New Parameter of CG-Method with Exact Line Search for Unconstraint Optimization

Mohammed M. Abdel Hady, Maha S. Younis

Department of Mathematics, College of Education for Pure Science, University of Mosul, Mosul, Iraq
Email: mohammedm202072@gmail.com

Abstract
In this paper, a new CG method has been introduced to solve nonlinear equations systems. This method achieved the conditions of descent and global convergence, using the exact line search. The numerical results were good compared to other methods in terms of the number of iterations and the number of functions evaluation.

Subject Areas
Mathematical Analysis

Keywords
Parameter, CG-Method, Optimization

1. Introduction
The conjugate gradient method is one of the important ways to find the minimum value of a function for unconstrained optimization.

The conjugate gradient method is widespread because its requirements are a small memory. Unconstrained optimization problem can be expressed as follows:

$$\min_{x \in \mathbb{R}^n} f(x)$$

(1)

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous and derivative function. The CG method generates frequent updates in this format.

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 1, 2, 3, 4$$

(2)

where $x_k$ is the current iteration point, $\alpha_k > 0$ is the positive step size using the “exact line search” as shown by the following:

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

(3)
and $d_k$ is the search direction, which we get as follows:

$$
d_k = \begin{cases} 
-g_k & \text{for } k = 0 \\
-g_k + \beta_k d_{k-1} & \text{for } k \geq 1 
\end{cases} \tag{4}$$

where $k$ is integer and that $g_k$ is the gradient of the function $f(x)$ and that $\beta_k$ is the coefficient of the conjugate gradient associated with the function $f(x)$ at the point $x_k$.

Some of the known conjugation methods are:

- $\beta^\text{FR}_k = \frac{g_k^T g_k}{\|g_{k-1}\|^2}$, $\beta^\text{PR}_k = \frac{g_k^T (g_k - g_{k-1})}{\|g_k\|^2}$
- $\beta^\text{HS}_k = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}$, $\beta^\text{DY}_k = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}}$
- $\beta^\text{LS}_k = \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T d_{k-1}}$, $\beta^\text{CD}_k = -\frac{g_k^T g_{k-1}}{d_{k-1}^T g_{k-1}}$

The coefficient gradient coefficient $\beta_k \in \mathbb{R}$ is a numerical constant, which determines the difference in different CG methods when $g_{k-1}, g_k$ denote the gradient of a function $f(x)$ at points $x_{k-1}, x_k$, respectively.

Above methods are known as:  

- Fletcher and Reeves (FR) \[1\], Polka and Ribiere (PR) \[2\], Hestenes and Steifel (HS) \[3\], Dai and Yuan (DY) \[4\], Liu and Story (LS) \[5\], Conjugate Descent (CD) by Fletcher \[6\].

These aforementioned methods behave strictly convex quadratic functions in a behavior that is completely different from what they do in non-quadratic general functions. In any case, most of these methods examine the properties of universal approach in the field of conjugated gradient.

However, in recent years, there have been many attempts that have been directed towards building new formulas for CG methods with good numerical performance and achieving the characteristics of global convergence.

2. The New Conjugate Gradient and Its Algorithm

It is well known that the methods of numerical optimization are iterative methods and there is no specific method suitable for all types of problems. Each method has its advantages and new features as well as some of the characteristics that are not good and are efficient for some types of problems and not efficient for other types of problems.

The new coefficient of gradient is

$$
\beta^\text{UE}_k = \frac{g_k^T g_k}{(g_k + d_{k-1})^T d_{k-1}} \tag{5}
$$

New method algorithm

Step (1): Set $\epsilon > 0, d_0 = -g_0, k = 0$ and choose an initial value $X_0$

Step (2): Calculate $\beta^\text{UE}$ from (5)
Step (3): Calculate 
\[ d_k = -g_k + \beta_k^{ME} d_{k-1} \]
In the case if \( \|g_k\| = 0 \), stop
Step (4): Calculate 
\[ \alpha_k = \min_{\alpha > 0} f(x_k + \alpha d_k) \]
Step (5): Calculate the new point with the following iterative formula:
\[ x_{k+1} = x_k + \alpha_k d_k \]  
(6)
Step (6): Test if it is
\[ f(x_{k+1}) < f(x_k) \]
And also
\[ \|g_k\| \leq \epsilon \quad \text{Stop} \]
Otherwise, go to step (1) with \( k = k + 1 \)
The coefficient \( \beta_k \) is chosen in such a way that \( d_{k+1} \) is G-conjugate to \( d_0, d_1, d_2, \ldots, d_k \).

