Mathematics and Statistics 10(3): 610-614, 2022 http://www.hrpub.org
DOI: 10.13189/ms.2022.100317

A Descent Conjugate Gradient Method With Global Converges Properties for Non-Linear Optimization

Salah Gazi Shareef

Department of Mathematics, Faculty of Science, University of Zakho, Kurdistan Region, Iraq

Received February 14, 2022; Revised April 22, 2022; Accepted May 23, 2022

Cite This Paper in the following Citation Styles

(a): *[1] Salah Gazi Shareef , "A Descent Conjugate Gradient Method With Global Converges Properties for Non-Linear Optimization," Mathematics and Statistics, Vol. 10, No. 3, pp. 610 - 614, 2022. DOI: 10.13189/ms.2022.100317.*

(b): *Salah Gazi Shareef (2022). A Descent Conjugate Gradient Method With Global Converges Properties for Non-Linear Optimization. Mathematics and Statistics, 10(3), 610 - 614. DOI: 10.13189/ms.2022.100317.*

Abstract Iterative methods such as the conjugate gradient method are well known methods for solving non-linear unconstrained minimization problems partially because of their capacity to handle large-scale unconstrained optimization problems rapidly, and partly due to their algebraic representation and implementation in computer programs. The conjugate gradient method has wide applications in a lot of fields such as machine learning, neural networks and many other fields. Fletcher and Reeves [1] expanded the approach to nonlinear problems in 1964. It is considered to be the first nonlinear conjugate gradient technique. Since then, lots of new other conjugate gradient methods have been proposed. In this work, we will propose a new coefficient conjugate gradient method to find the minimum of the non-linear unconstrained optimization problems based on parameter of Hestenes Stiefel. Section one in this work contains the derivative of new method. In section two, we will satisfy the descent and sufficient descent conditions. In section three, we will study the property of the global convergence of the new proposed. In the fourth section, we will give some numerical results by using some known test functions and compare the new method with Hestenes S. to demonstrate the effectiveness of the suggestion method. Finally, we will give conclusions.

Keywords Non-Linear Minimization, Algorithm of Conjugate Gradient, Descent property and Global Convergence Property

1. Introduction

Below is the nonlinear unconstrained minimization problem, consider it:

$$\text{Min. } f(x)\,;\; x \in R^n \tag{1.1}$$

where $f : R^n \to R$ is a continuously differentiable, real-valued function. We used the iterative method

$$x_{k+1} = x_k + \alpha_k\, d_k \tag{1.2}$$

for solving the problem (1.1), and the iterative method is starting with an initial guess x_1 belongs to R^n, where $v_k = \alpha_k d_k = x_{k+1} - x_k$, the positive step length α_k is computed by one dimensional line search and d_k is search direction. The search direction of the steepest descent, has the form

$$d_1 = -g_1 \tag{1.3}$$

The equation below is to calculate the next search directions:

$$d_{k+1} = -g_{k+1} + \beta_k\, d_k \tag{1.4}$$

Where, $g_k = \nabla f(x_k)$ and β_k is a scalar. The basic parameters of β_k , Hestenes S. (HS) [2], Polak R. Polyak (PRP) [3], Fletcher R.(FR) [4], Dai and Yuan (DY) [5], Dai and Liao [6], Perry [7], and Liu andStorey [8], which are shown below:

$$\beta_k^{HS} = \frac{g_{k+1}^T(g_{k+1}-g_k)}{d_k^T(g_{k+1}-g_k)} \tag{1.5}$$

$$\beta_k^{PRP} = \frac{g_{k+1}^T(g_{k+1}-g_k)}{\|g_k\|^2} \tag{1.6}$$

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \quad (1.7)$$

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k} \quad (1.8)$$

$$\beta_k^{DL} = \frac{g_{k+1}^T(y_k - tv_k)}{d_k^T y_k}, \ t > 0 \quad (1.9)$$

$$\beta_k^{Perry} = \frac{g_{k+1}^T(y_k - v_k)}{d_k^T y_k} \quad (1.10)$$

$$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k} \quad (1.11)$$

where $y_k = g_{k+1} - g_k$ and the Euclidean norm of vectors is denoted by the symbol. $\|.\|$. The property of the convergence about Fletcher R. (FR) method, Polak R. Polyak(PRP) method, Hestenes S.(HS) method, Dai Yuan (DY) method, and Liu Storey (LS) method can be seen from [9] [10][11][12][13] [8].

