A new formula for conjugate parameter computation based on the quadratic model

Basim A. Hassan
Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, Iraq

ABSTRACT
The conjugacy coefficient is the very basis of a diversity of the conjugate gradient methods. In this research, we derivation a new formula of conjugate gradient methods based on the quadratic model. Our arithmetical findings have revealed that, our new method has the most excellent performance contrast to the other standard CG methods. Also give proof viewing that this method converges globally.

Keywords:
Conjugate gradient method
Global convergence
Sufficient descent property

Copyright © 2019 Institute of Advanced Engineering and Science. All rights reserved.

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

1. INTRODUCTION
Methods of conjugate gradient are particularly important class because of their convergence features, a very simple application endeavor in computer performances and very good in solving big problems [1]. We are concerned with conjugate gradient methods for finding a local minimum of the function:

\[ \min \{ f(x) \mid x \in \mathbb{R}^n \} \] (1)

where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is a incessantly differentiable function.

Conjugate gradient methods for solving (1) are iterative methods of the form:

\[ x_0 \in \mathbb{R}^n, \quad x_{k+1} = x_k + \alpha_k h_k \] (2)

where \( \alpha_k > 0 \) is a step size and \( h_k \) is the search direction generated by:

\[ h_0 = -q_0, \quad h_{k+1} = -g_{k+1} + \beta_k h_k \] (3)

where \( q_{k+1} \) denotes gradient of \( f(x_{k+1}) \) at the point \( x_{k+1} \) and \( \beta_k \) is a scalar representing different methods. The best-known parameter for \( \beta_k \) is:

Journal homepage: http://iaescore.com/journals/index.php/ijeecs
A new formula for conjugate parameter computation based on the quadratic model (Basim A. Hassan)

\[ \beta_k^{FR} = \frac{q_k^T q_{k+1}}{q_k^T q_k} \]

introduced by Fletcher and Reeve, FR [2]. For other reviews of CG-classic methods see for instance [3-7]. As we can see in the parameter for \( \beta_k \) the difference \( f_k - f_{k+1} \) is not used at all. To obtain better conjugate gradient methods, many modified methods using value of objective function have been presented some of it:

Hideaki and Yasushi [8] made a modification on the CG parameter as follows:

\[ \beta_k^{HY} = \frac{q_k^T q_{k+1}}{2/\alpha_k (f_k - f_{k+1})} \]

Lately, Basim and Haneen [9], using quadratic function, modified CG parameter as follows:

\[ \beta_k^{BHQ} = \frac{q_k^T q_{k+1}}{\alpha_k (q_k^T h_k)^2 / (2(f_k - f_{k+1}))} \]

There a lot of are other modified CG-methods that we did not cover up in this paper. Commonly, steplength \( \alpha_k \) in (2) is selected to satisfy the Wolfe line search states:

\[ f(x_k) - f(x_k + \alpha_k h_k) \geq -\delta \alpha_k q_k^T h_k \]
\[ q^T h_k \geq \sigma q_k^T h_k \]

where \( 0 < \delta < \sigma < 1 \). In addition, the sufficient descent condition:

\[ q_{k+1}^T h_{k+1} \leq -\epsilon \| q_{k+1} \| \]

All from the Wolfe states and descent condition are good property to prove convergence. More details can be found in [10, 11].

A key factor of conjugate gradient methods is how to select the conjugancy coefficient \( \beta_k \). Below based on the quadratic model will introduced the new conjugate gradient methods. The resulting modified CG method retains global convergence, and performs slightly better than the FR-CG method on some test problems.

2. NEW FORMULA FOR CONJUGATE PARAMETER COMPUTATION AND ALGORITHM

In conjugate gradient (CG) methods the formula for the new step becomes:

\[ h_{k+1} = -q_{k+1} + \beta_k h_k \]

where \( \beta_k \) is found by imposing the condition that \( h_k^T R v_k \) and is given as:

\[ \beta_k = \frac{q_{k+1}^T R v_k}{h_k^T R v_k} \]

where \( R \in \mathbb{R}^{n \times n} \) is a non negative definite and if \( \alpha_k \) is the exact (\( q_{k+1}^T d_k = 0 \)) one dimensional minimizer given by:
More details about the conjugate gradient method can be found in [8, 12].