**Lemma (1)**

In the conjugate direction algorithm
\[ g_{k+1}^T d_i = 0 \quad \text{for all } k, 0 \leq k \leq n-1 \text{ and } 0 \leq i \leq k. \]

**Proposition:** In the conjugate gradient algorithm the direction \( d_0, d_1, \ldots, d_{n-1} \) are G-conjugate.

**Proof:** By using induction
We first show
\[ d_i^T G d_i = 0 \]
\[ d_0^T G d_i = d_0^T G(-g_i + \beta_i d_0) \]
\[ = -d_0^T G g_i + \beta_i d_0^T G d_0 \]
by ELS
\[ = \frac{\beta_i d_0^T (g_i - g_0)}{\alpha_0} \quad \text{when } \alpha_0 \text{ in (3)} \]
\[ = -\beta_i d_0^T g_0 \]
\[ = -\frac{g_i^T g_1}{(g_1 + d_0)^T d_0} \frac{d_0^T g_0}{\alpha_0} \quad \text{by Lemma (1) and ELS we get } = \text{zero} \]

Now we assume that \( d_{k-1}^T G d_k = 0 \) is correct. And we prove that \( d_{k}^T G d_{k+1} = 0 \)
\[ d_k^T G d_{k+1} = d_k^T G(-g_{k+1} + \beta_{k+1} d_k) \]
\[ = -d_k^T G g_{k+1} + \beta_{k+1} d_k^T G d_k \]
\[ = \frac{\beta_{k+1} d_k^T (g_{k+1} - g_k)}{\alpha_k} \quad \text{when } \alpha_k \text{ in (3)} \]
\[ = -\beta_{k+1} d_k^T g_k \]
\[ = -\frac{g_k^T g_{k+1}}{(g_{k+1} + d_k)^T d_k} \frac{d_k^T g_k}{\alpha_k} \]
By Lemma (1) and ELS we get \( d_k^T G d_{k+1} = 0 \).
The fulfillment of the descent condition \( g_k^T d_k < 0 \).
The new method is shown as follows:

\[
g_k^T d_k = -g_k^T g_k + \beta_k^r g_{k-1}^T d_{k-1}
\]

By ELS, we get

\[
g_k^T d_k = -g_k^T g_k = -\|g_k\|^2 < 0
\]
So \( g_k^T d_k < 0 \).
Thus the descent condition is held.

3. Global Convergence

An analysis of the overall convergence using the Exact Line search (ELS) demonstrates according to the following hypotheses:

1) In the neighborhood \( N \) of \( L \) the function \( f(x) \) is continuous, derivative, bound and defined at the level set \( L = \{ x, f(x) \leq f(x_0) \} \), when \( x_0 \) is an initial point.

2) The gradient is Lipschitz condition when there is a constant number \( L > 0 \) so that

\[
\|g(x) - g(y)\| \leq L \|x - y\|, \text{ for all } x, y \in N
\]

According to these assumptions we have the following taken by Zoutendijk [7].

**Lemma 2:** Assuming assumption 1) is correct, we consider the conjugate regression methods formulated in formula (3), where \( d_k \) is the descent search direction, \( \alpha_k \) fulfills the exact line search of the minimization rules, so the following condition defined by the Zoutendijk condition is held:

\[
\sum_{k=0}^{\infty} \left( \frac{g_k^T d_k}{d_k} \right)^2 < \infty
\]

From Lemma (2), we can obtain a convergence theorem of the conjugate gradient CG method using

\[
\beta_k^{MF} = \frac{g_k^T g_k}{(g_k + d_{k-1})^T d_{k-1}}
\]

**Theorem 1:** Suppose that the assumption 1) is satisfied. Consider every CG method in the form (4), where \( \alpha_k \) is obtained by the exact minimization rules. Then either

\[
\lim_{k \to \infty} \|g_k\| = 0 \text{ or } \sum_{k=0}^{\infty} \left( \frac{g_k^T d_k}{d_k} \right)^2 < \infty
\]

**Proof.** By contradiction, if theorem 1 is not true, there exists a constant \( c > 0 \) such that

\[
\|g_k\| > c
\]
\[ d_k = -g_k + \beta_k d_{k-1} \]
\[ d_k + g_k = \beta_k d_{k-1} \]

Squaring both sides
\[ \|d_k\|^2 + 2g_k^T d_k + \|g_k\|^2 = \|\beta_k d_{k-1}\|^2 \]
\[ \|d_k\|^2 = \|\beta_k d_{k-1}\|^2 - 2g_k^T d_k - \|g_k\|^2 \]  \hspace{1cm} (11)

But \( g_k^T d_k = -e_k \|g_k\|^2 \).

