Mathematics and Statistics 9(6): 882-888, 2021 http://www.hrpub.org
DOI: 10.13189/ms.2021.090603

A Modified Perry's Conjugate Gradient Method Based on Powell's Equation for Solving Large-Scale Unconstrained Optimization

Mardeen Sh. Taher[1,*], Salah G. Shareef[2]

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

Received August 21, 2021; Revised September 22, 2021; Accepted October 28, 2021

Cite This Paper in the following Citation Styles

(a): *[1] Mardeen Sh. Taher, Salah G. Shareef , "A Modified Perry's Conjugate Gradient Method Based on Powell's Equation for Solving Large-Scale Unconstrained Optimization," Mathematics and Statistics, Vol. 9, No. 6, pp. 882 - 888, 2021. DOI: 10.13189/ms.2021.090603.*

(b): *Mardeen Sh. Taher, Salah G. Shareef (2021). A Modified Perry's Conjugate Gradient Method Based on Powell's Equation for Solving Large-Scale Unconstrained Optimization. Mathematics and Statistics, 9(6), 882 - 888. DOI: 10.13189/ms.2021.090603.*

Abstract It is known that the conjugate gradient method is still a popular method for many researchers who are focused in solving the large-scale unconstrained optimization problems and nonlinear equations because the method avoids the computation and storage of some matrices so the memory's requirements of the method are very small. In this work, a modified Perry conjugate gradient method which fulfills a global convergence with standard assumptions is shown and analyzed. The idea of new method is based on Perry method by using the equation which is founded via Powell in 1978. The weak Wolfe–Powell search conditions are used to choose the optimal line search, under the line search and suitable conditions, we prove both descent and sufficient descent conditions. In particular, numerical results show that the new conjugate gradient method is more effective and competitive when compared to other of standard conjugate gradient methods including: - CG- Hestenes and Stiefel (H/S) method, CG-Perry method, CG- Dai and Yuan (D/Y) method. The comparison is completed under a group of standard test problems with various dimensions from the CUTEst test library and the comparative performances of the methods are evaluated by total the number of iterations and the total number of function evaluations.

Keywords Conjugate Gradient, Wolfe Line Search, Descent and Sufficient Descent Conditions, Analysis of Convergence

1. Introduction

Conjugate gradient (CG) methods have been used to solve large-scale unconstrained optimization problem which is presented as:

$$min\{h(x) \ : \ x \in R^n\} \tag{1}$$

Where $h(x)$ is smooth and twice continuously differentiable function over R^n , n is a number of variables for our problem. In general Conjugate gradient (CG) methods are generate a sequence x_i of estimates to the minimum of $h(x)$, by iterative formula: -

$$x_{i+1} = x_i + \lambda_i d_i \tag{2}$$

However, $\lambda_i > 0$ is a steplength and must be achieved the following Wolfe–Powell search conditions: -

$$h(x_i + \lambda_i d_i) - h(x_i) \le \xi_1 \lambda_i \nabla h_i^T d_i, \xi_1 \in \left(0, \frac{1}{2}\right) \tag{3}$$

$$\nabla h(x_i + \lambda_i d_i)^T d_i \ge \xi_2 \nabla h_i^T d_i \quad \xi_2 \in (\xi_1, 1) \tag{4}$$

And using a search direction:

$$\begin{cases} d_i = -g_i & i = 0 \\ d_{i+1} = -g_{i+1} + \beta_i d_i & i \ge 1 \end{cases} \tag{5}$$

where β_i is a scalar determines the conjugate gradient methods (CG-methods), and we denote$\nabla h(x_i)$ which is gradient of $h(x_i)$ by g_i.

