ROBUST IDENTIFICATION AND PREDICTION USING WILCOXON NORM AND PARTICLE SWARM OPTIMIZATION

Babita Majhi[1], G. Panda[1] and B. Mulgrew[2]

[1]Dept. of Electronics & Communication Engineering, National Institute of Technology, Rourkela – 769008, India
phone: +91-661-2462455, fax: +91-661-2462963, email: babita.majhi@gmail.com , ganapati.panda@gmail.com,
web:www.nitrkl.ac.in
[2]Institute of Digital Communication, University of Edinburgh, UK, email: B. Mulgrew@ed.ac.uk

ABSTRACT

The paper introduces a novel method of robust identification of complex plants and prediction of bench mark time series. It is assumed that training samples used contain strong outliers and the cost function chosen in the proposed model is a robust norm called Wilcoxon norm. The weights of the models are updated using population based PSO technique which progressively reduces the robust norm. To demonstrate the robust performance of the proposed technique standard identification and prediction problems are simulated and the results are compared with those obtained by conventional MSE norm based minimization method. A significant improvement in performance is observed in all cases.

1. INTRODUCTION

Identification of complex nonlinear plants finds many applications in control, power system, instrumentation and telecommunication [1]. Accurate and fast identification of such real time nonlinear processes is still a difficult problem. Further, building of proper models of a plant become challenging both for prediction and identification when outliers are present in the training sample. Under such adverse conditions the training of models becomes ineffective when conventional mean square error based on least mean square (LMS) or recursive least square (RLS) [2] type algorithms are used for training. Similarly robust time series prediction is important in many forecasting applications.

Various evolutionary computing tools such as genetic algorithm (GA) [3], particle swarm optimization (PSO) [4], bacterial foraging optimization (BFO) [5] and ant colony optimization (ACO) [6] have been reported and applied for optimization and identification tasks. In case of the derivative free algorithms conventionally the mean square error (MSE) is used as the fitness or cost function. Use of MSE as cost function leads to improper training of adaptive models when outliers are present in the training samples. Therefore there is a need for robust identification of complex plants in presence of strong outliers. It is known in statistics that linear regressors developed using Wilcoxon norm (W-norm) [7] are robust against outliers. Using such norm new robust machines have recently been reported for approximation of nonlinear functions [8]. In the present investigation we develop a new method of robust identification and prediction of complex time series by minimizing the W-norm of errors of model using a derivative free PSO technique. The identification and prediction performance of the new method is evaluated through simulation study and is compared with the results obtained from corresponding error square norm based PSO technique.

Section 2 deals with the basic principle of system identification and time series prediction where as the fundamental of particle swarm optimization is dealt in Section 3. Section 4 proposes the development of robust identification and prediction models using minimization of W-norm by PSO. Exhaustive simulation study for identification and prediction of benchmark problems using two different norms is carried out and the results are presented in Section 5. Finally concluding remarks are included in Section 6.

2. ADAPTIVE SYSTEM IDENTIFICATION AND PREDICTION USING ROBUST NORM

The block diagram of an adaptive system identification scheme is shown in Fig. 1.

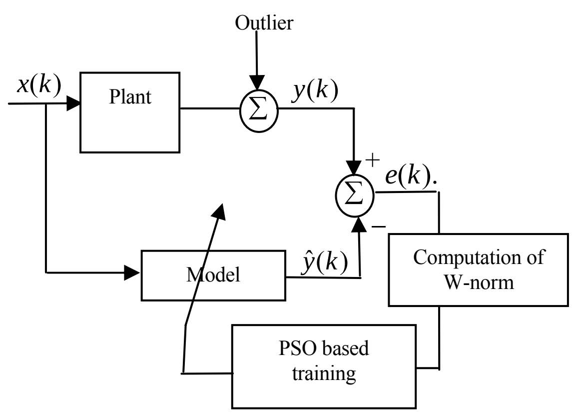


Fig. 1 Adaptive system identification model using W-norm and PSO algorithm

At any time instant k, $x(k), y(k)$ *and* $\hat{y}(k)$ represent the input, output of the plant and estimated output of the model respectively. The difference of these two output produces an error, $e(k)$. The output $y(k)$ is obtained by combining the system output and the outliers present at random locations.

Conventionally the MSE is employed as the cost function in deriving various iterative learning rules. It is observed that the learning rules exhibit poor training performance when strong outliers are present in the training samples. The W-norm of errors of the model has proven to be a robust norm and the resulting model is expected to be robust to outliers during training. In this paper recursive minimization of this norm by PSO is chosen to obtain an improved and robust identification model.