Dividing both sides of (11) by \( (g_k^T d_k)^2 \) given
\[ \frac{\|d_k\|^2}{(g_k^T d_k)^2} = \frac{\|\beta_k d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2g_k^T d_k}{(g_k^T d_k)^2} - \frac{\|g_k\|^2}{(g_k^T d_k)^2} \leq \frac{\|\beta_k\| \|g_{k-1}\|^2}{(g_k^T d_k)^2} + \frac{1}{\|g_k\|^2} \]
\[ \leq \left\{ \frac{g_k^T g_k}{(g_k + d_{k-1})^T d_{k-1}} \right\}^2 \|d_{k-1}\|^2 + \frac{1}{\|g_k\|^2} \]
\[ \leq \left\{ \frac{g_k^T g_k}{g_k^T d_{k-1} + \|d_{k-1}\|^2} \right\}^2 \|d_{k-1}\|^2 + \frac{1}{\|g_k\|^2} \]
\[ \leq \frac{\|g_k\|^2}{\|d_{k-1}\|^2} \|d_{k-1}\|^2 + \frac{1}{\|g_k\|^2} \leq \frac{1}{\|d_{k-1}\|^2} + \frac{1}{\|g_k\|^2} \]  \hspace{1cm} (12)

But note that \( \frac{1}{\|d_k\|^2} = \frac{1}{\|g_k\|^2} \), then from (12) we get
\[ \frac{\|g_k\|^2}{(g_k^T d_k)^2} \leq \frac{1}{\|g_{k-1}\|^2} + \frac{1}{\|g_k\|^2} \leq \sum_{i=0}^{\infty} \frac{1}{\|g_i\|^2} \]
\[ \therefore \frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{c^2}{k} \]  \hspace{1cm} (13)

From (10) and (13) we get
\[ \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \infty \]

This contradicts the Zoutendijk condition in lemma (2) which completes the proof. \( \square \)

4. Numerical Results

In this section we consider the numerical solution for this research. The conjugate gradient method of ME, Dai and Yuan, and Fletcher and Reeves were tested. Some test problems considered in Andrei [8]. We are selected based on the number of iteration and number of function evaluation (Table 1 and Table 2).
**Table 1.** Comparison of the algorithms for \( n = 100. \)

|       | ME              | DY              | FR              |
|-------|-----------------|-----------------|-----------------|
| \( F \) | NoI | Nof | NoI | Nof | NoI | Nof |
| \( F_1 \) | 13  | 24  | 18  | 34  | 19  | 35  |
| \( F_2 \) | 37  | 70  | 68  | 114 | 2001 | 2025 |
| \( F_3 \) | 5   | 13  | 10  | 21  | 32  | 64  |
| \( F_4 \) | 19  | 30  | 17  | 28  | 15  | 25  |
| \( F_5 \) | 61  | 105 | 389 | 694 | 2001 | 2103 |
| \( F_6 \) | 11  | 21  | 8   | 17  | 15  | 31  |
| \( F_7 \) | 37  | 59  | 63  | 98  | 63  | 98  |
| \( F_8 \) | 13  | 29  | 9   | 22  | 11  | 26  |
| \( F_9 \) | 15  | 27  | 38  | 58  | 40  | 65  |
| \( F_{10} \) | 295 | 526 | 1977 | 3735 | 2001 | 2355 |
| \( F_{11} \) | 11  | 20  | 13  | 24  | 13  | 24  |
| \( F_{12} \) | 41  | 72  | 86  | 134 | 121 | 218 |
| \( F_{13} \) | 53  | 104 | 28  | 150 | 69  | 1202 |
| \( F_{14} \) | 285 | 430 | 619 | 962 | 671 | 1066 |
| \( F_{15} \) | 67  | 137 | 40  | 567 | 34  | 57  |
| \( F_{16} \) | 14  | 23  | 10  | 19  | 20  | 33  |
| \( F_{17} \) | 8   | 17  | 7   | 15  | 12  | 25  |
| \( F_{18} \) | 234 | 395 | 486 | 759 | 439 | 743 |
| \( F_{19} \) | 32  | 7   | 3   | 9   | 3   | 7   |
| \( F_{20} \) | 11  | 21  | 8   | 17  | 15  | 31  |
| \( F_{21} \) | 5   | 7   | 9   | 11  | 9   | 11  |
| **Total** | 1267 | 2137 | 3906 | 7488 | 7604 | 10,244 |