At current, the most effective formulas of β_i are the following:

$$\beta_i^{H/S} = \frac{g_{i+1}^T(g_{i+1} - g_i)}{d_i^T(g_{i+1} - g_i)}$$

$$\beta_i^{F/R} = \frac{g_{i+1}^T g_{i+1}}{g_i^T g_i}$$

$$\beta_i^{\mathrm{PRP}} = \frac{g_{i+1}^T(g_{i+1} - g_i)}{g_i^T g_i}$$

$$\beta_i^{perry} = \frac{g_{i+1}^T(y_i - v_i)}{d_i^T(y_i)}$$

$$\beta_i^{D/Y} = \frac{g_{i+1}^T g_{i+1}}{d_i^T y_i}$$

In which, the definition of $\beta_i^{H/S}$ is due to Hestenes and Stiefe (H/S, 1952) [6], $\beta_i^{F/R}$ is due to Fletcher and Reeves (F/R, 1964) [4,5], β_i^{PRP} is due to Polak-Ribiere-Polyak (PRP, 1969) [7], β_i^{perry} is suggested by Perry(1978) [8] and in (1999) $\beta_i^{D/Y}$ is established by Dai-Yuan [2]. For iterative mathematics method the global convergence is one of requirement for it , and it mean by global convergence is any sequence generated via the iterative methods will either achieved after finite steps or consist of a subsequence which is converged to stationary point of the problem ($h(x)$) from a given initial point (x_0) , $\lim_{i\to\infty}(\|\nabla h(x_i + \lambda_i d_i)\|) = 0$) [11]. For Conjugate gradient the global convergence is investigated by researchers , Zoutendijk G. in 1970 proved that the Fletcher and Reeves method is global convergence in case that the line search is exact [13]. In Polak-Ribire method global convergence is described by Powell [9] when $h(x)$ is strongly convex and the line search is also exact, but after that Powell demonstrated that the Polak-Ribire strategy with exact line search could circle infinitely without convergent to a required point. The same result applies to the Hestenes-Stiefel method.

Many efforts have been made in few recent years to design new formulas for conjugate gradient method which are not only satisfied global convergence but also improve numerical performance for method. Remainder the conjugate gradient methods have many application in real life In our work, we found a new formula for CG-method which is satisfied the global convergence in section 3.3 and the new formula with weak Wolfe–Powell generate a descent direction at each iteration in section 3.2.

2. Derivation of the New Conjugacy Coefficient

In this section, we present a new formula for conjugate gradient method as a result improve the parameter of β_i^{perry} in (6) by using the equation which is proposed by Powell in (1978)[10].

Since

$$\beta_i^{perry} = \frac{g_{i+1}^T(y_i - v_i)}{d_i^T(y_i)} \tag{6}$$

The equation which is developed by Powell is:

$$y_i^{powell} = y_i + (1-\theta)(Gv_i - y_i) \tag{7}$$

Where, $y_i = g_{i+1} - g_i$ and $v_i = x_{i+1} - x_i$

Thus, we have the quasi-Newton equation:

$$Gv_i = \delta_i \; y_i \tag{8}$$

Now, suppose that

$$\delta_i = \frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)} \tag{9}$$

Where $\|.\|: R^n \to R$, is Euclidean norm, and ϖ is a machine accuracy, hence, from (8) and (9) we get:

$$Gv_i = \|v_i\| \frac{y_i}{2\sqrt{\varpi}(1+\|x_{i+1}\|)} \tag{10}$$

So, after putting (10) in (9) we obtain new equation y_i^{**}.

$$y_i^{**} = y_i + (1-\theta)\left(\|v_i\| \frac{y_i}{2\sqrt{\varpi}(1+\|x_{i+1}\|)} - y_i\right) \tag{11}$$

$$y_i^{**} = (1-\theta)\left(\|v_i\| \frac{y_i}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right) + \theta y_i \tag{12}$$

