Pattern Recognition Using Particle Swarm Optimization with Proposed a New Conjugate Gradient Parameter in Unconstrained Optimization

Ban Ahmed Mitras* and Suhaib Abdul-Jabbar**
**Department of Intelligent Techniques & Operation Research, College of Computer Sciences and Mathematics, University of Mosul.
**Department of Computer, University of Mosul Teacher/College of Education for pure science.
E-mail: dr.ban_mitras@yahoo.com.
E-mail: suhayb1975@yahoo.com.

Abstract

In this paper, we present modified conjugancy coefficient for the conjugate gradient method. This modification using the extention Dai and Yuan Method to solve non-linear programming problems. The algorithm of particle swarm optimization (PSO) is applied in this work, to coefficients extracted by features extraction techniques. The sufficient descent and the global convergence properties for the proposed algorithm are proved. The numerical results of our finding for the large scale optimization problem are very encouraging comparison with standard methods
The experimental results showed that PSO can generate excellent recognition results with the minimal set of selected features. Finally, the algorithm PSO based approaches are proposed and the influence of PSO parameters on the performance is evaluated.

Keywords: Particle Swarm Optimization, Pattern recognition, conjugate gradient, conjugancy coefficient, nonlinear programming, and unconstrained optimization.

1. Introduction

In unconstrained optimization, Minimizing on an objective function will done depends on real variables with no restrictions on the values of these variables. The unconstrained optimization problem is:

$$Min \quad f(x) : x \in R^n, \quad \text{.......} (1)$$

where $f : R^n \to R$ is a continuously differentiable function, bounded from below. A nonlinear conjugate gradient method generates a sequence $\{x_k\}, k$ is integer number, $k \geq 0$. Starting from an initial point x_0, the value of x_k calculate by the following equation:

$$x_{k+1} = x_k + \lambda_k d_k, \quad \text{.......} (2)$$

where the positive step size $\lambda_k > 0$ is obtained by a line search, and the directions d_k are generated as:

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \quad \text{.......} (3)$$

where d_k is a descent search direction and $0 < \rho \leq \sigma < 1$, where β_k is defined by one of the following formulas:

$$\beta_k^{(HS)} = \frac{y_k^T g_{k+1}}{y_k^T d_k} \text{(Hestenese and Stiefel[1])} \quad \text{.......} (4)$$

$$\beta_k^{(FR)} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \text{(Fletcher and Reeves[2])} \quad \text{.......} (5)$$

$$\beta_k^{(PRP)} = \frac{y_k^T g_{k+1}}{g_k^T g_k} \text{(Polak – Ribiere [3] and Polyak [4])} \quad \text{.......} (6)$$

$$\beta_k^{(CD)} = -\frac{g_{k+1}^T g_{k+1}}{g_k^T d_k} \text{(Conjugate descent [5])} \quad \text{.......} (7)$$

$$\beta_k^{(LS)} = -\frac{y_k^T g_{k+1}}{g_k^T d_k} \text{(Liu and Stoery [6])} \quad \text{.......} (8)$$

$$\beta_k^{(DY)} = \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} \text{(Dai and Yuan [7])} \quad \text{.......} (9)$$

Where $d_0 = -g_0$, the value of β_k is determine according to the algorithm of Conjugate Gradient (CG), and its known as a conjugate gradient parameter, $s_k = x_{k+1} - x_k$ and $g_k = \nabla f(x_k) = f'(x_k)$, consider $\|.\|$ is the Euclidean norm and $y_k = g_{k+1} - g_k$. The termination conditions for the conjugate gradient line search are often based on some version of the Wolfe conditions. The standard Wolfe conditions: [8][9]

$$f(x_k + \lambda_k d_k) - f(x_k) \leq \rho \lambda_k g_k^T d_k, \quad \text{.......} (10)$$

$$g(x_k + \lambda_k d_k)^T d_k \geq \sigma g_k^T d_k, \quad \text{.......} (11)$$

2. Particle Swarm Optimization Algorithm

The Particle Swarm Optimization (PSO) algorithm was originally designed by Kennedy and Eberhart in 1995. PSO is a population-based searching method which imitates the social behavior of bird flocks or fish schools. The population and the individuals are called a “swarm” and “particles”, respectively. Each particle moves in the swarm with a velocity that is adjusted according to its own flying experience and retains the best position it has ever encountered in memory. The best local and global positions ever encountered by all particles of the swarm are also communicated to all other particles. The advantages of PSO are that there is neither mutation calculation nor overlapping. The popular form of particle swarm optimizer is defined in the following equations and in the flow chart in Fig.(1) show New PSO Flow Chart with new conjugate gradient parameter[10][11][12].