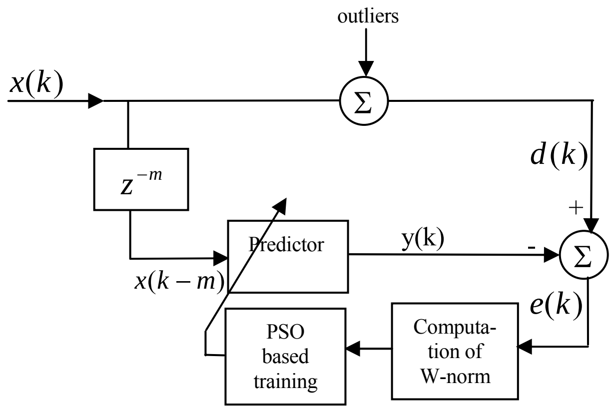


Fig. 2 Robust adaptive predictor using W-norm based PSO algorithm

Fig. 2 shows a robust adaptive predictor using PSO based training algorithm. The adaptive model employs the past samples, $x(k-m)$ of the times series and forecasts its future samples $x(k)$. It computes the robust W-norm of the error vector and progressively minimizes the same by updating its weights using PSO. The desired signal is generated by adding outliers at 10% to 50% random locations. The process of time series prediction consists of pre-processing of the data, selection of model to be used, parameter estimation of the model using training data and validation of the model using past data.

3. FUNDAMENTALS OF PARTICLE SWARM OPTIMIZATION

Particle swarm optimization (PSO) introduced by Eberhart and Kennedy is a population based optimization algorithm like the Genetic algorithm (GA). It has already been applied successfully to function optimization, image analysis, data clustering and structure optimization [9]. It imitates the "flying nature" of n-dimensional swarm (population) of particles (birds) – each of which provides a possible solution to the optimization problem - through a problem space, in search of a single optimum or multiple optima. The flying behavior can be mimicked as follows:

The swarms initially have a population of random solutions. Each potential solution, called a particle, is given a random velocity and is flown through the search space. The particles have reminiscence and each particle keeps tracks of the previous best position, called *pbest* and the corresponding fitness function is evaluated and stored. Again *gbest* denotes the position of the particle having highest fitness value for the current iteration. All the particles always tend to move in the direction of their *pbes*t and *gbes*t. The velocity and position of ith particle is changed according to (1) and (2) respectively so that the cost function defined in (8) is minimized progressively.

$$V_i(d) = wV_i(d) + c_1 * rand * (P_i(d) - X_i(d)) + c_2 * rand * (P_g(d) - X_i(d)) \tag{1}$$

$$X_i(d) = X_i(d) + V_i(d) \tag{2}$$

where $V_i(d)$ and $X_i(d)$ represent the velocity and position of the ith particle corresponding to dth dimension respectively and $rand$ is a uniform random number in the range [0,1]. $P_g(d)$ and $P_i(d)$ are the dth dimensional positions of the *gbest* and *pbest* respectively. w, c_1, c_2 are constants whose values are suitably chosen to achieve the best possible solution. The entire process is repeated for some fixed number of iterations until the global optimum is reached. At this stage *gbest* provides the desired solution.

4. NEW IDENTIFICATION AND PREDICTION MODELS USING WILCOXON NORM MINIMIZATION BY PSO

The Wilcoxon norm, a robust cost function is used for the development of identification and prediction models. The PSO is employed to iteratively minimize this norm of the errors of the model and hence the resulting model is expected to be robust.

Robust Cost Function (Wilcoxon Norm) [6, 7]

A score function is first defined as an increasing function $\phi(u) : [0,1] \rightarrow \Re$ such that

$$\int_0^1 \phi^2(u)du < \infty \tag{3}$$

The score function has the characteristics

$$\int_0^1 \phi(u)du = 0 \text{ and } \int_0^1 \phi^2(u)du = 1 \tag{4}$$

The score associated with the score function ϕ is defined as

$$a_\phi(i) = \phi\left(\frac{i}{l+1}\right), \quad i \in l \tag{5}$$

where l is a fixed positive integer.