Lots of parameters are suggested. Hussein A. Kh. and Salah G. Sh. [14][15] suggested some parameters of conjugate gradient methods. Hager and Zhang [16] suggested another conjugate gradient method called CG-DESCENT method.

2. Derivative of New Formula β_k^{New}

Consider the parameter of Hestenes Stiefel

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \quad (2.1)$$

Suppose that

$$y_k^* = g_{k+1} + (1 - p_1)\left(\left(\frac{g_{k+1}}{p_2}\right) - p_3 g_{k+1}\right) \quad (2.2)$$

Where, $p_1 \in (0,1)$, $p_2 = \frac{2\sqrt{\mu}}{\|v_k\|}(1 + \|x_{k+1}\|)$ and μ is the machine error, and $p_3 = 0.1$

Here we replace the vector y_k in the numerator of (1.1) by y_k^*, to get

$$\beta_k^{NEW} = \frac{g_{k+1}^T y_k^*}{d_k^T y_k} \quad (2.3)$$

Then,

$$\beta_k^{NEW} = \frac{g_{k+1}^T(g_{k+1} + (1-p_1)((\frac{g_{k+1}}{p_2}) - p_3 g_{k+1})}{d_k^T y_k} \quad (2.4)$$

Implies that

$$\beta_k^{NEW} = \frac{\|g_{k+1}\|^2 + (1-p_1)\frac{\|g_{k+1}\|^2}{p_2} - (1-p_1)p_3\|g_{k+1}\|^2}{d_k^T y_k} \quad (2.5)$$

2.1. Algorithm of the New Method

Step1: Select x_1 and $\partial = 10^{-5}$.
Step2: Set $d_1 = -g_1$, $g_k = \nabla f(x_k)$, Set $k = 1$.
Step3: Calculate the step size $\alpha_k > 0$ satisfy the Wolfe condition

$$f(x_k + \alpha_k d_k) - f(x_k) \le l_1 \alpha_k g_k^T d_k$$

$$|g_{K+1}^T d_k| \le l_2 |g_k^T d_k|$$

where, $0 < l_1 < l_2 < 1$.
Step4: Calculate $x_{k+1} = x_k + \alpha_k d_k$

$g_{k+1} = \nabla f(x_{k+1})$, $\quad If \ \|g_{k+1}\| \le \partial$, stop.

Step5: Calculate β_k^{NEW} by (2.5)
Step6: Compute $d_{k+1} = -g_{k+1} + \beta_k^{NEW} d_k$
Step7: If $|g_{k+1}^T g_k| > 0.2\|g_{k+1}\|^2$, then go to step2.
Else
$k = k + 1$, go to step3.

2.2. Descent Property of the New Method

The property of the descent is that any search direction (1.4) must satisfy the following condition:

$$d_{k+1}^T g_{k+1} \le 0.$$

Theorem 1:- If (1.2) generates the sequence $\{x_k\}$, then the descent condition is satisfied by the search direction in (1.4) with new conjugate gradient (2.5).

Proof:- From equations (1.4) and (2.5) we have the following

$$d_{k+1} = -g_{k+1} + \left(\frac{\|g_{k+1}\|^2 + (1-p_1)\frac{\|g_{k+1}\|^2}{p_2} - (1-p_1)p_3\|g_{k+1}\|^2}{d_k^T y_k}\right)d_k \quad (2.6)$$

Here,

$$d_{k+1} = -g_{k+1} + \left(\frac{\|g_{k+1}\|^2}{d_k^T y_k} + \frac{(1-p_1)}{p_2}\frac{\|g_{k+1}\|^2}{d_k^T y_k} - (1-p_1)p_3\frac{\|g_{k+1}\|^2}{d_k^T y_k}\right)d_k \quad (2.7)$$

Multiply above equation by g_{k+1} from right, to get

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{d_k^T y_k} d_k^T g_{k+1} + \frac{(1-p_1)}{p_2}\frac{\|g_{k+1}\|^2}{d_k^T y_k} d_k^T g_{k+1} - (1-p_1)p_3 \frac{\|g_{k+1}\|^2}{d_k^T y_k} d_k^T g_{k+1} \quad (2.8)$$