The idea of our new formula is based on replacing the equation y_i in Perry formula (3) with the equation y_k^{**}, and getting

$$\beta_i^{new} = \frac{g_{i+1}^T\left((1-\theta)\left(\|v_i\| \frac{y_i}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right) + \theta y_i \;] - v_i\right)}{d_i^T\left[(1-\theta)\left(\|v_i\| \frac{y_i}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right) + \theta y_i\right]}$$

$$= \frac{\left[(1-\theta)\left(\frac{\|v_i\|}{2\sqrt{k}(1+\|x_{i+1}\|)}\right) + \theta\right] g_{i+1}^T y_i - g_{i+1}^T v_i}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right) + \theta) d_i^T y_i}$$

$$= \frac{\left((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right) + \theta\right) g_{i+1}^T y_i}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right) + \theta) d_i^T y_i} -$$

$$- \frac{g_{i+1}^T v_i}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right) + \theta) d_i^T y_i} \tag{13}$$

Hence, we get the new parameter conjugacy: -

$$\beta_i^{new} = \frac{g_{i+1}^T y_i}{d_i^T y_i} - \frac{g_{i+1}^T v_i}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right) + \theta) d_i^T y_i} \tag{14}$$

we observe that , if the orthogonal condition is satisfied i.e. $g_{i+1}^T g_i = 0$ then the parameter β_k^{new} in (14) become as a following:

$$\beta_i^{new} = \frac{g_{i+1}^T(g_{i+1} - g_i)}{d_i^T y_i} - \frac{g_{i+1}^T v_i}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right) + \theta) d_i^T y_i} \tag{15}$$

$$\beta_i^{new} = \frac{\|g_{i+1}\|^2}{d_i^T y_i} - \frac{g_{i+1}^T v_i}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta)d_i^T y_i} \quad (16)$$

So, we obtain the modified Perry Conjugacy Coefficient, and get new direction:-

$$d_{i+1} = -g_{i+1} + \beta_i^{new} d_i \quad (17)$$

3. Algorithm, Descent Property and Convergence Analysis of New Method

3.1. The Algorithm of New Conjugate Gradient Method

The New CG-method is summarized below by following steps:

Step 0: Let $i = 0$, select $x_0 \in R^n$ is an initial point, $n \in \mathbb{Z}$ and $\varepsilon > 0$.

Step 1: Test a criterion for stopping, if $\|g_0\| < \varepsilon$ then stop.

else $d_0 = -g_0 = -\nabla f(x_0)$ go to step (2).

Step2: Do cubic line search to calculate steplength (λ_i) with Wolfe line searches (3) and (4),

Step3 : Determine $x_{i+1} = x_i + \lambda_i d_i$, and go to step(4).

Step4: Stop if $\|g_{i+1}\| < \varepsilon$ and $x_{i+1}$is a minimizer, else go to step (5)

Step5 : Calculate $d_{i+1} = -g_{i+1} + \beta_i d_i$, β_i is defined in (15) or (16).

Step6: If $|g_{i+1}^T g_i| > 0.2\, g_{i+1}^T g_{i+1}$, then go to step1 else $i := i + 1$, and go to Step 2.

3.2. Descent and Sufficient Descent Conditions

To ensure that the new algorithm is convergence we needed to prove the following theorems: -

Theorem3.2.1:- Consider the sequences of $\{x_i, d_i, g_i\}$ are generated by New CG- method then the search direction $\{d_i\}$satisfies the descent property

$$d_{i+1}^T g_{i+1} \le 0 \quad (18)$$

proof: - (prove by induction,)

when $= 0$, $d_0 = -g_0$, so

$$d_0^T g_0 \le -\|g_0\|^2. \quad (19)$$

Assume the induction hypothesis that for a particular (i)meaning $d_i^T g_i < 0$ is true for any > 0 ,

now we prove (18) is hold at $i + 1$, from the definition of new direction we have

$$d_{i+1} = -g_{i+1} + \left(\frac{g_{i+1}^T y_i}{d_i^T y_i} - \frac{g_{i+1}^T v_i}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta)d_i^T y_i}\right) d_i \quad (20)$$