From (4) it may be observed that $a_\phi(1) \le a_\phi(2) \le .... \le a_\phi(l)$. The Wilcoxon norm [6, 7] on $\Re'$ is defined as

$$C_1 = \sum_{i=1}^{l} a(R(v_i))v_i = \sum_{i=1}^{l} a(i)v_i, \quad v = [v_1, v_2, ..., v_l]^T \in \Re' \tag{6}$$

where $R(v_i)$ denotes the rank of v_i among $v_1, v_2, ..., v_l$, $v_{(1)} \le v_{(2)} \le ......v_{(l)}$ are the ordered values of $v_1, v_2, \ldots\ldots\ldots\ldots, v_l$, $a(i) = \phi[i/(l+1)]$. In statistics different types of score functions have been dealt but the commonly used one is given by $\phi(u) = \sqrt{12}(u - 0.5)$. The weight-updates of the models of Figs. 1 and 2 are carried out by minimizing the cost function of the errors defined in (6) using PSO algorithm. Subsequent steps involved are detailed as follows

Let the error vector of p th particle at k th generation due to application of N input samples to the model be represented as $[e_{1,p}(k), e_{2,p}(k), \ldots\ldots\ldots\ldots, e_{N,p}(k)]^T$. The errors are then arranged in an increasing manner from which the rank $R\{e_{n,p}(k)\}$ of each n th error term is obtained. The score associated with each rank of the error term is evaluated as

$$a(i) = \sqrt{12}\left(\frac{i}{N+1} - 0.5\right) \quad (7)$$

where $(1 \le i \le N)$ denotes the rank associated with each error term. At k th generation of each p th particle the Wilcoxon norm is then calculated as

$$C_p(k) = \sum_{i=1}^{N} a(i)\, e_{i,p}(k) \quad (8)$$

The learning strategy using PSO continues until the cost function in (8) decreases to the possible minimum value. At this stage the training is discontinued and the corresponding global best weight vector represents the optimal weights of these models.

5. SIMULATION STUDY

Simulation study is carried out to assess the prediction of standard time series and identification of some benchmark plants. The outliers are random values within some predefined range and are added at random locations (10% to 50%) of the training samples. The desired signal is obtained by adding outliers. The MSE and W-norm are used as the cost functions of scheme-1 and scheme-2 respectively. The performance of the proposed scheme is obtained from simulation study and compared with those obtained by scheme-1 (MSE-norm). In each example the number of population = 30, $c_1 = c_2 =$ 1.042 and linearly decreasing w from 0.9 to 0.4 are taken. These optimized values provide best performance in all cases.

Example 1 : Prediction of Mackey Glass series

The Mackey-Glass Series (MGS) is a standard benchmark system used for prediction purpose. This is a chaotic time series generated by solving the time-delay differential equation

$$\frac{dx(t)}{dt} = -bx(t) + a\frac{x(t-\tau)}{1 + x(t-\tau)^{10}} \quad (9)$$

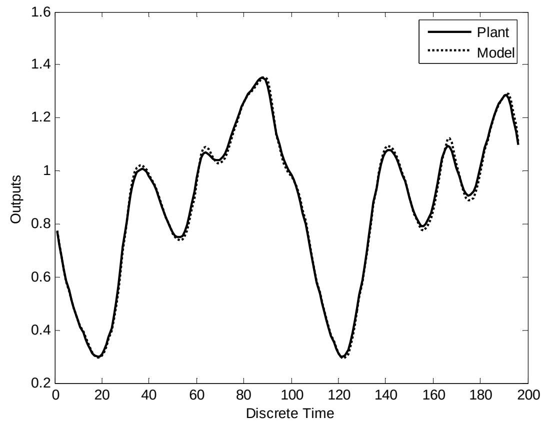
This MGS is periodic for $\tau < 17$ and is non-periodic otherwise. Initial values are taken as random values. The differential equation is solved using Euler's method. A set of 1100 samples is generated with b = 0.9, a = 0.2 and $\tau = 30$. The first 100 samples is discarded due to its random nature. Out of the remaining 1000 samples, 800 samples are used as training data and the rest 200 as test samples.

The model of the system is represented as

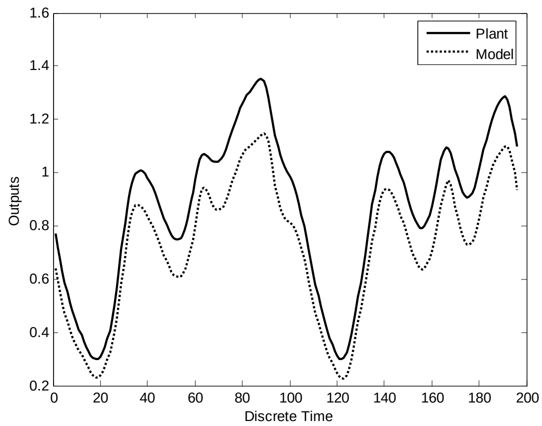
$$x(t+p) = f\{x(t), x(t-\tau), x(t-2\tau), \ldots\ldots, x(t-(N-1)\tau\} \quad (10)$$

where $p = 4$ and $N = 4$. In this example four sample ahead prediction of the MGS is made. Thus $p = 4$ and $N = 4$ are used.

