Proposed CG method to solve unconstrained optimization problems

H. A. Wasi ¹ and M. A. K. Shiker ¹, ²

¹Mathematics Department, College of Education for Pure Sciences, University of Babylon, Iraq.

¹mathman1972@gmail.com

²mmmtmmhh@yahoo.com

Abstract. The Conjugate Gradient (CG) method of unconstrained optimization algorithms possesses good properties like less requirement memory and global convergence properties. Many modified algorithms have been made to this method, as well as new suggestions for its work to obtain the best results. In this article, a modified of conjugate gradient method is introduced, so that global convergence was smoothly proven, and the numerical results of the proposed method were compared with other methods, and the results were better in its general form, this confirms the strength, durability, and effectiveness of the proposed new method.

Keywords. Conjugate gradient method, unconstrained optimization problem, Global convergence.

1. Introduction

Consider the following unconstrained optimization problem

\[ \min_{x \in \mathbb{R}^n} f(x), \quad (1.1) \]

where \( f: \mathbb{R} \rightarrow \mathbb{R}^n \) is differentiable and continuously. The unconstrained optimization problems appeared in many fields, actually the conjugated gradient method is considered one of the perfect methods for solving such problems due to its distinctive characteristics and ease of convergence to the optimal solution. The conjugate gradient method using the following iterative formula

\[ x_{k+1} = x_k + \alpha_k d_k, \quad k = 0,1,2,... \quad (1.2) \]

Where the step size \( \alpha_k \) is positive and \( x_k \) is the current iterative point.

The directions \( d_k \) are computed as [1]

\[ d_k = \begin{cases} -F_k & \quad \text{if } k = 0 \\ -F_k + \beta_k d_{k-1} & \quad \text{if } k \geq 1 \end{cases} \quad (1.3) \]

Where \( F_k = \nabla f(x_k), \beta_k \) is a scalar. The line search in CG algorithms is oftentimes based on the general Wolfe conditions [2]

\[ f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k d_k \quad (1.4) \]

\[ F_{k+1}^T d_k \geq \sigma F_k^T d_k \quad (1.5) \]
Where $d_k$ are a descent direction and $0 < \delta < 1$. However, for some conjugate gradient algorithms, stronger Wolfe (SW) condition used, defined by equation (1.4) and:

$$
|F_k^T d_k| \geq -\sigma F_k^T d_k
$$

(1.6)

Global convergence was proven in conjunction with using a strong wolf Powel (SWP) line search under appropriate conditions to achieve optimal synchronization [3].

The descent property of conjugate gradient algorithms holds if [4, 5]:

$$
F_k^T d_k < 0, \quad \forall k \geq 0.
$$

(1.7)

The descent condition is often used in the analytical aspect to prove the global convergence of a conjugate gradient method with inexact line search and maybe pivotal for conjugate gradient methods.

In recent decades a lot of efforts have been made to generate descent conjugate gradient methods independent of the line search used [6].

The authors had many papers in the field of optimization, transportation problems and line search methods see [7-20], but in this paper, we propose a modified conjugate gradient method to solve unconstrained optimization, the direction generated by the proposed method is always descent direction of the objective function, we use the parameter $\beta_k$ with special changes to get a prorate contribution to solving (1.1).

2. Suggested method

We introduced a new suggestion method by modifying the general CG algorithm by used a new $\beta_k$ and the new direction $d_k$ such that

$$
\beta_k^{md} = -\eta \frac{\|F_k + \beta_k w_k\|^2}{\|d_k\|^2 (F_{k+1} - F_k)}
$$

(2.1)

Where $F_k$ is the derivative of $f(x)$ at $x_k$ and $F_{k+1}$ is the derivative of $f(x_{k+1})$ at $x_{k+1}$, $\eta \in (0,1)$, also

$$
d_k = \begin{cases} 
-F_k & \text{if } k = 0 \\
-F_k + \beta_k w_k & \text{if } k \geq 1 
\end{cases}
$$

(2.2)

where $w_k = x_{k+1} - x_k$.

Now, we provide the following algorithm:

2.1. Algorithm

Step 1. Select a primary point $x_0 \in \mathbb{R}^n$, $\varepsilon = 1 \times 10^{-10}$, $\rho > 0$, $\sigma = 0$, $d_0 = -F_0 = -\nabla f(x_0)$, $k := 0$.

Step 2. if $\|F_{k-1}\| \leq \varepsilon$, then stop, Else go to the next step.