Multiply both sides of (20) by g_{i+1}^T from right to get,

$$d_{i+1}^T g_{i+1} = -g_{i+1}^T g_{i+1} + \left(\frac{g_{i+1}^T y_i}{d_i^T y_i} - \frac{g_{i+1}^T v_i}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta)d_i^T y_i}\right) d_i^T g_{i+1} \quad (21)$$

$$= -\|g_{i+1}\|^2 + \frac{g_{i+1}^T y_i}{d_i^T y_i} d_i^T g_{i+1} - \frac{g_{i+1}^T v_i}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta)d_i^T y_i} d_i^T g_{i+1} \quad (22)$$

$$= -\|g_{i+1}\|^2 + \frac{g_{i+1}^T y_i}{d_i^T y_i} d_i^T g_{i+1} - \frac{g_{i+1}^T \lambda_i d_i}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta)d_i^T y_i} d_i^T g_{i+1} \quad (23)$$

Since, λ_i is parameter greater than zero so:

$$d_{i+1}^T g_{i+1} = -\|g_{i+1}\|^2 + \frac{g_{i+1}^T y_i}{d_i^T y_i} d_i^T g_{i+1} - \frac{\lambda_i (d_i^T g_{i+1})^2}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta)d_i^T y_i} \quad (24)$$

it is clear that the first two terms of (24) refer to Hestenes and Stiefel method generates the descent directions, now we need to prove that the third term of (24) is less than or equal to zero.

Noticeably, all of λ_i , $(d_i^T g_{i+1})^2$, $\|v_i\|$, and $\|x_{i+1}\|$are positive and $\theta \in (0,1)$. Noteworthy that

$$d_i^T y_i = d_i^T (g_{i+1} - g_i) > (\xi_2 - 1) d_i^T g_i \ ,.$$

$(\xi_2 - 1)\, d_i^T g_i > 0$ therefore $d_i^T y_i > 0$. So, the third term of (24) is also less than of zero.

In case, when the orthogonal property is satisfy, we see

$$d_{i+1} = -g_{i+1} + \left(\frac{\|g_{i+1}\|^2}{d_i^T y_i} - \frac{g_{i+1}^T v_i}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta)d_i^T y_i}\right) d_i \quad (25)$$

Multiply both sides (25) by g_{i+1}^T, we get

$$d_{i+1}^T g_{i+1} = -\|g_{i+1}\|^2 + \frac{\|g_{i+1}\|^2}{d_i^T y_i} d_i^T g_{i+1} - \frac{\lambda_i (d_i^T g_{i+1})^2}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta)d_i^T y_i} \quad (26)$$

The terms $-\|g_{i+1}\|^2 + \frac{\|g_{i+1}\|^2}{d_i^T y_i} d_i^T g_{i+1}$ proved is the descent property by Dai and Yuan [2] and we proved term $\left(-\frac{\lambda_i (d_i^T g_{i+1})^2}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta)d_i^T y_i} < 0\right)$ in above . So the descent property is satisfied when $g_{i+1}^T g_i = 0$ Hence, we obtain

$$g_{i+1}^T d_{i+1} \le 0 \quad (27)$$

the proof is completed by induction.

Theorem 3.2.2: - let the sequences of $\{x_i\}$ and $\{d_i\}$ are generated by our proposed method of modified CG satisfies the descent property for all new CG- method is also hold the sufficient condition:

$$d_{i+1}^T g_{i+1} < -c\|g_{i+1}\|^2, \forall i \geq 0 \text{and} c > 0 \quad (28)$$

Proof

Observably, the Hestenes and Stiefel method generates descent directions so, the term $(-\|g_{i+1}\|^2 + \frac{g_{i+1}^T y_i}{d_i^T y_i} d_i^T g_{i+1})$ in (24) is less than or equal to zero, consequently we can rewrite the (24) as following:-

$$d_{i+1}^T g_{i+1} \leq -\frac{\lambda_i (d_i^T g_{i+1})^2}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta) d_i^T y_i} \quad (29)$$