$$V_{id}(K+1) = wv_{id}(k) + c_1 r_1 (pb_{id}(k) - x_{id}(k)) + c_2 r_2 (gb_{id}(k) - x_{id}(k)) \quad \text{.......... (12)}$$

$$x_{id}(k+1) = x_{id}(k) + V_{id}(k+1) \quad \text{............................ (13)}$$

Where:

V_{id}: is the velocity of particle i along dimension d.

x_{id}: is the position of particle i in dimension d.

c1: is a weight applied to the cognitive learning portion.

c2: is a similar weight applied to the influence of the social learning portion.

r1, r2: are separately generated random numbers in the range of zero and one.

w: is the inertia weight.

cost function is $\dfrac{(\sum_{j=1}^{n}(y_j - x_{ij})^2)^{1/3}}{n}$

In 2005 Mahamed used (PSO) in Pattern Recognition and Image Processing.

Konstantinos and Michael used Particle Swarm Optimization Method for Constrained Optimization Problems [11]

In 2013 MAJIDA used (PSO) in Handwritten Characters Recognition [10].

In 2011 Parvinder S. Sandhu, Shalini Chhabra used (PSO) with Conjugate Gradient Algorithms.

3. Modified PSO Algorithm

The modified PSO Algorithm is the same PSO algorithm but the change is normalize the Initialization the best by conjugate gradient Algorithm. The new PSO Flow Chart with new conjugate gradient parameter.

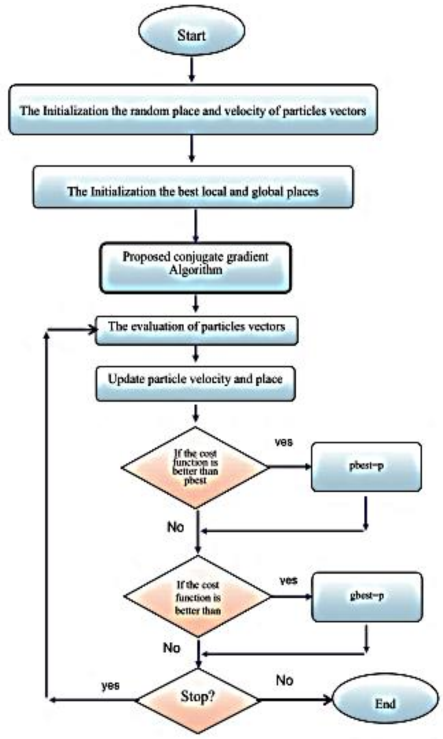

Fig.(1): Modified PSO Flow Chart with new conjugate gradient parameter.

4. Extension Dai and Yuan Method

By using extended of Dai and Yuan (DY) method they need to find new beta that produces a descent search direction .this requires that [13][14]:

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k < 0 \quad \text{........ (14)}$$

Letting τ_{k+1} be a positive parameter, then define

$$\beta_k = \frac{\|g_{k+1}\|^2}{\tau_{k+1}} \quad \text{.. (15)}$$

Equation (14) is equivalent to

$$\tau_{k+1} > g_{k+1}^T d_k \quad \text{.. (16)}$$

Taking the positivity of τ_{k+1} into consideration, they have

$$\tau_{k+1} > \max\{g_{k+1}^T d_k, 0\} \quad \text{.................................} (17)$$

Therefore, if condition (14) is satisfied for all k, the conjugate gradient method with equation (15), they can get various kinds of conjugate gradient methods by choosing various τ_{k+1}, where τ_{k+1} satisfying equation. (17) and prove global convergence of the proposed method. We note that the Wolfe condition in equation (11) guarantees $d_k^T y_k > 0$ and that

$$d_k^T y_k = d_k^T g_{k+1} - d_k^T g_k > d_k^T g_{k+1}$$

This implies that

$$d_k^T y_k > \max\{g_{k+1}^T d_k, 0\} \quad \text{..............................} (18)$$

By setting $\tau_k = d_k^T y_k$ formula (15) reduce to this DY method as:

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

It follows from (3) and (15) that

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k \\ &= -\tau_{k+1}\beta_{k+1} + \beta_{k+1} g_{k+1}^T d_k \\ &= (-\tau_{k+1} + g_{k+1}^T d_k)\beta_{k+1} \end{aligned}$$

The above relation can be rewritten as:

$$\beta_{k+1} = \frac{g_{k+1}^T d_{k+1}}{-\tau_{k+1} + g_{k+1}^T d_k} \quad \text{.............................} (19)$$

Recall that if put $\tau_k = d_k^T y_k$, this method reduces to the DY method.

