Evaluation of the ANN Based Nonlinear System Models in the MSE and CRLB Senses

M.V Rajesh, Archana R, A Unnikrishnan, R Gopikakumari, Jeevamma Jacob

Abstract— The System Identification problem looks for a suitably parameterized model, representing a given process. The parameters of the model are adjusted to optimize a performance function based on error between the given process output and identified process output. The linear system identification field is well established with many classical approaches whereas most of those methods cannot be applied for nonlinear systems. The problem becomes tougher if the system is completely unknown with only the output time series is available. It has been reported that the capability of Artificial Neural Network to approximate all linear and nonlinear input-output maps makes it predominantly suitable for the identification of nonlinear systems, where only the output time series is available. [1][2][4][5]. The work reported here is an attempt to implement few of the well known algorithms in the context of modeling of nonlinear systems, and to make a performance comparison to establish the relative merits and demerits.

Keywords—Multilayer neural networks, Radial Basis Functions, Clustering algorithm, Back Propagation training, Extended Kalman filtering, Mean Square Error, Nonlinear Modeling, Cramer Rao Lower Bound.

I. INTRODUCTION

The problem of system modeling and identification has attracted considerable attention during the past few years mostly because of a large number of applications in diverse fields like chemical processes, biomedical systems, transportation, ecology, electric power systems, hydrology, aeronautics and astronautics. An accurate on-line estimate of critical system states and parameters are needed in a variety of engineering applications like in automatic control, signal processing, echo cancellation, fault detection, tracking, image processing, speech recognition, and biomedical instrumentation. Neural network is a well recommends choice, when unknown dynamics is required to be constructively approximated. During the past few years, several authors have suggested a neural network implementation for nonlinear dynamical black box modeling [1][2][4].

When the mathematical model of the process cannot be derived with an analytical method, the only way for modeling is to represent the model function with a known function, which uses the relationship between input and output of the process. In the consequent approach, a neural network that emulates the behavior of the plant is trained based on the known non-linear model and the available time series output [2][4]. Thus dynamical system information is stored in the neural network function.

II. OBJECTIVES AND THE METHODOLOGIES

Objective of this paper is to implement the following Algorithms for nonlinear system identification and compare the performance of the models in order to evaluate the relative merits and demerits of the following algorithms

1. Back Propagation Algorithm (BPA) (Gradient Descent on Multilayer Perceptron (MLP) Networks)
2. Training based on the Extended Kalman Filter (EKF) algorithm. (BPA on MLP)
2. Back Propagation Training on Radial Basis Function (RBF) Network

All the models use the Nonlinear Auto Regressive with Exogenous input (NARX) model [4] for representing the system. Sec. 3.1 and 3.2 discusses the results of modeling using the illustrative examples of three nonlinear sets of data available with us viz.

1. Data generated using $y=sin(t+t^2)$
2. Nonlinear Source-A
3. Nonlinear Source-B.

(Measurements from some nonlinear under water systems)

The same model size and the network structure were used for all the three nonlinear systems, i.e. a Multi Layer Feed Forward Neural network (MLFFN) with one hidden layer, 14 input neurons and 15 hidden neurons [13][15]. The activation function used in the hidden neurons is "bipolar sigmoid" and the output neuron is linear.

M.V Rajesh is with Govt. Model Engineering College, Thrikkakara (PO), Cochin, Kerala, India-682 021,as Assistant Professor in Electronics Engineering; Phone: 91-4842575370-345; fax: 91-484-2577379; e-mail: mvrajesh@mec.ac.in, rajeshmaliyeckal@gmail.com).

Archana R is with the Department of Electrical Engineering, Federal Institute of Technology, Mookkannur, Angamali, Kerala, India, as Assistant Professor in Electrical Engineering (e-mail: archanasreenivasan@rediffmail.com).

Dr. A Unnikrishnan is with the Naval Physical & Oceanographic Laboratory, under Defense Research & Development Organization (DRDO-Govt. of India), Cochin, Kerala, India-682021, as the Associate Director. (e-mail: a_unni_in@yahoo.co.uk).

Dr. R Gopikakumari is with the Cochin University of Science & Technology, CUSAT(PO), Kerala, India-682022, as the Head, Division of Electronics, School of Engineering (SOE), (e-mail: gopika@cusat.ac.in)

Dr. Jeevamma Jacob is with the National Institute of technology (NIT), REC (PO), Calicut, Kerala, India as Assistant Professor in Electrical Engineering, e-mail: jeeva@nitc.ac.in

The amazing challenges in statistical estimation along with an opportunity to learn different techniques in solving the well known problem motivated the authors to compare the performance of these approaches. The model behavior and performance are evaluated in terms of the Mean Square Error (MSE) and the Cramer Rao Lower Bound (CRLB) for efficiency check [5].

III. PERFORMANCE ANALYSIS

A summary of the results and their comparisons obtained with the simulation and analysis are presented next. The final conclusions are made after a thorough evaluation of all the neural network models using the three different nonlinear systems mentioned above.