Now, we derive the new formulas for conjugate parameter computation. We shall think a different look of the denominator $h_k^T \gamma h_k$. Based on quadratic model and using $q_{k+1}^T h_k = 0$, we get:

\[
f(x_k) = f(x_{k+1}) + q_{k+1}^T(-v_k) + \frac{1}{2}(-v_k)^T \gamma (-v_k) = f(x_{k+1}) + \frac{1}{2} \alpha_k^2 h_k^T \gamma h_k
\]

which implies that:

\[
\alpha_k^2 h_k^T \gamma h_k = f(x_k) - f(x_{k+1}) - q_{k+1}^T h_k / 2
\]

\[
h_k^T \gamma v_k = (f(x_k) - f(x_{k+1})) / \alpha_k - q_{k+1}^T h_k / 2
\]

From this we define formula:

\[
\beta_k = \frac{\gamma_{k+1} y_k}{(f(x_k) - f(x_{k+1})) / \alpha_k - q_{k+1}^T h_k / 2}
\]

Quadratic function is optimal solution to the open problem known and is taken from Yuan [13]. More details can be found in [1].

For quadratic functions and under exact line searches, all the gradients of $f$ at the different iterates are mutually orthogonal. That is $q_{k+1}^T q_k = 0$. Formula (15) further reduces to:

\[
\beta_k^B = \frac{q_{k+1}^T q_{k+1}}{(f(x_k) - f(x_{k+1})) / \alpha_k - q_{k+1}^T h_k / 2}
\]

In order to adjust or extension the over formula as follow:

\[
\beta_k^{MB} = \frac{q_{k+1}^T q_{k+1}}{\max(h_k^T y_k, (f(x_k) - f(x_{k+1})) / \alpha_k - q_{k+1}^T h_k / 2)}
\]

With this new, we presenting algorithm as follows.

**New Algorithm:**

1. Given initial $x_1 \in R^n$ and estimate the $q_1$ and $d_1 = -q_1$.
2. If $\|q_{k+1}\| \leq 10^{-6}$, then stop.
3. Cubic search to estimate $\alpha_k$ and which satisfying the Wolfe conditions (7)-(8) and update the variables $x_{k+1} = x_k + \alpha_k h_k$.
4. Estimate $\beta_k^B$ which defined in (16) and (17).
5. Set $k = k + 1$ and repeat step 2 to step 5.

### 3. GLOBAL CONVERGENCE ANALYSIS BY SEVERAL LINE SEARCHES FOR $\beta_k^B$ METHOD

The aspire of this section is to study the worldwide convergence activities of new Algorithm.
For the sufficient state to hold, then:
\[ q_{k+1}^T h_{k+1} \leq -c \| h_{k+1} \|^2, \quad c > 0 \]  
(18)

If \( \alpha_k \) is vied by the Wolfe states (7) and (8), afterward the search direction \( \beta_k^B \) satisfies (16).

**Theorem 1.**
Consider the new \( \beta_k^B \) method. If \( \alpha_k \) is vied by the Wolfe states (7) and (8), afterward:
\[ q_{k+1}^T h_{k+1} \leq -\mu \| q_{k+1} \|^2 \]  
(19)

Moreover \( \beta_k^B > 0 \).

**Proof:**
The proof is by induction. For \( k = 0 \) then \( q_0^T d_0 = -\| q_0 \|^2 \leq -\mu \| q_0 \|^2 \). Suppose that (16) is satisfies for \( k \). Now we prove that (18) holds for \( k + 1 \). By multiplying \( q_{k+1}^T \) on both sides of (3), we obtain:
\[
q_{k+1}^T h_{k+1} = q_{k+1}^T (-q_{k+1} + \beta_k^B q_{k+1}^T h_k)
\]
\[
= -\| q_{k+1} \|^2 + \frac{\| q_{k+1} \|^2}{\| q_{k+1} \|^2} \left( f_k - f_{k+1} \right) + \frac{\| q_{k+1} \|^2}{\| q_{k+1} \|^2} \left( \alpha_k q_k^T h_k - 1/2 \right) q_{k+1}^T h_k
\]
(20)

From Wolfe states we get, \( f_k - f_{k+1} \) \( \geq -\alpha_k q_k^T h_k \) and \( g_{k+1}^T d_k \geq \alpha_k q_k^T h_k \). Put this value in the above equation to get:
\[
q_{k+1}^T h_{k+1} \leq -\| q_{k+1} \|^2 + \frac{\| q_{k+1} \|^2}{\| q_{k+1} \|^2} \left( \alpha_k q_k^T h_k - 1/2 \right) q_{k+1}^T h_k
\]
\[
\leq -\| q_{k+1} \|^2 - \frac{\| q_{k+1} \|^2}{\| q_{k+1} \|^2} \left( \alpha_k q_k^T h_k - \alpha_k q_k^T h_k \right)
\]
\[
\leq -\mu \| q_{k+1} \|^2
\]  
(21)

where \( \mu = 1 + \frac{\sigma}{\delta + 1/2} \). Therefore, (18) is fulfilled for \( \forall k \). Additional, from (21) analysis, we too obtain \( \beta_k^B > 0 \). By mathematical induction method, we obtain the desired result.