If the step size is determined by an accurate line search, $d_k^T g_{k+1} = 0$, here we get $d_{k+1}^T g_{k+1} \le 0$.
If we have an inexact line search which is $d_k^T g_{k+1} \ne 0$.
The equation (2.8) implies that

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \left[1 + \frac{(1-p_1)}{p_2} - (1-p_1)p_3\right]\frac{\|g_{k+1}\|^2}{d_k^T y_k} d_k^T g_{k+1} \quad (2.9)$$

Assume that

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + p_4 \frac{\|g_{k+1}\|^2}{d_k^T y_k} d_k^T g_{k+1} \quad (2.10)$$

Where, $p_4 = [1 + \frac{(1-p_1)}{p_2} - (1-p_1)p_3]$ which is positive

Since $(1 + \frac{(1-p_1)}{p_2}) > (1-p_1)p_3$, Now, we know that the parameters of HS satisfy the descent condition, then, the equation (2.10) is less than or equal to zero,

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + p_4 \frac{\|g_{k+1}\|^2}{d_k^T y_k} d_k^T g_{k+1} \leq 0. \quad (2.11)$$

2.3. Convergence Analysis

Assume that:

a) $S = \{x : x \in R^n, f(x) \leq f(x_1)\}$ is a bounded level set, where x_1 is the initial point.
b) f is continuously differentiable in the vicinity Ω of S, and its gradient g is Lipschitz continuously, i.e. there exist a constant L>0 such that

$$\|g(x) - g(x_k)\| \leq L\|x - x_k\| \ , \forall x, \quad (2.12)$$

On the basis of assumptions a and b on f there exists a constant $\rho \geq 0$, such that $\|g(x)\| \leq \rho, \quad \forall x \in S$.

Lemma1: Suppose that the assumptions (a) and (b) are correct and consider any conjugate gradient (1.2), (1.3) and (1.4), where is a descent direction d_k and α_k is obtained by the strong Wolfe conditions.

$$f(x_k + \alpha_k d_k) - f(x_k) \leq l_1 \alpha_k g_k^T d_k \quad (2.13)$$

$$|g_{K+1}^T d_k| \leq l_2 g_k^T d_k \quad (2.14)$$

If

$$\sum_{k\geq 1} \frac{1}{\|d_k\|^2} = \infty \quad (2.15)$$

Then

$$\lim_{k\to\infty} inf\|g_k\| = 0 \quad (2.16)$$

See ([12]).

There exists a constant ϑ >0 such that if f is a uniformly convex function:

$$\left(g(x) - g(y)\right)^T (x - y) \geq \vartheta\|x - y\|^2 \in \Omega \quad (2.17)$$

We can rewrite (2.16) in the following manner:

$$y_k^T v_k \geq \vartheta\|v_k\|^2 \quad (2.18)$$

Theorem 2: Suppose that the assumptions a and b hold and function is a uniformly convex. The new algorithm of the form (1.2), (1.4) and (2.3), where d_k satisfies the descent property and α_k is obtained by strong Wolfe conditions (2.13) and (2.14) satisfies the global convergence

(i.e.) $\lim_{k\to\infty} inf\|g_{k+1}\| = 0$

Proof: Since

$$d_{k+1} = -g_{k+1} + \beta_k^{New} d_k \quad (2.19)$$

$$|\beta_k^{New}| = \left|\frac{\|g_{k+1}\|^2 + (1-p_1)\frac{\|g_{k+1}\|^2}{p_2} - (1-p_1)p_3\|g_{k+1}\|^2}{d_k^T y_k}\right| \quad (2.20)$$

Implies that

$$|\beta_k^{New}| \leq \left|\frac{\|g_{k+1}\|^2}{d_k^T y_k}\right| + \left|\frac{(1-p_1)\|g_{k+1}\|^2}{p_2\, d_k^T y_k}\right| + \left|\frac{(1-p_1)p_3\|g_{k+1}\|^2}{d_k^T y_k}\right| \quad (2.21)$$

Since $y_k^T d_k \geq \frac{\vartheta\|v_k\|^2}{\alpha_k}$, then we can write equation (2.21) as follows

$$|\beta_k^{New}| \leq \frac{\alpha_k\|g_{k+1}\|^2}{\vartheta\|v_k\|^2} + \frac{\alpha_k(1-p_1)\|g_{k+1}\|^2}{p_2\, \vartheta\|v_k\|^2} + \frac{\alpha_k(1-p_1)p_3\|g_{k+1}\|^2}{\vartheta\|v_k\|^2} \quad (2.22)$$