5. The proposed Conjugancy Coefficient

Consider the following quadratic model as:

$$f(\mathrm{x}) = \frac{1}{2} x^T A x + b^T x + c \quad \text{.............................} (20)$$

Where $A \in R^{n \times n}$ is a symmetric positive definite matrix, $b \in R^n$ and $c \in R$. Then

$$y_k = g_{k+1} - g_k = A\, s_k .$$

$$f_{k+1} = \frac{1}{2} x_{k+1}^T A x_{k+1} + b^T x_{k+1} + c$$

Substitute $\mathrm{x}_{k+1} = \mathrm{x}_k + \mathrm{s}_k$ into (20), then obtain:

$$\begin{aligned} f_{k+1} &= \frac{1}{2}(x_k + s_k)^T A (x_k + s_k) + b^T (x_k + s_k) + c \\ &= \frac{1}{2} x_k^T A x_k + \frac{1}{2} s_k^T A s_k + b^T x_k + b^T s_k + c \end{aligned}$$

$$f_{k+1} = f_k + \frac{1}{2} s_k^T A s_k + b^T s_k \quad \text{...........................} (21)$$

From Taylor series b=g then:

$$\begin{aligned} f_{k+1} &= f_k + \frac{1}{2} s_k^T A s_k + (g_k)^T s_k - g_k^T s_k \\ &= f_k - f_{k+1} + \frac{1}{2} s_k^T A s_k \end{aligned}$$

Since $As_k = g_{k+1} - g_k$, then

$$\begin{aligned} -g_k^T s_k &= f_k - f_{k+1} + \frac{1}{2} s_k^T (g_{k+1} - g_k) - \frac{1}{2} g_k^T s_k \\ &= f_k - f_{k+1} + \frac{1}{2} s_k^T g_{k+1} \end{aligned}$$

Multiplying both sides by 2

$$-g_k^T s_k = 2(f_k - f_{k+1}) + s_k^T g_{k+1} \quad \text{...............} (22)$$

It follows from Perry's conjugacy conditions

$$d_{k+1}^T y_k = -g_{k+1}^T s_k$$

$$-g_k^T s_k = 2(f_k - f_{k+1}) - d_{k+1}^T y_k$$

From eq.(3) and now assume that

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

Then get

$$-g_k^T s_k = 2(f_k - f_{k+1}) - (-g_{k+1} + \beta_k^{DY} d_k)^T y_k \quad \text{.................................} (23)$$

$$-g_k^T s_k = 2(f_k - f_{k+1}) - (-g_{k+1} + \frac{\|g_{k+1}\|^2}{d_k^T y_k} d_k)^T y_k \quad \text{.................................} (24)$$

Additionally, from equation. (15) that is $\beta_k = \frac{\|g_{k+1}\|^2}{\tau_{k+1}}$ equation.(24) imply that:

$$-g_k^T s_k = 2(f_k - f_{k+1}) - (-g_{k+1} + \frac{\beta_k \tau_{k+1}}{d_k^T y_k} d_k)^T y_k$$

$$-g_k^T s_k = 2(f_k - f_{k+1}) + g_{k+1}^T y_k - \frac{\beta_k \tau_{k+1}}{d_k^T y_k} d_k^T y_k$$

$$-g_k^T s_k = 2(f_k - f_{k+1}) + g_{k+1}^T y_k - \beta_k \tau_{k+1}$$

Which yields:

$$\beta_k^{New} = \frac{g_{k+1}^T y_k + 2(f_k - f_{k+1}) + g_k^T s_k}{\tau_{k+1}} \quad \text{........ (25)}$$

Since $\tau_{k+1} > 0$ then we suppose that:

If we set $\tau_k = \lambda \|g_k\|^2 + (1-\lambda) d_k^T y_k$ then:

$$\beta_{k+1}^{New} = \frac{g_{k+1}^T y_k + 2(f_k - f_{k+1}) + g_k^T s_k}{\lambda \|g_k\|^2 + (1-\lambda) d_k^T y_k} \quad \text{....... (26)}$$

6. Outline of The New Extended CG-Method.

Step 1: Given $x_1 \in R^n$; ($\varepsilon > 0$); (k) is an index of the algorithm

Step 2: Set k=1; $d_k = -g_k$

Step 3: Set $x_{k+1} = x_k + \lambda_k d_k$; λ_k is satisfy Wolfe Condition.

Step 4: If Powell restarting, $g_k^T g_{k+1} > 0.2 \|g_k\|^2$, satisfied then set:

$d_{k+1} = -g_{k+1}$ else set $d_{k+1} = -g_{k+1} + \beta_k^{New} d_k$

β_k^{New} is defined in (25), go to Step 2.