The training data set is corrupted by adding random values from a uniform distribution of [−15, 15] to the uncorrupted data set. Simulation is carried out in presence of 10% to 50% of outliers in the training signal. The response matching obtained from the simulation is shown in Figs. 3(a) and (b) for 50% outliers. From these figures it is observed that scheme-2 based model makes significant better forecast in presence of 50% outliers in the training signal where as the scheme-1 based model fails to correctly predict future values.



(a) Using Scheme-2 (W-norm) learning with 50% outliers

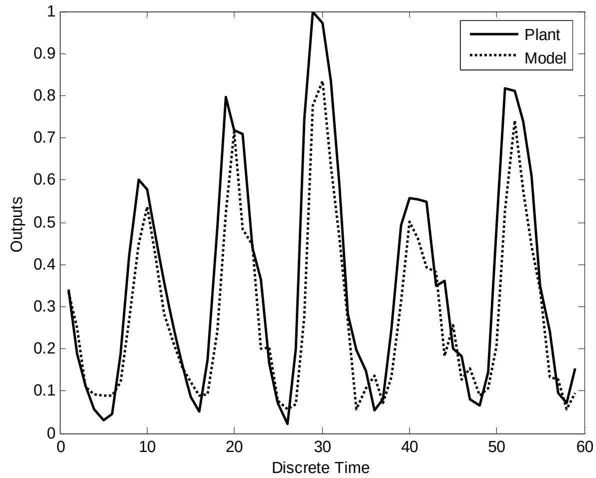


(b) Using Scheme-1(MSE-norm) learning with 50% outliers

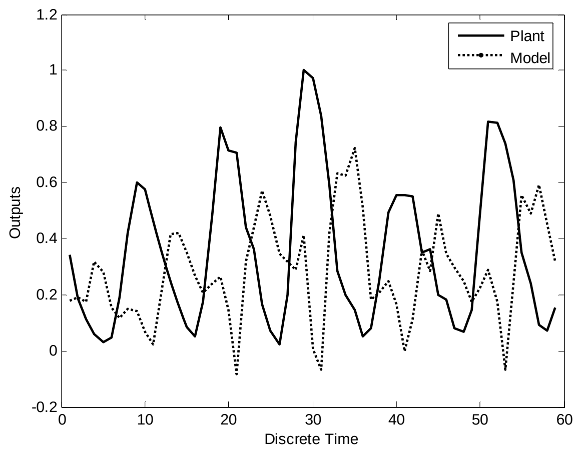
Fig. 3 Output response matching of Example 1

Example 2 : Forecasting of Sunspot time series

The series consists of 288 data points of yearly averages of sunspots starting from the year 1700 to the year 1987. The sunspots problem is a typical time series prediction problem, in which the task is to predict the sunspots number for the following year. In this example four sample ahead prediction of the sunspot is made. Out of the 288 data points first 225 data is used for training and rest 63 data used for testing purpose. The training data set is corrupted by adding random values from a uniform distribution defined between [−15, 15] to the uncorrupted data. Simulation is carried out in presence of 10% to 50% of outliers in the training signal. The response matching of the system with 40% outliers is given in Figs. 4(a) and (b). The forecasting performance of scheme-1 is severely degraded at 40% outliers. It is clearly observed that scheme-2 model provides superior prediction in comparison to the scheme-1 based model particularly in presence of outliers.