Step 3. Compute $\alpha_k$ from (1.4).

Step 4. Compute $x_{k+1} = x_k + \alpha_k d_k$, then, if $\|F_k\| \leq \varepsilon$, then stop.

Step 5. Compute the search direction $d_k$ by (2.2), where $\beta_k^{md}$ calculated by (2.1).

Step 6. Set $k := k+1$, go to step 3.

Theorem 1:

Suppose that the sequences $\{F_k\}, \{d_k\}$ are generated by algorithm 2.1 then

$$
F_k^T d_k \leq - \left(1 - \frac{1}{4\mu}\right) \|F_k\|^2, \quad \forall k \geq 0,
$$

(2.3)

where $\frac{1}{3} < \mu < 1$.

Proof: see [3]

The above theorem explains the Algorithm 2.1 possesses sufficient conditions in (1.7).

3. Convergence analysis

To prove the global convergence of the CG methods, the following assumptions are needed:

Assumption A:

The set $\Omega = \{x \in \mathbb{R}^n \mid \exists f(x) \leq f(x_0)\}$ is bounded where $x$ is the initial point.

Assumption B:
In some neighborhood $N$ of $\Omega$, assume $F$ be Lipchitz continuous on $\Omega$, i.e., $\exists$ a positive number $L > 0$, such that
\[
\|F[x] - F[y]\| \leq L\|x - y\|, \forall x, y \in \Omega
\] (3.1)

Lemma 1. Suppose the sequence $\{x_n\}$ obtained from the suggestion algorithm, if there exists $\varepsilon > 0$, such that:
\[
\|F_k\| \geq \varepsilon, \forall k
\] (3.2)
Then we have $\theta > 0$ such that:
\[
\|d_k\| \leq \theta
\] (3.3)
Proof:
Since $d_k$ is a descent direction of $f$ at $x_k$ and from (1.4) we have the sequence $\{f(x_k)\}$ is decreasing and from the same equation we get:
\[
\sum_{k=0}^{\infty} a_k \langle d_k, d_k \rangle < \infty
\] (3.4)
If $f$ is bounded then we have
\[
\lim_{k \to \infty} a_k \|d_k\|^2 = 0
\] (3.5)
Also, we can get from assumption A that there exists a constant $\sigma > 0$ such that:
\[
\|F_k\| \leq \sigma, \forall x \in \Omega
\] (3.6)
By the definition of $\beta_k^{md}$, we can simplify its expression as:
\[
\beta_k^{md} = \eta \frac{\|\|F_{k+1}\|\|^2}{\|d_k\|^2 \|F_{k+1}\| + \|F_k\|} = \eta \frac{\|d_k\|^2 \|F_{k+1}\|}{\|F_{k+1}\|^2 + \|F_k\|} < \mu \eta \frac{\|d_k\|^2 \|F_{k+1}\|}{\|F_{k+1}\|^2 + \|F_k\|}
\] (3.7)
By the definition of $d_k = -F_k + \beta_k d_{k-1}$ and the relations (3.5) and (3.7), we get
\[
\|d_k\| = \|F_k\| + \beta_k \|d_{k-1}\| \leq \sigma + \mu \eta \|f_{k+1}\|^2 + \|F_k\| \leq \sigma + \mu \eta \|f_{k+1}\|^2 + \|F_k\| \|d_{k-1}\| \|d_{k-1}\|
\]
Since $\lim_{k \to \infty} \|d_k\|^2 = 0$, then there exists a constant $a \in (0, 1)$ and an integer $k_0$
Such that the next inequality hold for $k \geq k_0$
\[
\|d_k\| \leq \sigma + \mu \eta \|f_{k+1}\|^2 + \|F_k\| \leq a
\]
For all $k \geq k_0$ we have
\[
\theta = \max \left\{\|d_1\|, \|d_2\|, \ldots, \|d_{k_0}\|, \frac{\theta}{\|d_k\|} + \|d_{k_0}\|\right\}, \forall k \geq k_0
\]
Then
\[
\|d_k\| \leq \theta, \forall k \geq k_0
\]