**Global convergence of the \( \beta_k^B \) method**

To study the worldwide convergence of \( \beta_k^B \) - method, the a number of basic assumption.

B1. The level set \( \mathcal{A} = \{ x \in \mathbb{R}^n \mid f(x) \leq f(x_i) \} \) is bounded.

B2. There exists a constant \( L > 0 \) such that for any :
\[ \| f(x) - f(y) \| \leq L \| x - y \|, \quad \forall x, y \in U \]

Within this subdivision, we make the convergence of the \( \beta_k^B \) method. Initially, the researcher manifests that \( \beta_k^B \) has the same features to the \( \beta_k^{DY} \) method and Zoutendijk state, that is very much employed to demonstrate worldwide convergence.
The $\beta_k^B$ method has the same features of the $\beta_k^{DY}$ method, that is considered of importance to the worldwide convergence of the following investigations.

**Theorem 2**

Consider $\{x_k\}$ generated by new method. Then for every $k$, the relations $0 < \beta_k^B \leq \frac{q_{k+1}^T h_{k+1}}{q_k^T h_k}$ always hold.

**Proof:**

From Theorem 1, we know $0 < \beta_k^B$. Multiplying (3) by $q_{k+1}^T$ with new formula we obtain:

$$q_{k+1}^T h_{k+1} = -q_{k+1}^T q_{k+1} + \beta_k q_{k+1}^T h_k$$

$$= (-[(f_k - f_{k+1})/\alpha_k - (q_k^T h_k / 2)] + q_{k+1}^T h_k) \beta_k.$$  \hfill (22)

Moreover, by Theorem 1, we have that:

$$[(f_k - f_{k+1})/\alpha_k - (q_k^T h_k / 2)] - q_{k+1}^T h_k \geq h_k^T y_k - q_{k+1}^T h_k$$

$$= q_{k+1}^T h_k - q_k^T h_k - q_{k+1}^T h_k$$

$$= -q_k^T h_k > 0$$  \hfill (23)

This also shows that $\beta_k^B \leq \frac{q_{k+1}^T h_{k+1}}{q_k^T h_k}$. Therefore the proof is complete.

**Lemma 1**

Supposition (B1)-(B2) holds. Let the methods in the form of (2)-(3), where $d_k$ satisfy (18) and $\alpha_k$ satisfies the (7)-(8) states. Then we:

$$\sum_{k=1}^{\infty} \frac{(q_{k+1}^T h_{k+1})^2}{\|h_{k+1}\|^2} < \infty$$  \hfill (24)

We ascertain the worldwide convergence of the $\beta_k^B$ method.

**Theorem 3**

Supposition (B1)-(B2) holds. Consider $\{x_k\}$ be generated by New Algorithm. Then $\lim_{k \to \infty} |q_{k+1}| = 0$.

**Proof:**

We carry on by disagreement. Presume that $\|q_{k+1}\|^2 > \gamma$ for $\gamma > 0$. By (3), it follows so as to

$$h_{k+1} + q_{k+1} = \beta_k^B h_k.$$ This jointly by theorem 2, imply:

$$\|h_{k+1}\|^2 = (\beta_k^B \|q_k\|^2 - 2h_{k+1}^T q_{k+1} - \|q_{k+1}\|^2)^2$$

$$\leq \left( \frac{q_{k+1}^T h_{k+1}}{q_k^T h_k} \right)^2 \|d_k\|^2 - 2h_{k+1}^T q_{k+1} - \|q_{k+1}\|^2$$  \hfill (25)

Dividing both sides of (25) by $(h_{k+1}^T q_{k+1})^2$, we obtain:

Indonesian J Elec Eng & Comp Sci, Vol. 13, No. 3, March 2019 : 954 – 961
A new formula for conjugate parameter computation based on the quadratic model (Basim A. Hassan)

\[
\left| \frac{\|h_k\|^2}{(h_k^T q_k)^2} \right| \leq \frac{\|h_k\|^2}{(h_k^T q_k)^2} - \frac{2}{(h_k^T q_k)^2} \cdot \left( \frac{\|h_k\|^2}{(h_k^T q_k)^2} + \frac{1}{\|h_k\|^2} \right)
\]

(26)

\[
\leq \left( \frac{\|h_k\|^2}{(h_k^T q_k)^2} + \frac{1}{\|h_k\|^2} + \frac{1}{\|h_k\|^2} \right)
\]

(27)

Noting that \( \frac{\|h_k\|^2}{(h_k^T q_k)^2} = \frac{1}{\|q_k\|^2} \), by recurrence formula above (26), we have:

\[
\left| \frac{\|h_{k+1}\|^2}{(h_{k+1}^T q_{k+1})^2} \right| \leq \left| \frac{\|h_k\|^2}{(h_k^T q_k)^2} - \frac{1}{\|h_k\|^2} \right|
\]

\[
\leq \left( \frac{\|h_k\|^2}{(h_k^T q_k)^2} + \frac{1}{\|h_k\|^2} + \frac{1}{\|h_k\|^2} \right) \leq \frac{k}{\gamma}
\]

Thus \( \left( \frac{h_{k+1}^T q_{k+1}}{\|h_{k+1}\|^2} \right)^2 \geq \frac{\gamma}{k} \), and this implies that:

\[
\sum_{k=1}^{\infty} \left( \frac{h_{k+1}^T q_{k+1}}{\|h_{k+1}\|^2} \right)^2 = \infty
\]

(28)

it contradicts Lemma 1. Therefore, the desired result holds.

4. ARITHMETICAL FINDINGS AND DISCUSSION

In this part, arithmetical findings are reported. We test and compare the new methods with FR method whose results be given by [2].

Using Fortran 90 to code this methods. In our application, we select the following parameters : \( \delta = 0.001 \) and \( \sigma = 0.9 \). The examination problems are selected from ref. [15, 16]. Optimization problems exist in many areas [17]. The stopping state is :

\[
\|q_{k+1}\| \leq 10^{-6}
\]

(29)

The arithmetical findings are cataloged in table 1, where the column “Problem” stands for the label of the examined problem. “Dim” refers to the dimension of the test problems. The results are denoted by NI and NF refer to the table of iterations and function estimations successively.

In summary, the arithmetical findings show that New methods are more efficient than the FR method and provides an efficient method for solving unconstrained optimization problems.

Fail: The algorithm fail to converge. Problems numbers indicant for

1. is the Extended Rosenbrock,
2. is the Extended White & Holst,
3. is the Extended Beale,
4. is the Extended Tridiagonal 1,
5. is the Extended Three Expo Terms,
6. is the Generalized Tridiagonal 2,
7. is the Extended Powell,
Table 1. Numerical Results of new Algorithms and FR-CG algorithm

| P. No. | n   | FR algorithm | B algorithm | MB algorithm |
|-------|-----|--------------|-------------|--------------|
|       | n   | NI | NF | NI | NF | NI | NF |
| 1     | 100 | 47 | 93 | 39 | 79 | 37 | 75 |
|       | 1000| 78 | 131| 37 | 78 | 38 | 81 |
| 2     | 100 | 43 | 88 | 37 | 84 | 37 | 84 |
|       | 1000| 46 | 92 | 35 | 79 | 28 | 62 |
| 3     | 100 | 32 | 52 | 17 | 32 | 13 | 26 |
|       | 1000| 22 | 42 | 13 | 24 | 12 | 24 |
| 4     | 100 | 32 | 64 | 11 | 23 | 10 | 26 |
|       | 1000| 77 | 129| 13 | 26 | 16 | 31 |
| 5     | 100 | 15 | 25 | 8  | 13 | 17 | 25 |
|       | 1000| Fail| Fail|28 | 342| 25 | 474|
| 6     | 100 | 37 | 67 | 40 | 62 | 41 | 63 |
|       | 1000| 73 | 115| 60 | 98 | 64 | 101|
| 7     | 100 | 180| 313| 61 | 115| 81 | 115|
|       | 1000| Fail| Fail|78 | 149| 83 | 159|
| 8     | 100 | 124| 231| 47 | 81 | 55 | 96 |
|       | 1000| 445| 711| 181| 313| 160| 281|
| 9     | 100 | 71 | 110| 26 | 50 | 31 | 59 |
|       | 1000| 47 | 84 | 26 | 52 | 26 | 51 |
| 10    | 100 | 130| 196| 111| 167| 111| 174|
|       | 1000| 364| 593| Fail|Fail|471|753|
| 11    | 100 | 13 | 25 | 13 | 26 | 13 | 26 |
|       | 1000| 15 | 29 | 14 | 29 | 12 | 25 |
| 12    | 100 | 121| 218| 80 | 120| 84 | 133|
|       | 1000| 345| 634| 219| 347|249|388|
| 13    | 100 | 74 | 123| 89 | 134|81 |125 |
|       | 1000| 370| 616| 273| 454|244|410|
| 14    | 100 | 122| 156| 12 | 23 | 12 | 23 |
|       | 1000| 130| 166| 14 | 23 | 12 | 23 |
| 15    | 100 | 23 | 45 | 19 | 34 | 17 | 33 |
|       | 1000| 27 | 55 | 20 | 45 | 24 | 53 |
| Total | 2739| 4610|1515|2611|1541|2613|