(a) Using Scheme-2 (W-norm) learning with 40% outliers

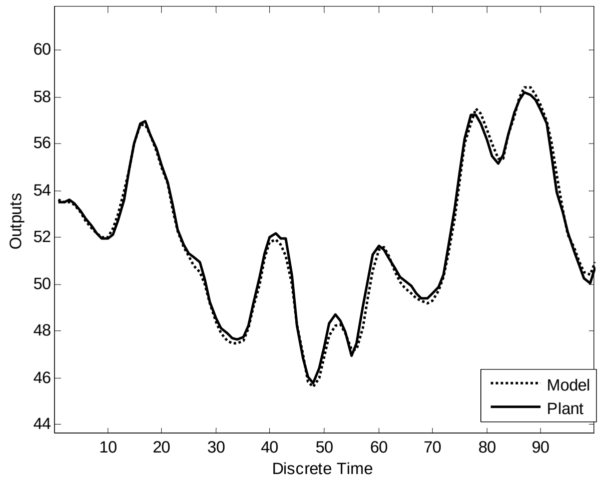


(b) Using Scheme-1 (MSE-norm) learning with 40% outliers

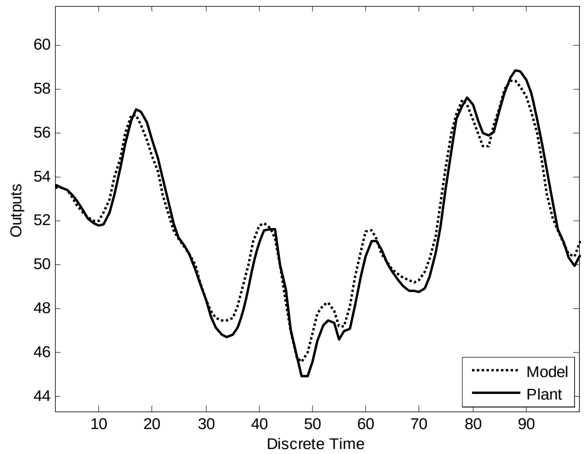
Fig. 4 Output response matching of Example 2

Example 3 : Identification of Box-Jenkin's System

The 296 input-output samples are generated with a sampling period of 9s. The gas combustion process has one variable, gas flow, $x(k)$, and one output variable , the concentration of CO_2, $y(k)$. The output $y(k)$ is influenced by four past output samples $y(k-1), y(k-2), y(k-3)$ and $x(k-1)$ .Uniformly distributed random values between [−3, 3] is added at 10% to 50% random locations of the desired samples. Figs. 5 (a) and (b) display the actual and estimated output values obtained by using scheme-2 and scheme-1 methods of training respectively. It is evident from these figures that scheme-2 provides better identification performance in presence of strong outliers in the training signal in comparison to scheme-1 based method.



(a) Using Scheme-2 (W-norm) learning with 50% outliers



(b) Using Scheme-1 (MSE-norm) learning with 50% Outliers

Fig. 5 Output response matching of Example 3

Example 4 : Identification of SISO dynamic system [10]

The plant in this case is described by the difference equation

$$y(k+1) = f[y(k), y(k-1), y(k-2), x(k), x(k-1)] \quad (11)$$

where the unknown nonlinear function f is given by

$$f[a_1, a_2, a_3, a_4, a_5] = \frac{a_1 a_2 a_3 a_5 (a_3 - 1.0) + a_4}{1.0 + a_2^{\ 2} + a_3^{\ 2}} \quad (12)$$

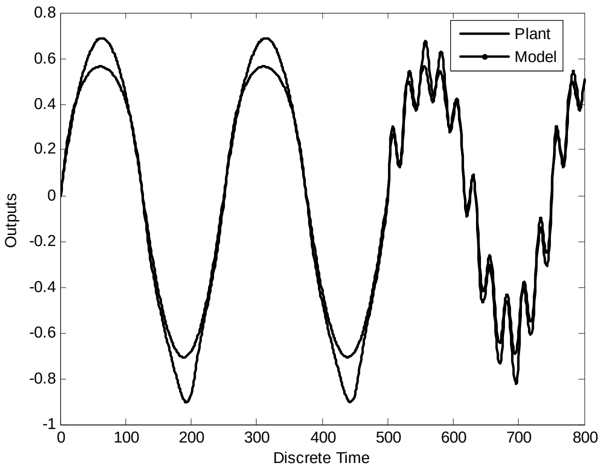
The series-parallel model used for identification of this plant is given as

$$\hat{y}(k+1) = \mathrm{N}[y(k), y(k-1), y(k-2), x(k), x(k-1)] \quad (13)$$

In case of scheme-1 and scheme-2 the basic model is a functional link artificial neural network (FLANN) [11] structure. Its input and output are expanded to six and nine terms respectively using trigonometric expansion. An uniformly distributed random signal in the interval [-1, 1] is used as input. The outliers are uniformly distributed random values within the range of -1 to +1 and are added at random locations (10% to 50%) of the training samples. During the testing phase, the effectiveness of the proposed models are evaluated by using the test signal

$$x(k) = \begin{cases} \sin \dfrac{2\pi k}{250} \;\; for\ k \le 250 \\ 0.8 \sin \dfrac{2\pi k}{250} + 0.2 \sin \dfrac{2\pi k}{25} \;\; for\ k > 250 \end{cases} \quad (14)$$

