New Trends in Mathematical Sciences

http://dx.doi.org/10.20852/ntmsci.2022.483

A New Version Coefficient of Three-Term Conjugate Gradient Method to Solve Unconstrained Optimization

Alaa Luqman Ibrahim[1] *and Banaz Hamza Jahwar*[2]

[1]Department of Mathematics, Faculty of Science, University of Zakho, Zakho, Kurdistan Region, Iraq.
[2]Department of Mathematics, College of Science, University of Duhok, Duhok, Kurdistan Region, Iraq.

Received: 9 February 2022, Accepted: 1 July 2022
Published online: 31 October 2022.

Abstract: This paper presents the new three-term conjugate gradient method for solving unconstrained optimization problem. The main aim is to upgrade the search direction of conjugate gradient method to present a more active new three term method. Our new method satisfies the descent and the sufficient descent conditions and global convergent property. Furthermore, the numerical results show that the new method has a better numerical performance in comparison with the standard (PRP) method from an implementation of our new method on some test functions of unconstrained optimization according to number of iterations (NOI) and the number of functions evaluation (NOF).

Keywords: Unconstrained optimization, three-term Conjugate gradient method, descent and sufficient descent conditions and global convergent.

1 Introduction

The main idea of the unconstrained optimization problems is minimizing an objective function that depends on real variable, with no restrictions at all on the values of these variables. The mathematical formulation is

$$min f(x) \qquad \forall x \in R^n \tag{1}$$

where $f : R^n \to R$ is a smooth function and its gradient g is available, we start from the initial guess point x_0 to solve the problem (1). A nonlinear conjugate gradient (CG) method is an iterative scheme that usually generated a sequence $\{x_i\}$ of an approximation to the solution of the problem (1), using the repetition:

$$x_{i+1} = x_i + \lambda_i d_i, \quad i = 0, 1, 2, \ldots, \tag{2}$$

where $\lambda_i > 0$ is called a step size determined by some line searches (Goldstein condition, wolfe condition, curvature condition and sufficient decent condition) and the search direction d_i is designed by:

$$d_{i+1} = \begin{cases} -g_{i+1} & if \quad i = 0 \\ -g_{i+1} + \beta_i \, d_i & if \quad i \geq 1 \end{cases} \tag{3}$$

where $g_i = \nabla f \, (x_i)$ and β_i is an important parameter. The different choices for the parameter β_i correspond to different CG methods. For examples, Polak-Ribiere-Polyak (PRP) [8], Hestenes and Stiefel (HS) [5], Fletcher and Reeves (FR) [4], Alaa et al. method [2], Some well-known formulas for β_i given below:

$$d_{i+1} = -g_{i+1} + (\beta_i^{PRP} = \frac{g_{i+1}^T y_i}{g_i^T g_i}) \, d_i \tag{4}$$

* Corresponding author e-mail: alaa.ibrahim@uoz.edu.krd,banaz.jahwar@uod.ac

$$d_{i+1} = -g_{i+1} + (\beta_i^{HS} = \frac{g_{i+1}^T y_i}{d_i^T y_i})\, d_i \tag{5}$$

$$d_{i+1} = -g_{i+1} + (\beta_i^{FR} = \frac{g_{i+1}^T g_{i+1}}{g_i^T g_i})\, d_i \tag{6}$$

$$d_{i+1} = -g_{i+1} + (\beta_k = \frac{g_{i+1}^T y_i}{g_i^T g_i} - t\frac{g_{i+1}^T v_i}{g_i^T g_i})\, d_i \tag{7}$$

where $y_i = g_{i+1} - g_i$ and $v_i = \lambda_i d_i = x_{i+1} - x_i$.In the convergence analysis and implementation of conjugate gradient method, one often requires the exact and inexact line search such as the Wolfe conditions or the strong Wolfe conditions. The Wolfe line search is to find λ_i such that

$$f(x_i + \lambda_i d_i) \le f(x_i) + \theta \lambda_i g_i^T d_i \tag{8}$$

$$d_i^T g(x_i + \lambda_i d_i) \ge \sigma g_i^T d_i \tag{9}$$

with $0 < \theta < \sigma$. The strong Wolfe line search is to find λ_i such that