Theorem 2: suppose the sequences $\{x_n\}$ be generated by Algorithm (2.1), then:
\[
\lim_{k \to \infty} \|F_k\| = 0
\] (3.8)
Proof: by contradiction assume that the theorem is not true, then there exists a constant $\varepsilon > 0$ such that
\[
\|F_k\| \geq \varepsilon, \forall k = 0, 1, 2, \ldots
\] (3.9)
If $\lim_{k \to \infty} \inf \alpha_k > 0$ Then from (3.4) we get $\inf \lim_{k \to \infty} \|F_k\| = 0$, that is mean contradicts with (3.9).
Now if $\lim_{k \to \infty} \inf \alpha_k = 0$ there exists an infinite index set $K$ such that:
\[
\lim_{k \in K} \frac{\alpha_k}{\|d_k\|^2} = 0, \text{ since } \alpha_k = \max \{\rho^j, j = 0, 1, 2, \ldots\} \text{ by (1.4) we have}
\]
\[
f(x_k + \rho^{-1} \alpha_k d_k) > f(x_k) + \delta (\rho^{-1} \alpha_k)^2 \|d_k\|^4
\] (3.10)
there exists a constant $\omega \in (0, 1)$ such that:
\[
f(x_k + \rho^{-1} \alpha_k d_k) - f(x_k) = \rho^{-1} F(x_k + \omega \rho^{-1} \alpha_k d_k)^T d_k
\]
Now from (3.1), (2.3) and the concept of the mean – value and with some algebraic processing we get:

$$\rho^{-1} \alpha_k F_k^T d_k + L \rho^{-2} \alpha_k^2 \|d_k\|^2 \leq - \left(1 - \frac{1}{4 \mu}\right) \rho^{-1} \alpha_k \|F_k\|^2 + L \rho^{-2} \alpha_k^2 \|d_k\|^2$$

Substituting the above inequality in (3.10) and from (3.3) we have

$$\left(1 - \frac{1}{4 \mu}\right) \|F_k\|^2 \leq L \rho^{-1} \alpha_k \|d_k\|^2 + \delta \rho^{-1} \alpha_k^2 \|d_k\|^2$$

By dividing both sides of this inequality by $\left(1 - \frac{1}{4 \mu}\right) > 0$ we get:

$$\|F_k\|^2 \leq \frac{\rho^{-1}(L + \delta \alpha_k^2)}{(1 - \frac{1}{4 \mu})} \alpha_k \|d_k\|^2.$$ 

Since $\lim_{k \to 0} \alpha_k \|d_k\|^2 = 0$, then the last inequality implies $\lim_{k \to \infty} \inf \|F_k\| = 0.$

This contradicts with (3.9), then (3.8) is true and $\lim_{k \to \infty} \|F_k\| = 0.$  □

4. Numerical Results

In this part, we present the numerical results of our proposed method coded SM, and compared them with the results of other famous algorithms, which are as follows:

HG [1], AK [3], QD [4]. The results showed a remarkable difference in terms of functions evaluations, number of iterations and processing time.

The following parameters were used: $\rho = 0.7, \sigma = 0.3, \epsilon = 10^{-10}, \eta = 0.02$, and the stop condition $\|F_{k-1}\| \leq 10^{-8}$ with Dim = 500000.

All algorithms including the proposed algorithm have been implemented with MATLAB R2014 and run on PC with the following specifications, 2.5 GHz CPU processor, 12 GB RAM and Windows operation system.

The results are shown in the following Tables 1 and 2:

| problem | f eval | Iter |
|--------|--------|------|
| SM     | HG     | AK   |
| P1     | 42     | 407  | 41   | 163  | 10   | 40   | 10   | 22   |
| P2     | 42     | 407  | 41   | 163  | 10   | 40   | 10   | 22   |
| P3     | 36     | 399  | 58   | 153  | 8    | 38   | 15   | 19   |
| P4     | 20     | 387  | 99   | 21   | 4    | 36   | 27   | 6    |
| P5     | 42     | 407  | 41   | 163  | 10   | 40   | 10   | 22   |
| P6     | 45     | 113  | 107  | 12   | 11   | 15   | 30   | 2    |
| P7     | 36     | 399  | 58   | 153  | 8    | 38   | 15   | 19   |
| P8     | 36     | 119  | 68   | 270  | 8    | 17   | 18   | 31   |
| P9     | 31     | 323  | 222  | 78   | 7    | 32   | 64   | 24   |
| P10    | 40     | 110  | 160  | 597  | 9    | 20   | 43   |      |
| P11    | 52     | 99   | 117  | 947  | 12   | 6    | 31   | 114  |
| P12    | 36     | 130  | 135  | 279  | 8    | 19   | 36   | 44   |
| P13    | 40     | 276  | 169  | 389  | 9    | 31   | 46   | 65   |
| P14    | 32     | 92   | 117  | 138  | 7    | 15   | 31   | 32   |
| P15    | 353    | 2827 | 360  | 261  | 96   | 188  | 109  |      |
| P16    | 2196   | 2883 | 303  | 268  | 629  | 202  | 93   | 64   |