Step 5: If $\|g_{k+1}\| < \varepsilon$, stop else set k=k+1 go to Step 3.

7. The Convergence Analysis

7.1 Theoretical Properties for the New CG-Method.

In this section, the convergence behavior on the β_k^{New} method with exact line searches are explain. Hence, the following basic assumptions on the objective function is depend to find modify CG-Method.

Assumption (1) [15]

f is bounded below in the level set $L_{x_0} = \{x \in R^n | f(x) \le f(x_0)\}$; in some neighborhood U of the level set L_{x_0}, f is continuously differentiable and its gradient ∇f is Lipschitz continuous in the level set L_{x_0}, namely, there exists a constant L> 0 such that:

$$\|\nabla f(x) - \nabla f(y)\| \le L \|x - y\| \quad \text{for all x,y} \in L_{x_0} \quad \text{................................ (27)}$$

7.2 Sufficient Descent Property:

In this section will show that the proposed algorithm which defined in equations (26) and (3) satisfy the sufficient descent property which satisfy the convergence property.

Theorem (1):

The search direction d_k that generated by the proposed algorithm of modified CG satisfy the descent property for all k, when the step size λ_k satisfied the Wolfe conditions (10), (11).

Proof:

By use the indication to prove the descent property, for $k = 0$,

$d_0 = -g_0 \Rightarrow d_0^T g_0 = -\|g_0\| < 0$, after that then proved the theorem is true for $k = 0$, now assume that

$\|s_k\| \le \eta;\ \|g_{k+1}\| \le \gamma \quad and\ \|g_k\| \le \eta 2$ and assume that the theorem is true for any k i.e. $d_k^T g_k < 0$ or $s_k^T g_k < 0$ since $s_k = \lambda_k d_k$, now will prove that the theorem is true for $k+1$ then:

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

i.e.

$$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T y_k + 2(f_k - f_{k+1}) + g_k^T s_k}{\lambda \|g_k\|^2 + (1-\lambda) d_k^T y_k} d_k \quad \text{......... (28)}$$

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k + 2(f_k - f_{k+1}) + g_k^T s_k}{\lambda \|g_k\|^2 + (1-\lambda) d_k^T y_k} g_{k+1}^T d_k$$

$$g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2 = \frac{g_{k+1}^T y_k + 2(f_k - f_{k+1}) + g_k^T s_k}{\lambda \|g_k\|^2 + (1-\lambda) d_k^T y_k} g_{k+1}^T d_k$$

$$g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2 = \frac{g_{k+1}^T y_k . g_{k+1}^T d_k}{\lambda\|g_k\|^2 + (1-\lambda) d_k^T y_k} + \frac{2(f_k - f_{k+1}) g_{k+1}^T d_k}{\lambda\|g_k\|^2 + (1-\lambda) d_k^T y_k} + \frac{g_k^T s_k . g_{k+1}^T d_k}{\lambda\|g_k\|^2 + (1-\lambda) d_k^T y_k}$$

Using Wolfe condition then get:

$$\leq \frac{g_{k+1}^T y_k . g_{k+1}^T d_k}{\lambda\|g_k\|^2 + (1-\lambda) d_k^T y_k} + \frac{-2\rho\lambda . g_k^T d_k g_{k+1}^T d_k}{\lambda\|g_k\|^2 + (1-\lambda) d_k^T y_k} - \frac{g_k^T s_k . \rho g_k^T d_k}{\lambda\|g_k\|^2 + (1-\lambda) d_k^T y_k} \quad \text{.................................} (29)$$

By using the relation ($u^T v = \frac{1}{2}(\|u\|^2 + \|v\|^2)$)

$$\leq \frac{g_{k+1}^T y_k . g_{k+1}^T d_k}{\lambda\|g_k\|^2 + (1-\lambda) d_k^T y_k} - \frac{2\rho\lambda . \left(\frac{1}{2}(\|g_k\|^2 + \|d_k\|^2\right)\left(\frac{1}{2}(\|g_{k+1}\|^2 + \|d_k\|^2\right)}{\lambda\|g_k\|^2 + (1-\lambda) d_k^T y_k}$$

$$g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2 \leq \frac{g_{k+1}^T y_k . g_{k+1}^T d_k}{\lambda\|g_k\|^2 + (1-\lambda) d_k^T y_k} \quad \text{.................................} (30)$$

By divide equation (30) on $\|g_{k+1}\|^2$

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \leq \frac{g_{k+1}^T y_k . g_{k+1}^T d_k}{(\lambda\|g_k\|^2 + (1-\lambda) d_k^T y_k)\|g_{k+1}\|^2}$$

