Weak Wolfe-Powell Line Search," *Hindawi Abstract and Applied Analysis*, 2017, pp. 1-6.

[24] *Li D. H. and Fukushima M., " A derivative-free line search and global convergence of Broyden-like methods for nonlinear equations," Optimization Methods and Software, 13 (2000), pp.181-201.*

[25] *Babaie-K. S. ., " A modified BFGS algorithm based on a hybrid secant equation," Sci. China Math. 54 (2011), pp. 2019–2036.*