Table 1. Functions evaluations (f - eval) & iterations (Iter).
Table 2. CPU-Time (in seconds).

| problem | CPU-Time |
|---------|----------|
| SM      | HG       | AK       | QD       |
| P1      | 0.234375 | 2.15625  | 0.21875  | 0.71875  |
|         | 0.15625  | 2.015625 | 0.171875 | 0.6875   |
|         | 0.125    | 2.125    | 0.234375 | 0.734375 |
|         | 0.21875  | 2.109375 | 0.28125  | 0.65625  |
|         | 0.09375  | 2.0625   | 0.421875 | 0.078125 |
|         | 0.28125  | 2.171875 | 0.234375 |          |
| P2      | 0.28125  | 0.625    | 0.515625 | 0.65625  |
|         | 0.140625 | 2.1875   | 0.296875 | 0.03125  |
|         | 0.125    | 0.6875   | 0.40625  | 1.171875 |
|         | 0.140625 | 1.96875  | 1.09375  | 0.4375   |
|         |          | 0.15625  | 0.171875 | 0.65625  |
| P3      | 0.046875 | 0.15625  | 0.15625  | 0.90625  |
|         | 0.0625   | 0.203125 | 0.171875 | 0.265625 |
|         | 0.03125  | 0.296875 | 0.25     | 0.421875 |
|         | 0.046875 | 0.15625  | 0.125    | 0.15625  |
|         | 0.25     | 0.296875 |          | 0.171875 |
| P4      | 1.53125  | 1.578125 | 0.21875  | 0.28125  |
|         | 0.1875   | 1.640625 | 0.296875 | 0.15625  |
|         | 0.28125  | 1.5625   | 0.234375 | 0.25     |
|         | 0.5      | 1.59375  | 0.1875   | 0.15625  |
|         | 0.34375  |          | 0.90625  | 79.46875 |
| P5      | 0.359375 | 3.359375 | 0.90625  | 145.3438 |
|         | 0.25     | 3.796875 | 0.609375 | 14.65625 |
|         | 0.296875 | 3.8125   | 0.953125 | 37.40625 |
|         | 0.28125  | 3.734375 | 0.765625 | 28.78125 |

5. Conclusions

In this paper, we suggested a modified general conjugate gradient algorithm. The global convergence of this modification has been proven. The comparison of the numerical results of the proposed method with other three famous algorithms proved the success of the proposed algorithm, where, in general, the numbers of iterations, functions evaluation and the CPU time needed in the new method to reach the required solution are less than that needed in the other algorithms.
References

[1] Andrei, N., (2008), A Dai–Yuan conjugate gradient algorithm with sufficient descent and conjugacy conditions for unconstrained optimization, Applied Mathematics Letters, vol. 21, pp. 165–171.

[2] Abbo, K. K and Abdul Wahid, L A., (2017), Generalized Dai–Yuan nonlinear conjugate gradient method for unconstrained optimization, International Journal of recent Scientific research, vol. 8, pp. 17993-17999.

[3] Li, C., (2013), A Modified Conjugate Gradient Method for Unconstrained Optimization, TELKOMNIKA, vol.11, pp. 6373-6380.

[4] Wasi, H. A. and Shiker, M. A. K. (2020), A new conjugate gradient method for solving large scale systems of monotone equations, International Journal of Advanced Science and Technology, vol. 29: 4, pp.2303-2314.

[5] http://sersc.org/journals/index.php/IJAST/article/view/20318

[6] Wasi, H. A. and Shiker, M. A. K., (2020), A new hybrid CGM for unconstrained optimization problems, “in press”, accepted paper for publication in Journal of Physics: Conference Series, 1st International Virtual Conference on Pure Sciences (IVCPS)- Iraq.