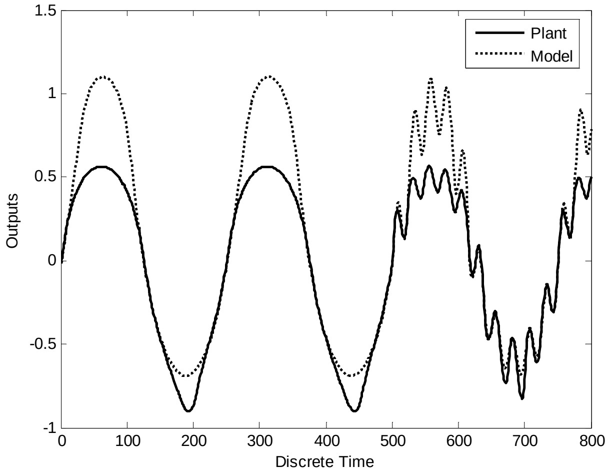
Figs. 6(a)-(b) show the comparative performance of the output response of two models. The simulation results indicate that the identification performance is better in the proposed model than the MSE-norm based model.



(a) Using Scheme-2 (W-norm) learning with 40% outliers

6 CONCLUSION

The adaptive identification and prediction tasks have been formulated as optimization problems. Instead of using conventional MSE as fitness function, robust W-norm of errors is chosen for minimization using PSO technique. Robust prediction of Mackey Glass and Sunspot time series when strong outliers are present in training set is carried out through simulation. Similarly a PSO based method for robust identification of standard plants is also suggested by way of minimizing W-norm. The proposed techniques is robust because it provides excellent prediction and identification performance of complex plants even when the training signal of the model contains up to 50% of outliers. The introduction of the new cost function in the model and PSO based minimization of these cost function has contributed to robust and improved performance compared to those obtained from standard squared error norm based model.



(b)Using Scheme-1 (MSE-norm)learning with 40% outliers

Fig. 6 Comparison of response of the dynamic plant of Example 4

REFERENCES

[1] Chia-Feng Juang, "A TSK-type recurrent fuzzy network for dynamic systems processing by neural network and genetic algorithms", IEEE Trans. on Fuzzy Systems, vol. 10, no. 2, pp. 155-170, April 2002.

[2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Second Edition, Pearson.

[3]D.H.Goldberg, *Genetic algorithms in search, optimization and machine learning*, Addition-Wesley,1989.

[4]G. Panda, D. Mohanty, Babita Majhi and G. Sahoo, "Identification of Nonlinear Systems using Particle Swarm Optimization Technique", in Proc. of IEEE Congress on Evolutionary Computation(CEC-2007), Singapore 25-28, September, 2007, pp.3253-3257

[5] B. Majhi and G. Panda, "Bacterial Foraging based Identification of Nonlinear Dynamics System", in Proc. of IEEE Congress on Evolutionary Computation(CEC-2007**),** Singapore 25-28, September, 2007, pp.1636-1641

[6] Marco Dorigo and Thomas Stützle, Ant colony optimization, MIT Press, June 2004

[7]Joseph W. McKean, "Robust analysis of Linear models", Statistical Science, vol. 19, no. 4, pp. 562-570, 2004.

[8] Jer-Guang Hsieh, Yih-Lon Lin and Jyh-Horng Jeng, "Preliminary study on Wilcoxon learning machines", IEEE Trans. on neural networks, vol. 19, no. 2, pp. 201-211, Feb. 2008.

[9] J. Kennedy and R. C. Eberhart, *Swarm Intelligence*, San Mateo, CA: Mogan Kaufmann, 2001

[10]K. S Narendra. and K. Parthasarathy, "Identification and control of dynamical systems using neural networks", IEEE Trans. on neural networks, vol. 1, no. 1, pp. 4-27, 1990.

[11] J. C. Patra, R. N. Pal, R. Baliarsingh and G. Panda, "Nonlinear channel equalization for QAM signal constellation using artificial neural networks", IEEE Trans. on Systems, Man and Cybernetics-Part B, vol. 29, no.2, pp. 262-271, April 1999.