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \leq \frac{g_{k+1}^T y_k . g_{k+1}^T d_k}{(\lambda\|g_k\|^2 + d_k^T y_k)\|g_{k+1}\|^2}$$

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \geq \frac{y_k^T g_{k+1}}{\lambda\|g_k\|^2 + \|d_k\|\|y_k\|} . \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2} \quad \text{.................................} (31)$$

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \leq \frac{\|y_k\|\|g_{k+1}\|}{\lambda\|g_k\|^2 + \|d_k\|\|y_k\|} \frac{\|g_{k+1}\|\|d_k\|}{\|g_{k+1}\|^2}$$

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \leq \frac{\|y_k\|\|g_{k+1}\|}{\lambda\|g_k\|^2 + \|d_k\|\|y_k\|} \frac{\|g_{k+1}\|\|d_k\|}{\|g_{k+1}\|^2} \quad \text{.................................} (32)$$

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \leq \frac{\|y_k\|\|d_k\|}{\lambda\|g_k\|^2 + \|d_k\|\|y_k\|} \quad \text{.................................} (33)$$

$$\frac{\|g_{k+1}\|^2}{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2} \leq \frac{\lambda\|g_k\|^2 + \|d_k\|\|y_k\|}{\|y_k\|\|d_k\|} = \delta > 1 \quad \text{.................................} (34)$$

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \leq \frac{1}{\delta}$$

$$g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2 \leq \frac{1}{\delta}\|g_{k+1}\|^2 \quad \text{........................} (35)$$

$$g_{k+1}^T d_{k+1} \leq -(1-\frac{1}{\delta})\|g_{k+1}\|^2 \quad \text{.............................} (36)$$

Let $c = (1-\frac{1}{\delta})$

Then

$$g_{k+1}^T d_{k+1} \leq -c\|g_{k+1}\|^2 \quad \text{...............................} (37)$$

7.3 Global Convergence Property:

***Lemma 1:* [16]**

Let assumptions (i) and (ii) hold and consider any conjugate gradient method (2) and (3), where d_k is a descent direction and λ_k is obtained by the strong Wolfe line search. If

$$\sum_{k\geq 1}\frac{1}{\|d_k\|^2} = \infty \quad \text{...} (38)$$

Then

$$\liminf_{k\to\infty}\|g_k\| = 0 \quad \text{...} (39)$$

For uniformly convex functions which satisfy the above assumptions, the norm of d_{k+1} given by equation (28) can prove bounded above. Assume that the function f is a uniformly convex function, i.e. there exists a constant $\mu \geq 0$ such that for all $x, y \in S$,

$$(g(x) - g(y))^T (x - y) \geq \mu\|x - y\|^2, \quad \text{................................} (40)$$

***Theorem 2:* [17]**

Suppose that the assumptions (i) and (ii) hold. Consider the algorithm (2), (26). If $\|s_k\|$

tends to zero and there exists nonnegative constants $\eta 1$ and $\eta 2$ such that:

$$\|g_k\|^2 \geq \eta 1 \|s_k\|^2 , \quad \ldots\ldots (41)$$

And *f* is a uniformly convex function, then

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \quad \ldots\ldots (42)$$

Proof:
From

$$\beta_{k+1}^{New} = \frac{g_{k+1}^T y_k + 2(f_k - f_{k+1}) + g_k^T s_k}{\lambda \|g_k\|^2 + (1-\lambda) d_k^T y_k} \quad \ldots\ldots (43)$$

From equation. (43) Then get:

$$\left|\beta_{k+1}^{New}\right| = \left| \frac{g_{k+1}^T y_k + 2(f_k - f_{k+1}) + g_k^T s_k}{\lambda \|g_k\|^2 + (1-\lambda) d_k^T y_k} \right| \leq \frac{\|g_{k+1}\| \|y_k\| + 2|(f_k - f_{k+1})| + \|g_k\| \|s_k\|}{\lambda \|g_k\|^2 + (1-\lambda) \|d_k\| \|y_k\|} \quad \ldots\ldots (44)$$

But $\|y_k\| \leq L \|s_k\|$,. Then

$$= \frac{\|g_{k+1}\| L. \|s_k\| + 2|(f_k - f_{k+1})| + \|g_k\| \|s_k\|}{\lambda \eta 1 \|s_k\|^2 + (1-\lambda) \frac{\|s_k\|}{\lambda} L \|s_k\|} \leq \frac{L\gamma\eta + 2|(f_k - f_{k+1})| + \eta 2\eta}{\lambda \eta 1 \eta \|s_k\| + (1-\lambda) \frac{\|s_k\|}{\lambda} L\eta} \quad \ldots\ldots (45)$$