[7] Mahdi, M. M. and Shiker, M. A. K., (2020), Three-Term of New Conjugate Gradient Projection Approach under Wolfe Condition to Solve Unconstrained Optimization Problems, Journal of Advanced Research in Dynamical and Control Systems, vol. 12: 7, pp. 788- 795. 10.5373/JARDCS/V12I7/20202063

[8] Shiker, M. A. K., & Sahib, Z. (2018). A modified trust-region method for solving unconstrained optimization. Journal of Engineering and Applied Sciences, vol. 13: 22, pp. 9667–9671. https://doi.org/10.3923/jeasci.2018.9667.9671

[9] Hassan, Z. A. H. H. and Shiker, M. A. K., (2018), Using of generalized baye’s theorem to evaluate the reliability of aircraft systems, Journal of Engineering and Applied Sciences, (Special Issue 13), pp. 10797-10801. https://doi.org/10.36478/jeasci.2018.10797.10801

[10] Shiker, M. A. K. and Amini, K., (2018), A new projection-based algorithm for solving a large scale nonlinear system of monotone equations, Croatian operational research review, vol. 9: 1, pp. 63-73. https://doi.org/10.17535/corr.2018.0006

[11] Mahdi, M. M. and Shiker, M. A. K., (2020), A new projection technique for developing a Liu-Storey method to solve nonlinear systems of monotone equations, Journal of Physics: Conference Series, 1591, The Fifth International Scientific Conference of Al-Khwarizmi Society (FISCAS) 26-27 June 2020, Iraq. 10.1088/1742-6596/1591/1/012030

[12] Mahdi, M. M. and Shiker, M. A. K., (2020), Three terms of derivative free projection technique for solving nonlinear monotone equations, Journal of Physics: Conference Series, 1591, The Fifth International Scientific Conference of Al-Khwarizmi Society (FISCAS) 26-27 June 2020, Iraq. 10.1088/1742-6596/1591/1/012031

[13] Mahdi, M. M. and Shiker, M. A. K., (2020), Solving systems of nonlinear monotone equations by using a new projection approach, “in press”, accepted paper for publication in Journal of Physics: Conference Series, 1st International Virtual Conference of the University of Babylon (IVCUB)- Iraq.

[14] Mahdi, M. M. and Shiker, M. A. K., (2020), A New Class of Three-Term Double Projection Approach for Solving Nonlinear Monotone Equations, “in press”, accepted paper for publication in Journal of Physics: Conference Series,1st International Virtual Conference on Pure Sciences (IVCPS)- Iraq.

[15] Dwail, H. H. and Shiker, M. A. K., (2020), Using a trust region method with nonmonotone technique to solve unrestricted optimization problem, “in press”, accepted paper for publication in Journal of Physics: Conference Series, 1st International Virtual Conference on Pure Sciences (IVCPS)- Iraq.

[16] Dwail, H. H., Mahdi, M. M., Wasi, H. A., Hashim, K. H., Dreeb, N. K., Hussein, A. H. and Shiker,
M. A. K. (2020), A new modified TR algorithm with adaptive radius to solve a nonlinear systems of equations, "in press", accepted paper for publication in Journal of Physics: Conference Series, 1st International Virtual Conference of the University of Babylon (IVCUB)- Iraq.

[17] Dwail, H. H. and Shiker, M. A. K., (2020), Using a new trust region algorithm with nonmonotone adaptive radius for solving nonlinear systems of equations, International Journal of Advanced Science and Technology, 29: 4, pp.2351- 2360. http://sersc.org/journals/index.php/IJAST/article/view/20325

[18] Dwail, H. H. and Shiker, M. A. K., (2020), Reducing the time that TRM requires to solve systems of nonlinear equations, “in press”, accepted paper for publication in Journal of Physics: Conference Series, 2nd International Scientific Conference of Al-Ayen University (ISCAU)- Iraq.

[19] Hussein, H. A. and Shiker, M. A. K., (2020), A modification to Vogel’s approximation method to Solve transportation problems, Journal of Physics: Conference Series, Volume 1591, The Fifth International Scientific Conference of Al-Khwarizmi Society (FISCAS) 2020 26-27 June 2020, Iraq. 10.1088/1742-6596/1591/1/012029

[20] Hussein, H. A., Shiker, M. A. K. and Zabiba, M. S. M., (2020), A new revised efficient of VAM to find the initial solution for the transportation problem, Journal of Physics: Conference Series, Volume 1591, The Fifth International Scientific Conference of Al-Khwarizmi Society (FISCAS) 2020 26-27 June 2020, Iraq. 10.1088/1742-6596/1591/1/012032

[21] Hussein, H. A. and Shiker, M. A. K., (2020), Two New Effective Methods to Find the Optimal Solution for the Assignment Problems, Journal of Advanced Research in Dynamical and Control Systems, 12 (7), p. 49- 54. 10.5373/JARDCS/V12I7/20201983