$$f(x_i + \lambda_i d_i) \le f(x_i) + \theta \lambda_i g_i^T d_i \tag{10}$$

$$\left| d_i^T g(x_i + \lambda_i d_i) \right| \le \sigma g_i^T d_i \tag{11}$$

where $0 < \theta < \sigma < 1$ are constants, by Li and Weijun, [6]. There are other types of conjugate gradient methods called three-term conjugate gradient methods. That it is one of the most reliable and attractive method and this method numerically stronger than classical conjugate gradient method. Among the generated three-term conjugate gradient methods in the literature we have the three-term conjugate methods proposed by Nazareth [7] present a computationally efficient three-term nonlinear conjugate gradient method with the search direction:

$$d_{i-1} = -y_i + \frac{y_i^T y_i}{y_i^T d_i} d_i - \frac{y_{i-1}^T y_i}{y_{i-1}^T d_{i-1}} d_{i-1} \tag{12}$$

Alaa L. I., and Salah G. Sh., in (2019) [1], submit a new class of three-term conjugate gradient method as:

$$d_{i+1} = \begin{cases} -g_{i+1} & if \quad i = 0 \\ -g_{i+1} + \beta_i^{HS} d_i - t_i (\frac{g_{i+1}^T d_i}{d_i^T y_i}) y_i & if \quad i \ge 1 \end{cases} \tag{13}$$

where $t_i = \gamma \frac{\|y_i\|}{\|s_i\|} + (1-\gamma) \frac{s_i^T y_i}{\|s_i\|^2}$ and the parameter $\beta^k = \beta_i^{PRP}$, β_i^{HS} or β_i^{FR}.

Another important class of the three-term conjugate gradient method proposed by Zhang et al. [9, 10] by considering a descent modified PRP and also a descent modified HS conjugate gradient method as

$$d_{i+1}^{ZPRP} = \begin{cases} -g_{i+1} & if \quad i = 0 \\ -g_{i+1} + \frac{g_{i+1}^T y_i}{g_i^T g_i} d_i - \frac{g_{i+1}^T d_i}{g_i^T g_i} y_i & if \quad i \ge 1 \end{cases} \tag{14}$$

and

$$d_{i+1}^{ZHS} = \begin{cases} -g_{i+1} & if \quad i = 0 \\ -g_{i+1} + \frac{g_{i+1}^T y_i}{d_i^T y_i} d_i - \frac{g_{i+1}^T d_i}{d_i^T y_i} y_i & if \quad i \ge 1 \end{cases} \tag{15}$$

This paper is organized as follow: in Section 2, we will present a new three-term conjugate gradient method as we prove the descent condition and sufficient descent condition of new method then we will study the global convergence property of our new three term method. In section 3, some numerical experiments to this new three term conjugate gradient method are reported. In section 4, we will present the conclusion.

2 A New Three-Term Conjugate Gradient Method

In this section, we will derive our suggestion based on the search direction that is present by Alaa et al. [2]. We suggested a three-term conjugate gradient method, in which the search direction is:

$$d_{i+1}^{NEW} = -g_{i+1} + \beta_i^{PRP}\, d_i + \beta_i^{NEW}\, y_i \tag{16}$$

Now, from equation (7) and equation (16), we have

$$-g_{i+1} + \beta_i^{PRP}\, d_i + \beta_i^{NEW}\, y_i = g_{i+1} + \frac{g_{i+1}^T y_i}{g_i^T g_i}\, d_i - t\frac{g_{i+1}^T v_i}{g_i^T g_i}\, d_i \tag{17}$$

Multiplying both sides of above equation by y_i , we obtain

$$\beta_i^{NEW} y_i^T y_i = -t\frac{g_{i+1}^T v_i}{g_i^T g_i} d_i^T y_i \tag{18}$$

This implies that,

$$\beta_i^{NEW} = -t\frac{g_{i+1}^T v_i\, d_i^T y_i}{g_i^T g_i\, y_i^T y_i} \tag{19}$$

To achieve balance, we will suppose that the value of parameter t is
$t = \gamma g_{i+1}^T y_i$ where $\gamma > 0$,
so,

$$d_{i+1}^{NEW} = -g_{i+1} + \beta_i^{PRP}\, d_i + \beta_i^{NEW}\, y_i$$

or

$$d_{i+1}^{NEW} = -g_{i+1} + \frac{g_{i+1}^T y_i}{g_i^T g_i}\, d_i - \gamma\frac{g_{i+1}^T v_i\, d_i^T y_i g_{i+1}^T y_i}{g_i^T g_i\, y_i^T y_i}\, y_i \tag{20}$$

Remarks:

1. Note that if $\gamma = 0$, or the line search is exact, then the method reduces to the PR method.
2. If the objective function is quadratic convex $g_{i+1}^T g_i = 0$ and line search is exact $g_{i+1}^T d_i = 0$, then the new method in the search direction (16) will reduce to the Fletcher and Reeves (FR) direction.