Since $\|g(x)\| \leq \rho$, then, we have

$$|\beta_k^{New}| \leq \frac{\alpha_k\rho^2}{\vartheta\|v_k\|^2} + \frac{\alpha_k(1-p_1)\rho^2}{p_2\, \vartheta\|v_k\|^2} + \frac{\alpha_k(1-p_1)p_3\rho^2}{\vartheta\|v_k\|^2}$$

Here, $\|d_{k+1}\| \leq \|g_{k+1}\| + |\beta_k^{New}|\|d_k\|$

Then,

$$\|d_{k+1}\| \leq \rho + \left(\frac{\alpha_k\rho^2}{\vartheta\|v_k\|^2} + \frac{\alpha_k(1-p_1)\rho^2}{p_2\, \vartheta\|v_k\|^2} + \frac{\alpha_k(1-p_1)p_3\rho^2}{\vartheta\|v_k\|^2}\right)\|d_k\| \quad (2.23)$$

We know that $\|d_k\| = \frac{\|v_k\|}{\alpha_k}$, so, we get

$$\|d_{k+1}\| \leq \rho + \left(\frac{\rho^2}{\vartheta\|v_k\|} + \frac{(1-p_1)\rho^2}{p_2\, \vartheta\|v_k\|} + \frac{(1-p_1)p_3\rho^2}{\vartheta\|v_k\|}\right)$$

Since, $\|v_k\| = \|x - x_k\|$,

$$D = \max\{\|x - x_k\|\}, \forall x, x_k \in R\}$$

Here,

$$\|d_{k+1}\| \leq \rho + \left(\frac{\rho^2}{\vartheta D} + \frac{(1-p_1)\rho^2}{p_2\, \vartheta D} + \frac{(1-p_1)p_3\rho^2}{\vartheta D}\right) \quad (2.24)$$

Above inequality becomes

$$\|d_{k+1}\| \leq \rho + \left(\frac{\rho^2}{\vartheta D} + \frac{(1-p_1)\rho^2}{p_2\, \vartheta D} + \frac{(1-p_1)p_3\rho^2}{\vartheta D}\right) = \varphi$$

Then,

$$\sum_{k\geq 1} \frac{1}{\|d_{k+1}\|^2} \geq \sum_{k\geq 1} \frac{1}{\varphi^2} = \sum_{k\geq 1} 1 = \infty$$

And

$\sum_{k\geq 1} \frac{1}{\|d_{k+1}\|^2} = \infty$. By using lemma (1), we get $\lim_{k\to\infty} inf\|g_{k+1}\| = 0$.

3. Numerical Results

The suggested method is compared to the Conjugate Gradient algorithm (H/S), with nonlinear unconstrained problems (standard test function) of different dimensions. Programs are written in FORTRAN 90, and the requirement for stopping is $\|g_{k+1}\| \leq 10^{-5}$ in all cases. The number of the functions (NOF) and the number of the iterations (NOI) are shown in tables (1) and (2). Tables (1) and (2) show that the suggested technique outperforms classical Conjugate Gradient methods (HS).

Table 1. Algorithms Performance in Comparison (HS and Suggestion Method)

Test functions	Dim.	HS Algorithm		Suggestion Algorithm	
		NOI	NOF	NOI	NOF
Powell	4	38	108	18	47
	100	40	122	23	63
	500	41	124	22	65
	1000	41	124	22	66
	3000	41	124	22	64
	5000	41	124	22	61
Cubic	4	12	35	14	45
	100	13	37	10	31
	500	13	37	10	31
	1000	13	37	10	31
	3000	13	37	8	26
	5000	13	37	10	31
Edger	4	5	14	5	14
	100	5	14	5	14
	500	6	16	5	14
	1000	6	16	5	14
	3000	6	16	5	14
	5000	6	16	5	14
Rosen	4	30	83	27	90
	100	30	83	20	59
	500	30	83	22	64
	1000	30	83	16	51
	3000	30	83	16	51
	5000	30	83	17	53
GCentral	4	22	159	12	67
	100	22	159	21	179
	500	23	171	16	120
	1000	23	171	16	121
	3000	27	234	22	209
	5000	28	148	21	196
Miele	4	28	85	23	81
	100	33	114	21	75
	500	40	146	17	58
	1000	46	176	16	62
	3000	54	211	17	60
	5000	54	211	17	59
Non-Digonal	4	24	64	20	64
	100	29	79	19	56
	500	F	F	27	85
	1000	29	79	24	74
	3000	29	79	21	66
	5000	30	81	24	77
Wolfe	4	11	24	16	33
	100	49	99	43	87
	500	52	105	47	96
	1000	70	141	49	100
	3000	170	351	156	329
	5000	165	348	140	296
Wood	4	30	68	27	62
	100	30	68	28	64
	500	30	68	26	61
	1000	30	68	28	65
	3000	30	68	28	64
	5000	30	68	28	64
Total		1821	5479	1309	4043