**Table 2.** Comparison of the algorithms for \( n = 1000. \)

|       | ME              | DY              | F/R              |
|-------|-----------------|-----------------|-----------------|
| \( F \) | NoI | Nof | NoI | Nof | NoI | Nof |
| \( F_1 \) | 14  | 31  | 38  | 65  | 38  | 65  |
| \( F_2 \) | 137 | 210 | 179 | 292 | 2001 | 2005 |
| \( F_3 \) | 13  | 28  | 15  | 29  | 77  | 129 |
| \( F_4 \) | 13  | 25  | 11  | 19  | 127 | 3531 |
| \( F_5 \) | 139 | 229 | 436 | 878 | 2001 | 2073 |
| \( F_6 \) | 6   | 13  | 7   | 15  | 8   | 17  |
| \( F_7 \) | 41  | 66  | 67  | 105 | 67  | 105 |
| \( F_8 \) | 15  | 35  | 8   | 21  | 16  | 125 |
| \( F_9 \) | 27  | 42  | 39  | 59  | 43  | 68  |
| \( F_{10} \) | 2001 | 2012 | 2001 | 3773 | 2001 | 2066 |
Continued

|   |   |   |   |   |   |
|---|---|---|---|---|---|
|   | 12 | 21 | 14 | 25 | 14 |
| F1 |   |   |   |   |   |
| F2 | 112| 190| 220| 345| 345|
| F3 | 66 | 136| 45 | 642| 98 |
| F4 | 1925 | 2240| 2001| 3140| 2001 | 2897|
| F5 | 63 | 130| 157| 4477| 142 |
| F6 | 17 | 28 | 9 | 18 | 19 | 35|
| F7 | 8 | 17 | 7 | 15 | 11 | 23|
| F8 | 2001| 2098| 2001| 3126| 2001 | 2851|
| F9 | 23 | 9 | 4 | 11 | 4 | 9|
| F10 | 6 | 13 | 7 | 15 | 8 | 17|
| F11 | 6 | 8 | 9 | 11 | 9 | 11|
| Total | 6645| 7581| 7275| 17,081| 11,031 | 22,269|

5. Conclusion

A new kind of parameter in the conjugate gradient method for large scale unconstrained optimization problems is proposed. Numerical results are detected that the new method is superior in practice with competitive DY and FR methods.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Fletcher, R. and Reeves, C.M. (1964) Function Minimization by Conjugate Gradients. The Computer Journal, 7, 149-154. https://doi.org/10.1093/comjnl/7.2.149
[2] Polak, E. and Ribiere, G. (1969) Note sur la convergence de méthodes de directions conjuguées. ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique, 3, 35-43. https://doi.org/10.1051/m2an/196903R100351
[3] Hestenes, M.R. and Stiefel, E. (1952) Methods of Conjugate Gradients for Solving Linear Systems. Journal of Research National Bureau Standards, 49, 409-436. https://doi.org/10.6028/jres.049.044
[4] Dai, Y.-H. and Yuan, Y. (1999) A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property. SIAM Journal on Optimization, 10, 177-182. https://doi.org/10.1137/S1052623497318992
[5] Liu, Y. and Storey, C. (1991) Efficient Generalized Conjugate Gradient Algorithms, Part 1: Theory. Journal of Optimization Theory and Applications, 69, 129-137. https://doi.org/10.1007/BF00940464
[6] Fletcher, R. (1987) Practical Methods of Optimization, Vol. 1, Unconstrained Optimization. Wiley, New York.
[7] Sun, J. and Zhang, J. (2001) Global Convergence of Conjugate Gradient Methods without Line Search. Annals of Operations Research, 103, 161-173.
A List of Test Function

F₁ Extended Trigonometric Function.
F₂ Diagonal 2 function.
F₃ Extended Tridiagonal −1 function.
F₄ Extended Three Exponential Terms.
F₅ Generalized PSC1 function.
F₆ Extended PSC1 Function.
F₇ Extended Block Diagonal BD1 function.
F₈ Extended Quadratic Penalty QP1 function.
F₉ Extended Tridiagonal −2 function.
F₁₀ Nondquar (CUTE).
F₁₁ DIXMAANC (CUTE).
F₁₂ DIXMAANE (CUTE).
F₁₃ EDENSCH function (CUTE).
F₁₄ STAIRCASE S1/F₁₂ VARDIM function (CUTE).
F₁₅ ENGVAL1 (CUTE).
F₁₆ DENSCHNA (CUTE).
F₁₇ DENSCHNB (CUTE).
F₁₈ DIGGSB1 (CUTE).
F₁₉ Diagonal 7.
F₂₀ SINCOS.
F₂₁ HIMMELBG (CUTE).