Multiply and divide the right side of (29) by $\|g_{i+1}\|^2$, we obtain

$$d_{i+1}^T g_{i+1} \leq -\frac{\lambda_i (d_i^T g_{i+1})^2 \|g_{i+1}\|^2}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta) d_i^T y_i \|g_{i+1}\|^2} \quad (30)$$

$$= -c\|g_{i+1}\|^2$$

$c = \frac{\lambda_i (d_i^T g_{i+1})^2}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta) d_i^T y_i \|g_{i+1}\|^2}$, and according to theorem (3.2.1), we see > 0 .

Therefore

$$d_{i+1}^T g_{i+1} \leq -c\|g_{i+1}\|^2 \quad (31)$$

Therefore, new method satisfied the sufficient descent condition.

3.3. Global Convergence Analysis for New Method

In order to prove the global convergence of new nonlinear conjugate gradient methods, under the Wolfe line search. We will impose the following assumptions and lemma about the objective function $h(x)$.

Assumptions (3.3.1)

(i). Let $h(x): R^n \to R$ be twice continuously differentiable, and the level set $\mathcal{L}_{\breve{x}} = \{x: x \in R^n, and h(x) \leq h(\breve{x})\}$ is closed and bounded.

(ii). In some neighborhood $\mathcal{B}$ of $\mathcal{L}_{\breve{x}}$, the gradient of $h(x)$ *is* ,$\nabla h(x)$ and satisfying Lipschitz condition, namely there exists a constant $\ell > 0$ such that:

$$\|\nabla h(\text{a}) - \nabla h(\text{b})\| \leq \ell\|\text{a} - \text{b}\| \ \forall\, a \text{ and } \text{b} \in \mathcal{L}_{\breve{x}} \quad (32)$$

Noticeably that the $\{ x_i \}$ is generating by new CG-method (14) which is in a bounded and under assumptions (3.3.1) on $h(x)$, there exists a constant $\gamma > 0$ such that $\|\nabla h(\breve{x})\| \leq \gamma , \forall \breve{x} \in \mathcal{B}$.

lemma (3.3.1): let hypotheses (3.3.1) is satisfied and the λ_i is taken by the Wolfe conditions(3)and (4), if

$$\sum_{i\geq 1} \frac{1}{\|d_{i+1}\|^2} = \infty \quad (33)$$

Then,

$$\lim_{i\to\infty}(\inf(\|g_{i+1}\|))=0 \quad (34)$$

Theorem (3.3.1)

Let x_i be generated by new CG-method (14) and function $h(x)$ is uniformly convex, if assumptions (3.3.1) holds then we have

$$\lim_{i\to\infty}(\inf(\|g_i\|)) = 0 \quad (35)$$

Proof

From (14), we have

$$|\beta_i^{new}| = \left| \frac{g_{i+1}^T y_i}{d_i^T y_i} - \frac{g_{i+1}^T v_i}{((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right)+\theta) d_i^T y_i} \right| \quad (36)$$

Let $\Gamma = ((1-\theta)\left(\frac{\|v_i\|}{2\sqrt{\varpi}(1+\|x_{i+1}\|)}\right) + \theta$, thus (36) becomes

$$|\beta_i^{new}| = \left| \frac{g_{i+1}^T y_i}{d_i^T y_i} + (-\lambda_i) \frac{g_{i+1}^T d_i}{\Gamma d_i^T y_i} \right| \quad (37)$$

we know that the $g_{i+1}^T d_i < d_i^T y_i$ and using triangle and Cauchy–Schwarz inequalities in (37) and we obtain

$$|\beta_i^{new}| \leq \frac{\|g_{i+1}\|\|y_i\|}{\|d_i\|\|y_i\|} + \frac{\lambda_i}{\Gamma} \quad (38)$$

and return to hypothesis (ii), and using $\|g_{i+1}\| \leq \gamma$, we get

$$|\beta_i^{new}| \leq \frac{\gamma}{\|d_i\|} + \frac{\lambda_i}{\Gamma} \quad (39)$$