Let

$$A = (f_k - f_{k+1})$$

$$\left|\beta_k^N\right| \leq \frac{L\eta\gamma + 2A + \eta 2\eta}{\lambda \eta 1 \eta \|s_k\| + (1-\lambda) \frac{\|s_k\|}{\lambda} L\eta} \quad \ldots\ldots (46)$$

Hence,

$$\|d_{k+1}\| \leq \|g_{k+1}\| + \left|\beta_k^N\right| \|s_k\| \quad \ldots\ldots (47)$$

$$\|d_{k+1}\| \leq \gamma + \frac{L\eta\gamma + 2A + \eta 2\eta}{\lambda \eta 1 \eta \|s_k\| + (1-\lambda) \frac{\|s_k\|}{\lambda} L\eta} \|s_k\| \quad \ldots\ldots (48)$$

$$= \gamma + \frac{L\eta\gamma + 2A + \eta 2\eta}{\lambda \eta 1 \eta + (1-\lambda) \frac{L\eta}{\lambda}}$$

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

$$\frac{1}{\left(\gamma + \frac{L\eta\gamma + 2A + \eta 2\eta}{\lambda \eta 1 \eta + (1-\lambda) \frac{L\eta}{\lambda}}\right)^2} \sum_{k\geq 1} 1 = \infty \quad \ldots\ldots (49)$$

8. Results and Discussion

The practical side includes some numerical results which are obtained with the implementation of new algorithm (on asset of unconstrained optimization).

Large scale unconstrained optimization problems are selected (generalized form).

For each test function, n=1000, 10000 is the number of variable consider as numerical experiment. By using the standard wolfe conditions (4) and (5) with stopping criteria is $\|g_{k+1}\| \leq 10^{-6}$.

All the computations in this part are carried out by using Fortran 90 Language.

Method and FR method (7) are compared in this research. The preliminary numerical results of tests are show in Tables (1) and (2). The first column is "test fun" (name of test function), the second column "NOI" denoted the number of iterations, the third column "NOF" denoted the number of calculated functions and the fourth column "MIN" denoted the minimum values.

Table (1)
Comparison between standard method and modified method with respect to (NOI and NOF) for n = 1000.

Test Functions	*Fletcher - Reeves*			*Dai- Yuan*			*modified method*		
	NOI	*NOF*	*MIN*	*NOI*	*NOF*	*MIN*	*NOI*	*NOF*	*MIN*
Extended Rosenbrock SROSENBR (CUTE)	fail	fail	fail	fail	fail	fail	62	123	6.32E-15
Extended White & Holst	fail	fail	fail	fail	fail	fail	79	152	1.20E-15
Extended Beale	1108	1154	1.04E-12	1126	1173	7.18E-13	31	56	8.84E-14
Penalty	24	61	8.83E+02	24	60	8.83E+02	13	37	8.83E+02
Generalized Tridiagonal	65	776	9.97E+02	131	3005	2.00E+03	49	731	9.97E+02
Generalized Tridiagonal 2	233	280	9.58E-01	233	280	9.58E-01	54	88	1.16E-14
Diagonal	641	677	3.69E-13	661	698	1.71E-13	17	32	2.00E-16
Extended Himmelblau	30	61	8.22E-15	30	61	8.29E-15	18	33	5.89E-16
Extended Maratos	fail	fail	fail	fail	fail	fail	69	140	-5.00E+02
Extended Wood WOODS (CUTE)	fail	fail	fail	fail	fail	fail	282	525	1.21E-13
Extended Hiebert	fail	fail	fail	fail	fail	fail	99	217	1.57E-12
Extended Quadratic Penalty QP2	fail	fail	fail	fail	fail	fail	50	104	6.69E-15
ARWHEAD (CUTE)	1546	2433	0.00E+00	1540	2516	0.00E+00	42	426	0.00E+00
NONDIA (CUTE)	fail	fail	fail	fail	fail	fail	25	47	3.44E-17
DQDRTIC (CUTE)	1589	1632	1.22E-13	1596	1639	1.08E-13	171	285	2.21E-13
Broyden Tridiagonal	83	127	1.36E-14	83	127	1.41E-14	37	62	1.48E-14
LIARWHD (CUTE)	fail	fail	fail	fail	fail	fail	57	107	6.22E-15
DENSCHNA (CUTE)	23	37	8.86E-14	25	38	1.28E-13	15	27	9.09E-15
DENSCHNC (CUTE)	132	170	8.79E-14	132	170	8.82E-14	35	70	5.12E-14
Extended Block-Diagonal BD2	130	166	1.93E-13	132	169	1.21E-13	35	70	1.72E-14
Generalized quartic GQ1	10	24	1.30E-13	12	28	3.66E-14	9	22	1.38E-15
Generalized quartic GQ2	118	153	1.22E-13	117	152	2.23E-13	45	71	2.65E-13
FLETCHCR (CUTE)	32	62	2.09E-16	32	62	2.12E-16	32	56	1.12E-16
HIMMELBH (CUTE)	21	43	-5.00E+02	21	43	-5.00E+02			-5.00E+02