2.2 Algorithm of New Method

1. Given an initial point $x_0 \in R^n$
2. set $d_0 = -g_0,\ i = 0$.
3. If $\|g_i\| = 0$ then stop, otherwise go to Step 4.
4. Compute the step size λ_i by using minimize $f(x_i + \lambda_i d_i)$.
5. Set $x_{i+1} = x_i + \lambda_i d_i$.
6. Determine g_{i+1}, if $\|g_{i+1}\| \le 10^{-5}$ stop, else go to Step 7.
7. Compute d_{i+1} by (**20**).
8. If $\|g_{i+1}\|^2 \le \frac{|g_i^T g_{i+1}|}{0.2}$ is satisfied go to step 3,

else $i = i + 1$ and go to step 4.

Theorem 2.1: Assume that the sequence $\{x_i\}$ is generated by (2), then the search direction in (20) satisfy the descent condition, i.e

$$g_{i+1}^T d_{i+1} \le 0 \tag{21}$$

Proof: Multiply both sides of above equation by g_{i+1} , to obtain

$$g_{i+1}^T d_{i+1}^{NEW} = -g_{i+1}^T g_{i+1} + \frac{g_{i+1}^T y_i}{g_i^T g_i} g_{i+1}^T\, d_i - \gamma\frac{g_{i+1}^T v_i\, d_i^T y_i g_{i+1}^T y_i}{g_i^T g_i\, y_i^T y_i} g_{i+1}^T\, y_i \tag{22}$$

If the above search direction is exact, then it satisfies the descent condition i.e.

$$g_{i+1}^T d_{i+1}^{NEW} = -\|g_{i+1}\|^2 \leq 0.$$

However, if the search direction (22) is inexact (i.e.) $g_{i+1}^T d_i \neq 0$.We concludes that the first two terms of equation (22) are satisfies the descent condition i.e.

$$-g_{i+1}^T g_{i+1} + \frac{g_{i+1}^T y_i}{g_i^T g_i} g_{i+1}^T d_i \leq 0, \tag{23}$$

because the PRP method satisfies the (21) condition.

Therefore, we can present the equation (22) as

$$g_{i+1}^T d_{i+1}^{NEW} \leq -\gamma \frac{g_{i+1}^T v_i \, d_i^T y_i \, (g_{i+1}^T y_i)^2}{g_i^T g_i \, y_i^T y_i} \tag{24}$$

Since from wolfe condition we have $-\sigma \, d_i^T g_i \leq g_{i+1}^T d_i$, by multiplying both side by -1 we get $-g_{i+1}^T d_i \leq d_i^T g_i$. So, the equation (24), can be written as

$$g_{i+1}^T d_{i+1}^{NEW} \leq \gamma \frac{\sigma \, \lambda_i \, d_i^T g_i \, d_i^T y_i \, (g_{i+1}^T y_i)^2}{g_i^T g_i \, y_i^T y_i}$$

Clearly, $\gamma, \lambda_i, \sigma$, $d_i^T y_i, (g_{i+1}^T y_i)^2$, $g_i^T g_i$ and $y_i^T y_i$ are positive. But $d_i^T g_i \leq 0$
So, we have

$$g_{i+1}^T d_{i+1}^{NEW} \leq 0 \blacksquare$$

Theorem 2.2: Suppose $\{x_i\}$ is a sequence generated by (2). Then the search directions defined by (20), satisfy the sufficient descent condition,

$$g_{i+1}^T d_{i+1} \leq -C \|g_{i+1}\|^2 \tag{25}$$

Proof: It is obvious from theorem 2.1, after multiplying the new search direction (20) by g_{i+1}^T that the first two terms of equation (20) are less than or equal to zero and from equation (24), we have

$$g_{i+1}^T d_{i+1}^{NEW} \leq -[\gamma \frac{\lambda_i \, (d_i^T y_i)^2 \, (g_{i+1}^T y_i)^2}{g_i^T g_i \, y_i^T y_i \|g_{i+1}\|^2}] \, \|g_{i+1}\|^2 \tag{26}$$