Taking norm for both sides of new direction in (17):

$$\|d_{i+1}\| = \|g_{i+1} + \beta_i^{new} d_i\| \quad (40)$$

Apply preliminary of triangle inequality in (40), and we get

$$\|d_{i+1}\| \leq \|g_{i+1}\| + |\beta_i^{new}|\|d_i\| \quad (41)$$

From (38) and (41), we obtain

$$\|d_{i+1}\| \leq \gamma + \left(\frac{\gamma}{\|d_i\|} + \frac{\lambda_i}{\Gamma}\right)\|d_i\| \quad (42)$$

implies that

$$\|d_{i+1}\| \leq \gamma + \left(\gamma + \frac{\|v_i\|}{\Gamma}\right) \quad (43)$$

and $\|v_i\| = \|x_{i+1} - x_i\|$, suppose $\mathcal{M} = \max\{\|x_{i+1} - x_i\| : x_{i+1} , x_i \in R^n\}$

$$\|d_{i+1}\|^2 < \mathcal{H}^2 \ , \ \mathcal{H} = \gamma + \left(\gamma + \frac{\mathcal{M}}{\Gamma}\right) \quad (44)$$

Thus ,

$$\sum_{i\geq 1} \frac{1}{\|d_{i+1}\|^2} > \sum_{i\geq 1} \frac{1}{\mathcal{H}^2} \quad (45)$$

which indicates

$$\sum_{i\geq 1} \frac{1}{\|d_{i+1}\|^2} = \infty \quad (46)$$

By Lemma (3.1.1), we obtain

$$\lim_{i\to\infty}(\inf(\|g_{i+1}\|)) = 0 \quad (47)$$

Table 1. Comparing performance of the three method s (New formula $\boldsymbol{\beta_i^{new}}$, $\boldsymbol{\beta_i^{perry}}$ and classical Hestenes and Stiefel $\boldsymbol{\beta_i^{H/S}}$)

Test	**N**	$\boldsymbol{\beta_i^{new}}$	$\boldsymbol{\beta_i^{H/S}}$	$\boldsymbol{\beta_i^{perry}}$
		NOI- NOF	**NOI- NOF**	**NOI- NOF**
Powell (3,-1,0,1)	4	32-87	38-108	35-89
	100	39-114	40-122	43-105
	500	33-92	41-124	43-105
	1000	38-105	41-124	45-120
	3000	38-104	41-124	46-122
	5000	41-122	41-124	46-122
Wood (-3,-1,-3,-1)	4	27-62	30-68	30-68
	100	28-64	30-68	30-68
	500	28-64	30-68	30-68
	1000	28-64	30-68	30-68
	3000	28-64	30-68	30-68
	5000	28-64	30-68	30-68
Rosen (-1.2,1;…)	4	30-83	30-83	30-83
	100	30-83	30-83	30-83
	500	30-83	30-83	30-83
	1000	30-83	30-83	30-83
	3000	30-83	30-83	30-83
	5000	30-83	30-83	30-83
Powell (0,1,2;..)	4	15-33	16-36	16-36
	100	16-36	16-36	16-37
	500	16-36	16-36	16-37
	1000	16-36	16-36	16-37
	3000	16-36	16-36	16-37
	5000	16-36	16-36	16-37
Gcantrel (1,2,2,2)	4	19-128	22-159	19-128
	100	21-154	22-159	21-154
	500	21-154	23-171	21-154
	1000	22-167	23-171	22-167
	3000	22-167	27-234	22-167
	5000	22-167	28-248	22-167
Cubic (-1.2,1;...)	4	12-35	12-35	12-35
	100	13-37	13-37	13-37
	500	13-37	13-37	13-37
	1000	13-37	13-37	13-37
	3000	13-37	13-37	13-37
	5000	13-37	13-37	13-37
Sum (2;...)	4	3-11	3-11	3-11
	100	14-81	14-81	15-84
	500	21-124	21-124	22-132
	1000	23-128	23-128	23-125
	3000	29-150	27-128	32-168
	5000	30-147	31-159	34-178
Doxin (-1;...)	4	17-36	17-36	17-36
	100	483-1091	7678-15360	7735-15474
	500	565-1255	490-1093	6284-12573
	1000	526-1182	4005-8015	3391-6787
	3000	474-1072	6202-12409	515-1129
	5000	479-1068	481-1080	6357-12717
Miele (1,2,2,2)	4	23-73	28-85	34-133
	100	35-121	33-114	46-169
	500	39-141	40-146	52-198
	1000	46-171	46-176	58-229
	3000	50-202	54-211	58-229
	5000	48-185	54-211	64-261
Wolfe (-1;...)	4	11-24	11-24	11-24
	100	49-99	49-99	49-99
	500	52-105	52-105	52-105
	1000	61-123	70-141	70-141
	3000	134-281	170-351	170-351
	5000	153-317	165-348	166-350
OSP (1;..)	4	6-26	8-44	8-46
	100	47-171	49-185	51-193
	500	105-314	112-353	107-327
	1000	154-460	156-473	156-473
	3000	196-597	197-602	197-602
	5000	256-774	256-774	256-774