Fail: The algorithm fail to converge.

Table 2. Averages Efficiency of the New Algorithms

|          | FR algorithm | B algorithm | MB algorithm |
|----------|--------------|-------------|--------------|
| NI       | 100 %        | 55.32 %     | 56.26 %      |
| NF       | 100 %        | 56.64 %     | 56.68 %      |

Indonesian J Elec Eng & Comp Sci, Vol. 13, No. 3, March 2019 : 954 – 961
5. CONCLUSIONS
In this research, we have derived a new CG-methods based on the quadratic model. Arithmetical findings have been accounted, which explained the usefulness of our method.

REFERENCES
[1] Andrei N. Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization. Bulletin of the Malaysian Mathematical Sciences Society, 2011; 34: 319-330.
[2] Fletcher R, Reeves C. Function minimization by conjugate gradients. Computer J, 1964; 7, 149-154.
[3] Dai H, Yuan Y. A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optimization, 1999; 177-182.
[4] Fletcher R. Practical Method of Optimization. 2nd Edition, John Wiley and Sons. New York. 1989.
[5] Hestenes R, Stiefel L. Method of conjugate gradients for Solving linear systems. Journal National Standards, 1952; 49: 409-436.
[6] Polak E, Ribiere G. Note for Convergence Direction Conjugate. Revue Francaise Informant, Reserche. Operationelle, 1969: 35-43.
[7] Liu Y, Storey C. Efficient generalized conjugate gradients algorithms. Part 1: Theory. J. Optimization Theory and Applications, 1991; 69: 129-137.
[8] Yasushi N, Hideaki I. Conjugate gradient methods using value of objective function for unconstrained optimization Optimization Letters. Springer, 2011; 6(5): 941-955.
[9] Basim H, Haneen A. New Nonlinear Conjugate Gradient Formulas for Solving Unconstrained Optimization Problems. Al-Mustansiriyah Journal of Science, 2016; 3: 82-88.
[10] Wolfe P. Convergence conditions for ascent methods. SIAM Review, 1969; 11: 226-235.
[11] NLi C. Fang L. and Cao X. Global convergence of a kind of conjugate gradient method. TELKOMNIKA (Telecommunication, Computing, Electronics and Control), 2013; 11: 544-549.
[12] Nocedal S. Theory of algorithms for unconstrained Optimization. Acta Numerica. 1999.
[13] Yuan Y. Some problems in nonlinear programming. (Ed.) Numerical Linear Algebra and Optimization, Science Press, Beijing / New York, 2003: 90-110.
[14] ZZoutendijk G. Nonlinear programming computational algorithms. In: Integer and Nonlinear programming, Abadie, J.(ED). North-Holland, Amsterdam, 1970: 37-86.
[15] Andrie N. An Unconstrained Optimization Test functions collection. Advanced Modeling and optimization. 2008; 10: 147-161.
[16] LLi C. A modified conjugate gradient method for unconstrained optimization. TELKOMNIKA (Telecommunication, Computing, Electronics and Control), 2013; 11: 6373-6380.
[17] SSeman K, Puspita FM, Taib BM, Shafii Z. An improved optimization model of internet charging scheme in multi service networks. TELKOMNIKA (Telecommunication, Computing, Electronics and Control). 2012; 10(3): 592-598.