Notes: a) F in the above table means failure, and for the summation we took F = 50 for NOI and F = 100 for NOF.

b) $p_3 = 0.1$ for all the test functions in the above table except Powell where we took $p_3 = 1$.

Table 2. Percentage of the New Method Improving

	HS Algorithm	Suggestion Algorithm
NOI	100 %	71.8835804503%
NOF	100 %	73.7908377441%

4. Conclusions

We proposed in this paper, a new coefficient conjugate gradient algorithm for unconstrained nonlinear minimization. We studied the proposed method descent condition and global convergence, ran numerical tests on low and high dimensionality problems, and made comparisons between different test functions. The new method has proven its efficiency through results in tables (1) and (2).

REFERENCES

[1] Fletcher R., Reeves C. M., "Function minimization by conjugate gradients," Comput. J., 7, 2, 149–154, 1964.

[2] Steifel, M.R.H. & E., "Method of conjugate gradient for solving linear equations," J.Res. Nat. Bur. Stand., 49, 409–436, 1952

[3] Polak E., G.R., "Note sur la convergence de directions conjugees," Rev. Fr. Inform. Rech. Oper., 3,35–43, 1969.

[4] Reeves, R.F., C.M., "Function minimization by conjugate gradients," Comput. Journal, 7, 149–154, 1964.

[5] Dai Y.H., Yuan, Y., "A nonlinear conjugate gradient with a strong global convergence properties," SIAM J. Optim., 10, 177–182, 1999.

[6] Dai Y.H. L., "New conjugacy conditions and related nonlinear conjugate gradient methods," Appl. Math. Optim., 43, 87–101, 2001.

[7] Perry, A., "A Modified Conjugate Gradient Algorithm," JSTOR, 26, 1073–1078, 1978.

[8] Liu Y., Storey, C., "Efficient generalized conjugate gradient algorithms," Part 1: Theory. JOTA, 69, 129–137, 1991.

[9] Birgin E.G., J.M.M., "A spectral conjugate gradient method for unconstrained optimization," Appl. Math. Optim., 43, 117–128, 2001.

[10] Gilbert J.C., J.N., "Global Convergence properties of conjugate gradient methods for optimization," SIAM J. Optim., 2, 21–24, 1992,

[11] Sun J., J.Z., "Convergence of conjugate gradient methods without line search," Ann. Oper. Res., 103, 161–173, 2001.

[12] Zhang L., Zhou, W., D.L., "Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search," Numer. Math., 104, 561–572, 2006.

[13] Raydan, M., "The Barzilain and Borwein gradient method for the large unconstrained minimization problem," SIAM J. Optim., 7, 26–33, 1997.

[14] Hussein A. Kh., Salah G.S., "A New Parameter Conjugate Gradient Method Based on Three Terms Unconstrained Optimization," Gen. Lett. Math., 7,1, 39–44, 2019.

[15] Hussein A. Kh., Salah G.S., "A New conjugate gradient method for unconstrained optimization problems with descent property," Gen. Lett. Math., 9,2, 101-105, 2020.

[16] Hager W., Z.H.C., "A new conjugate gradient method with guaranteed descent and an efficient line search," SIAM J. Optim., 16, 170–192,2005,

[17] Zhang L., Zhou W.J., L.D.H., "A descent modified Polak-Ribiere-Polyak conjugate gradient method and its global convergence," IMA J. Numer. Anal., 26, 629–640, 2006.

[18] Zhang L., Zhou W.J., L.D.H., "Global convergence of a modified Fletcher-Reeves conjugate method with Armijo type line search," Numer. Math., 104, 561-572, 2006.