Table 2. Comparing performance profiles of New method $\boldsymbol{\beta_i^{new}}$ and Dai and Yuan method $\boldsymbol{\beta_i^{D/Y}}$

Test	**N**	$\boldsymbol{\beta_i^{new}}$	$\boldsymbol{\beta_i^{D/Y}}$
		NOI- NOF	**NOI- NOF**
Powell (-3,-1,0,1)	4	36-96	50-128
	100	37-97	51-130
	500	37-97	51-130
	1000	37-97	51-130
	3000	37-98	52-132
	5000	37-98	52-132
Wood (-3,-1,-3,-1)	4	27-63	28-65
	100	27-63	28-65
	500	27-63	29-68
	1000	27-63	29-68
	3000	29-68	29-68
	5000	29-68	29-68
Rosen	4	27-63	27-63
	100	28-65	28-65
	500	28-65	28-65
	1000	28-65	28-65
	3000	30-69	30-69
	5000	30-69	30-69
Gcantrel (1,2,2,2)	4	14-85	18- 127
	100	28-240	20 -153
	500	16-109	23 -192
	1000	16-109	23 -192
	3000	17-125	24-205
	5000	17-125	24-205
Dixon (-1;...)	4	13-28	13-28
	100	460-994	466-1021
	500	430-927	503-1085
	1000	451-977	484-1048
	3000	432-937	462-1005
	5000	456-989	510-1115
Miele (1,2,2,2)	4	34-109	36-115
	100	38-127	46-156
	500	33-111	53-188
	1000	57-207	60-222
	3000	46-160	66-257
	5000	63-249	66-257
Wolfe (-1;...)	4	11-23	11-23
	100	44-89	45-91
	500	49-99	48-97
	1000	49-99	52-105
	3000	112-237	125-263
	5000	132-284	159-327
OSP (1;...)	4	8- 44	8- 44
	100	51-173	52 -180
	500	136-427	138- 439
	1000	187-576	196- 607
	3000	387-1308	421-1388
	5000	529-1806	555 1857

4. Numerical Results

It is important to take note that the theoretical prove it is not adequate to show the effectiveness or robust of any conjugate gradient method. Thus, there would be need to study the method numerical by evaluate the performance method on group of test problems and evaluation the number of iteration or computation time (CPU time).