Let $C = \gamma \frac{\lambda_i \, (d_i^T y_i)^2 \, (g_{i+1}^T y_i)^2}{g_i^T g_i \, y_i^T y_i \|g_{i+1}\|^2}$ which is positive, then

$$g_{i+1}^T d_{i+1}^{NEW} \leq -C \, \|g_{i+1}\|^2 \blacksquare$$

Some necessary assumptions for the objective function to prove the global convergence property are given as follows.
Assumption 2.1: The level set $\delta = \{x \mid f(x) \leq f(x_i)\}$ at x_i is bounded, there exists a constant $a > 0$ such that

$$\|x\| \leq a, \forall x \in \delta. \tag{27}$$

In some neighborhood N of δ, f is continuously differentiable, and its gradient is Lipschitz continuous with Lipschitz constant $L > 0$, i.e.

$$\|g(r) - g(t)\| \leq \delta \, \|r - t\| \ \forall r, t \in \delta \tag{28}$$

From the above assumptions, that there exists a positive constant b such that

$$\|g(b)\| \leq b \forall x \in \delta. \tag{29}$$

If function f is strongly convex, then there exists a constant $\mu > 0$ such that

$$\mu \|x-y\|^2 ? (\nabla f(x) - \nabla f(y))^T (x-y), forallx, y \in \delta \tag{30}$$

.

Lemma 2.1: [3, 5] Let assumption (2.1) hold. Consider the methods (1) and (2), where d_i is a descent direction and λ_i satisfies the e standard Wolfe line search. If

$$\sum_{i\geq 1} \frac{1}{\|d_i\|^2} = \infty. \tag{31}$$

Then,

$$\lim_{i\to\infty} inf \ \|g_i\| = 0. \tag{32}$$

Theorem 2.3: Suppose the assumption (2.1) hold and consider the new method in (2) and (20) where λ_i is computed by the strong Wolfe conditions satisfies the global convergence.

Proof: $d_{i+1}^{NEW} = -g_{i+1} + \frac{g_{i+1}^T y_i}{g_i^T g_i} d_i - \gamma \frac{g_{i+1}^T v_i \, d_i^T y_i g_{i+1}^T y_i}{g_i^T g_i \, y_i^T y_i} y_i;$

$$\|d_{i+1}^{NEW}\| \leq \|-g_{i+1}\| + \left\| \frac{g_{i+1}^T y_i}{g_i^T g_i} \right\| \|d_i\| + \left\| \gamma \frac{g_{i+1}^T v_i \, d_i^T y_i g_{i+1}^T y_i}{g_i^T g_i \, y_i^T y_i} \right\| \|y_i\| \tag{33}$$

Since $g_{i+1}^T v_i \leq \lambda_i d_i^T y_i$. so, we have

$$\|d_{i+1}^{NEW}\| \leq b + \frac{\|g_{i+1}\| \|y_i\|}{\|g_i\|^2} \|d_i\| + \gamma\lambda_i \frac{\|d_i^T y_i\|^2 \|g_{i+1}\| \|y_i\|}{\|g_i\|^2 \|y_i\|^2} \|y_i\| \tag{34}$$

Hence, form Lipschitz condition we have $\|y_i\| \leq \delta \|v_i\|$ and from equation (3) we have $d_i = -g_i$, also we have this form $v_i = \lambda_i d_i = -\lambda_i g_i$.

$$\|d_{i+1}^{NEW}\| \leq b + \frac{\|g_{i+1}\| \delta\lambda_i \|g_i\|}{\|g_i\|^2} \|d_i\| + \gamma\lambda_i \frac{\|d_i\|^2 \delta\lambda_i^2 \|g_i\|^2 \|g_{i+1}\|}{\|g_i\|^2} \tag{35}$$

Hence

$$\|d_{i+1}^{NEW}\| \leq b + b\delta\lambda_i + \gamma\delta\lambda_i^3 b \|d_i\|^2 = M \tag{36}$$

$$\therefore \sum_{i\geq 1} \frac{1}{\|d_{i+1}^{NEW}\|^2} \geq \sum_{i\geq 1} \frac{1}{M} = \infty$$

$$\Rightarrow \sum_{i\geq 1} \frac{1}{\|d_{i+1}^{NEW}\|^2} = \infty$$

By using lemma 2.1, we get

$$\lim_{i\to\infty} inf \ \|g_{i+1}\| = 0 . \blacksquare$$

3 Numerical results and discussion

In this section, we will study the implementation of the new three term conjugate gradient method. The tests include familiar nonlinear problems with different dimension. The code of the method is written by FORTRAN 95 language with given initial points. Table (1) illustrate that the numerical results of our new method is more effective than PRP method in the same field with respect to the number of iterations (NOI) and the number of functions evaluation (NOF).