As shown in Table (1) a comparison between standard method (Fletcher and Reeves, Daiand Yuan) and modified method with respect to (NOI and NOF) for n=1000, The results are obtained in modified method are better than the results in standard method (Fletcher and Reeves, Daiand Yuan).

Table (2)
Comparison between standard method and modified method with respect to (NOI and NOF) for n = 10000.

Test Functions	*Fletcher - Reeves*			*Dai- Yuan*			*modified method*		
	NOI	*NOF*	*MIN*	*NOI*	*NOF*	*MIN*	*NOI*	*NOF*	*MIN*
Extended Rosenbrock SROSENBR (CUTE)	fail	fail	fail	fail	fail	fail	54	109	4.37E-17
Extended White & Holst	fail	fail	fail	fail	fail	fail	91	177	1.61E-16
Extended Beale	1194	1243	1.15E-12	1222	1272	6.06E-13	37	65	1.02E-12
Penalty	25	71	9.45E+03	24	65	9.45E+03	13	43	9.45E+03
Generalized Tridiagonal	34	64	1.00E+04	2001	2015	3.39E-05	27	59	1.00E+04
Generalized Tridiagonal 2	1458	1527	3.41E+00	1479	1557	3.41E+00	82	132	9.58E-01
Diagonal	701	740	3.11E-13	701	740	3.30E-13	17	32	7.59E-15
Extended Himmelblau	32	65	6.87E-15	32	65	6.93E-15	18	34	1.46E-16
Extended Maratos	fail	fail	fail	fail	fail	fail	303	661	-5.00E+03
Extended Wood WOODS (CUTE)	fail	fail	fail	fail	fail	fail	243	447	1.65E-13
Extended Hiebert	fail	fail	fail	fail	fail	fail	2001	2097	4.70E-05
Extended Quadratic Penalty QP2	fail	fail	fail	fail	fail	fail	53	120	4.15E-14
ARWHEAD (CUTE)	fail	fail	fail	fail	fail	fail	42	645	0.00E+00
NONDIA (CUTE)	fail	fail	fail	fail	fail	fail	30	60	1.38E-13
DQDRTIC (CUTE)	1591	1634	1.17E-13	1598	1641	1.03E-13	206	341	1.53E-13
Broyden Tridiagonal	1169	1217	8.74E-01	1169	1217	8.74E-01	124	207	3.97E-01
LIARWHD (CUTE)	fail	fail	fail	fail	fail	fail	59	116	6.83E-19
DENSCHNA (CUTE)	25	40	2.02E-13	30	48	2.08E-13	16	28	6.66E-16
DENSCHNC (CUTE)	148	191	7.35E-14	137	179	1.33E-13	35	68	1.86E-14
Extended Block-Diagonal BD2	142	181	9.80E-14	140	179	1.64E-13	36	72	1.11E-13
Generalized quartic GQ1	13	32	1.12E-14	13	32	1.37E-14	11	28	4.78E-15
Generalized quartic GQ2	116	154	1.67E-13	119	158	1.45E-13	48	75	1.60E-13
FLETCHCR (CUTE)	233	2075	4.96E+01	245	2490	4.96E+01	90	585	4.96E+01
HIMMELBH (CUTE)	31	183	-5.00E+03	22	46	-5.00E+03	14	130	-5.00E+03

Table (2) shows a comparison between standard method (Fletcher and Reeves, Daiand Yuan) and modified method with respect to (NOI and NOF) for n=10000, The results are obtained in modified method are better than the results in standard method (Fletcher and Reeves, Daiand Yuan).

Table (3)
Result of PSO with new CG recognized ear image of database.

Number of subject	*No of features*	*Recognized ear image*	*Recognition rate (%)*
120	6	120	100%
140	7	140	100%

Table (4)
Result of PSO with new CG recognized ear image of test.

Number of subject	*No of features*	*Recognized ear image*	*Recognition rate (%)*
50	6	33	66
50	7	35	70

Table (3) show two number of subject 18 person, Table (3) cluster subject 120(6) is shows the (6) features data ear image for person and 140(7) shows the (7) features data ear image for person. We shows the recognized ear image with the recognition rate is (100%). While Table (4) explain recognized ear image (33) for 50(6) and (35) for 50(7) with unrecognized ear image

17 and 15 respectively with recognized ear rate 66% and 70% respectively.