In this section, is recorded some numerical results by comparing the suggestion method with classical conjugate gradient methods. All of the methods are coded in Fortran95 and stopping the iteration when this statement $\|g_{i+1}\| < 10^{-5}$ is true. For testing we used the well-known nonlinear problems with dimension ranging between 4 to 5000 [1], all methods are using cubic fit mothed to find the steplength $\lambda_i > 0$ under conditions (3) and (4) with $\xi_1 = 0.001\ and\ \xi_2 = 0.1$.

The following abbreviates in tables meaning:

N:- the number of variable in test problem.

NOI:- the number of iterations.

NOF:- the number of function evaluations.

In table 1, the results show that the performance of the new method is better than the HS and Perry methods based on the total numbers of function evaluations and its gradient evaluations.

As a final note, table 2 demonstrations the performance of new method when the orthogonal condition is available $\langle g_{k+1}, g_k \rangle = 0$. During the numerical experiments outcome from comparing between New method $\boldsymbol{\beta}_i^{new}$ and Dai and Yuan method $\boldsymbol{\beta}_i^{\mathbf{D/Y}}$ and see that the new method is top perform to Dai and Yuan method.

5. Conclusions

In this work, we present a modified Perry's conjugacy coefficient for conjugate gradient method. The proposed new method is apposite for solving large-scale unconstrained optimization problems because of its lower storage requirement. Furthermore, and under suitable conditions we show that the new method is satisfy the descent, sufficient descent conditions and it is global convergent also the new method gives auspicious numerical results.

In future it can be used the new method to train the neural network in order to improve performance and in the image processing.

REFERENCES

[1] Andrei, Neculai. "An Unconstrained Optimization Test Functions Collection." Adv. Model. Optim 10.1: 147-161, 2008.

[2] Dai Y. H. and Yuan Y., "A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Propert". SIAM Journal on Optimization, vol. 10, no. 1, pp. 177–182, 1999.

[3] Dai, Y. and Yuan, Y. A "nonlinear conjugate gradient with a strong global convergence properties". SIAM Journal on Optimization, 10, 177-182. doi: 10.1137/S1052623497318992

[4] Fletcher, R. "Practical Method of Optimization", Vol I: Unconstrained Optimization. 2nd Edition, Wiley, New York. 1997

[5] Fletcher, R.M., Reeves, C.M. "Function Minimization by Conjugate Gradients". Computer Journal, 7, 149-154. doi:10.1093/comjnl/7.2.149

[6] Hestenes, M.R., Stiefel, E.: "Methods of Conjugate Gradients for Solving Linear Systems". Journal of Research National Bureau of Standards. 49(6), 409–439, 1952.

[7] Liu, Dongyi, and Genqi Xu. "A Perry Descent Conjugate Gradient Method with Restricted Spectrum". Optimization Online, Nonlinear Optimization (unconstrained optimization):1-19.

[8] Perry, A.: "A Modified Conjugate Gradient Algorithm". Oper. Res., Tech. Notes 26(6), 1073–1078, 1978.

[9] Polak, E., Ribière, G.: "Note Sur La Convergence De Méthodes De Directions Conjuguées". Rev. Fr. Inform. Rech. Oper. 3(16), 35–43, 1969.

[10] Powell M. J. D.: "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations", Lecture Notes in Mathematics, Vol. 630, 1, pp. 144-157, 1978.

[11] Sulaiman I. M., Mamat M., Waziri M. Y., Yakubu U. A., Malik M., "The Performance Analysis of a New Modification of Conjugate Gradient Parameter for Unconstrained Optimization Models," Mathematics and Statistics, Vol. 9, No. 1, pp. 16-23, 2021. DOI: 10.13189/ms.2021.090103.

[12] Wolfe P. : "Convergence conditions for ascent methods", SIAM Rev., 11, pp. 226–235,1969.

[13] Zoutendijk G., "Nonlinear Programming, Computational Methods in Integer and Nonlinear Programming", J. Abadie, ed., North-Holland, Amsterdam, pp. 37–86, 1970.