Table 1: The results of the new three term method with PRP method

		Standard form PRP		New Formula TT	
test function	i	NOI	NOF	NOI	NOF
Powell	5	40	120	29	78
	10	40	120	34	92
	100	43	135	32	90
	500	46	150	35	91
	1000	46	150	39	112
	5000	50	180	36	104
Mile	5	37	116	30	94
	10	37	116	30	96
	100	44	148	37	128
	500	44	148	36	119
	1000	50	180	49	185
	5000	50	180	49	178
Central	5	22	159	22	159
	10	22	159	22	159
	100	22	159	22	159
	500	23	171	23	170
	1000	23	171	23	170
	5000	30	270	23	182
Cubic	5	15	45	15	45
	10	16	47	16	47
	100	16	47	16	47
	500	16	47	13	38
	1000	16	47	13	38
	5000	16	47	13	44
Wolfe	5	14	29	14	29
	10	32	65	32	65
	100	49	99	45	91
	5000	58	117	46	93
	1000	64	129	44	90
5000	99	214	55	112	
Sum	5	6	39	6	39
	10	6	34	6	34
	100	14	80	12	61
	500	21	123	19	93
	1000	23	127	20	103
	5000	31	145	26	124
Wood	5	29	67	29	67
	10	29	67	29	67
	100	30	69	30	69
	500	30	69	30	69
	1000	30	69	30	69
	5000	30	69	30	69
Total		1539	5233	1309	4384

Table 2: The percentage of improvement between the new three term with PRP method.

Tools	PRP	New TT
NOI	100%	85.0552%
NOF	100%	83.7760%

Table 2 shows the improve percentage of new three term method. The new method has improvement as compared to PRP method with 14.9448% in NOI and 16.224% in NOF and in general the rate of improvement in the new three term method is 15.5844% in NOI and NOF.

4 Conclusion

In this paper, we present a new method of three-term conjugate gradient method for solving unconstrained optimization. The main idea is improving the search direction of conjugate gradient method to new search direction . The descent and sufficient descent conditions of our new method are proved. Also, we study the global convergent property under standard assumptions. The numerical results on low and high dimensionality problems reported that the new method is more efficient than the PRP method.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors have contributed to all parts of the article. All authors read and approved the final manuscript.

References

[1] Alaa L. I., and Salah G. Sh., (2019) *A new class of three-term conjugate gradient methods for solving unconstrained minimization problems,* , General Letters in Mathematics. pp.187-193. Vol 7 (2) 79-86. https://doi.org/10.31559/glm2019.7.2.4.

[2] Alaa L. I., Muhammad A. S. and Salah G. Sh., (2019), *A New Conjugate Gradient Coefficient for Unconstrained Optimization Based On Dai-Liao,* , JUOZ , Vol. 7, No.1. https://doi.org/10.25271/sjuoz.7.1.525.

[3] 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.

[4] Fletcher, R. and Reeves, C.M., (1964),*Function minimization by conjugate gradients,*, The Computer Journal. 7, 149-154.

[5] Hestenes, M. R. and Stiefel, E., (1952),*Methods of conjugate gradients for solving linear systems,*, Journal of Research of the National Bureau of Standards. 49, 409-436.

[6] Li, Z. and Weijun, Z., (2008),*Tow descent hybrid conjugate gradient methods for optimization,*,Journal of computational and Appl. Math, 216, 251.

[7] Nazareth L., (1977), *A conjugate direction algorithm without line searches,*,Journal of Optimization, Theory and Applications, 23, 373-387.

[8] Polak, E. and Ribiere, G., (1969), *Note surla convergence des mÃ©thodes de directions conjuguÃ©es,*3(16), 35-43.

[9] Zhang, L, Zhou, W, Li, D-H., (2006), *A descent modified Polak-RibiÃ¨re-Polyak conjugate gradient method and its global convergence.* IMA J. Numer. Anal. 26(4), 629-640.

[10] Zhang, L, Zhou, W., Li, D., (2007), *Some descent three-term conjugate gradient methods and their global convergence.* .Optim. Methods Softw. 22(4), 697-711.