PSO based feature selection algorithm found to generate excellent recognition result with the minimal set of selected features.

9. Conclusion

PSO is computation paradigm based on the idea of collaboration behavior inspired by the social behavior of bird flocking or fish schooling. Feature selection algorithm is utilized to search the feature space for the optimal feature subset where features are carefully selected to a well defined in terms of maximizing the class separation. The application of the proposed clustering algorithm to the problem of segmentation of images is investigated. A clustering algorithm with minimal user in deference is developed in this work.

The numerical results of our finding of standard method and modified method are very encouraging by using the proposed algorithm. Through the results and minimum errors for the large scale optimization problem. These out come of results is to show that modified method are more effective than the standard method. And the accuracy of the modified method gives a good optimization problem.

References

[1] Hestenes, M. R., Stiefel, E.L, "Methods of Conjugate gradient for solving linear systems", J. Res. Ntal. Bur. stant., 49, 409-346, 1952.

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

[3] Polak, E. and Ribiere, G., "Note sur la convergence de directions conjugees", Rev. Francaise Informat Recherche Opertionelle, 3e Annee 16, 35-43, 1969.

[4] Polyak, B. T., "The conjugate gradient method in extreme problems", USSR Comp Math. Phys., 9, 94-112, 1969.

[5] Fletcher, R., "Practical Methods of Optimization" vol. 1: Unconstrained Optimization, John Wiley & Sons, New York, 1987.

[6] Liu, Y. and Storey, C., "Efficient generalized conjugate gradient algorithms", Part 1: Theory, J. Optim. Theory Appl., 69, 129-137, 1991.

[7] Dai, Y. H. and Yuan, Y., "A nonlinear conjugate gradient method with a strong global convergence property", SIAM J. Optim., 10, 177-182, 1999.

[8] Wolfe, P., "Convergence conditions for ascent methods" I, SIAM Review, 11, 226-235, 1969.

[9] Wolfe, P., "Convergence conditions for ascent methods" II, Some corrections, SIAM Review, 13, 185-188, 1971.

[10] Omran, M G. H., "Particle Swarm Optimization Methods for Pattern Recognition and Image Processing" thesis, University of Pretoria, 2005.

[11] Abed, M Ali and Alasadi, H Ali, "Simplifying Handwritten Characters Recognition Using a Particle Swarm Optimization Approach", european academic research, I(1), 2286-4822, 2013.

[12] Bratton, Daniel and Kennedy, James, "Defining a Standard for Particle Swarm Optimization", IEEE Swarm Intelligence Symposium, 2007.

[13] Yabe, Hiroshi and Sakaiwa Naoki,. "A New Nonlinear Conjugate Gradient Method for Unconstrained Optimization", Tokyo University of Science Hitachi Information Systems, 48(4), 284-296, 2005.

[14] Dai, Y. H. and Liu, G.H., Sun, D.F., Yin, X., and Yuan, Y., "Convergence properties of nonlinear conjugate gradient thods. SIAM J. Optim. 10, 348-358, 1999.

[15] Glibert, J.C. and Nocedal, J,. "Global Convergence Properties of Conjugate gradient methods for optimization", SIAM J. optimization, 2(1), 21-42, 1992.

[16] Kiwiel, K.C., Murty, K., "Convergence of the steepest descent method for minimizing quasiconvex functions. J. Optim. Theory Appl. 89(1), 221-226, 1996

[17] Andrei, N, "Scaled conjugate gradient algorithms for unconstrained optimization" Comput. Optim. Appl. 38, 401-416, 2007.

الخلاصة

تم في هذا البحث اشتـقاق معامل ترافق محسن لطريقة المتـجهات المترافقة. هذا التحسين استخدام توسيع طريقة (Dai and Yuan Method) الحل مسائل البرمجة غير الخطية، كما تم تطبيق خوارزمية الطيور في هذا العمل، لتحديد المعاملات باستخدام مميزات التقنيات المحددة. تم إثبات خاصية الانـحدار الكافي (sufficient descent) وخاصية التقارب الشامل للخوارزمية المقترحة، تم الحصول على نتائج عددية مشجعة جدا لمسائل الأمثلة ذات القياس العالي مقارنة مع الطرق القياسية.

النتائج التجريبية وضحت ان طريقة خوارزمية الطيور كانت نتائجها مميزة مع اقل خطأ في المميزات المختارة، أخيرا طريقة خوارزمية الطيور المقترحة التي اعتمده التقريبات ومعاملات أداء الطريقة قد قيمت.