A. Performance Analysis with the Back Propagation Algorithm

The initial value of the weight, W is chosen randomly. The example cases (Fig. 3.1 and 3.3) clearly indicate that all the models bring down the error in modeling to less than 0.01. The model also reaming stable during validation and it is observed that the stability is better, when more samples are used for training.
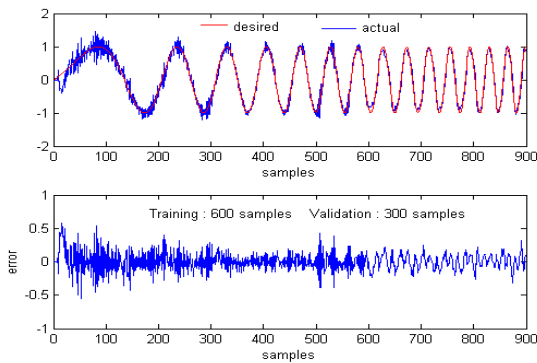

Fig.3.1: Comparison of the time series generated by $y=sin(t+t^2)$ and the modeled output generated by Back Propagation Algorithm along with the instantaneous error
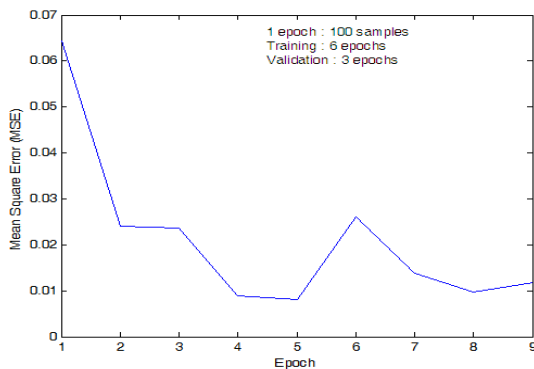

Fig 3.2: The Mean Square Error (MSE) between the given time series and the modeled output (BPA algorithm)

B. Performance Analysis with the Extended Kalman Filter Algorithm

The same nonlinear systems as used in BPA are modeled using EKF. The EKF model takes the parameters (i.e. the weights W of the neural network) as the state and model the system using the following update equations [2][3][8][13].

Let,

$$W = [w_1, w_2, ... w_m]^T \qquad (1)$$

as the weight vector and,

$$y_k = h \left(\Sigma_i w_i y_{k-i} + \Sigma_j w_j v_j \right), \qquad (2)$$

as the modeling function, as given by the NARX model. Then the steps for updation are given by;

1. $W_{k+1/k} = W_{k/k} + \omega_k \qquad (3.1)$

    (Process equation for EKF, ω_k is the process noise)

2. $W_{k/k} = W_{k/k-1} + L_k (y_k - h ()) \qquad (3.2)$

    Where h is the non linearity in the neural network;
    L_k is the Kalman Gain given by,

3. $L_k = P_{k/k} H_k^T (H_k P_{k/k} H_k^T + R)^{-1} \qquad (3.3)$

    Where,

    $H_k = \partial h / \partial W; \qquad (3.4)$

    $P_{k/k} = E (Wk_{/k} W^T_{k/k}); \qquad (3.5)$

    Where R is the noise variance in the model process and $P_{k/k}$ is the state covariance.

4. $P_{k+1/k+1} = [I - W_{k+1})H_{k+1}]P_{k+1/k}, \qquad (3.6)$

The learning process is started with an initial value of $P_{1/1}$ and R and W are updated for every term in the given time series in the order of its arrival. The learning carried out for 1000 number of samples, so that the error is sufficiently lowered and thereafter the process is validated with the remaining samples in the time series. The neural network model size and structure used are same as in the Back Propagation Algorithm i.e. one hidden layer, 14 input neurons and 15 hidden neurons. The activation function used in hidden neurons is "bipolar sigmoid" and output neuron is linear. The results of the simulation with this updation strategy are presented in Fig 3.3 and 3.4 for the nonlinear source-A.
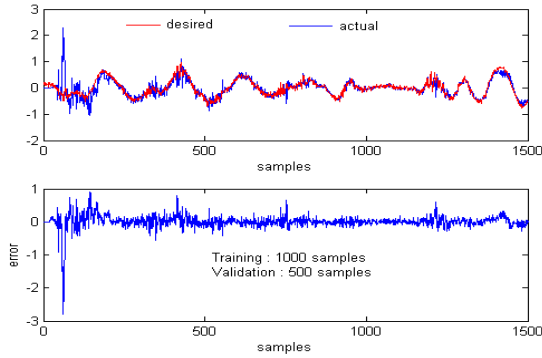
Fig 3.3: Comparison of the given time series (*Nonlinear Source-A)* and the modeled output in the EKF algorithm along with the instantaneous error.
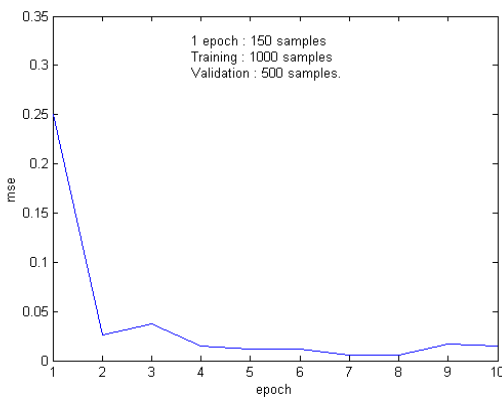

Fig 3.4: The Mean Square Error (MSE) between the given time series and the modeled output. (EKF algorithm)

The performance of the EKF model is superior to that of back propagation algorithm for all the three nonlinear systems. Also it is consistent for the three different time series with different types of nonlinearity behaviors. This indicates that the EKF Algorithm is well suited for nonlinear system identification in general. The dependence of mean square error on initialization of states, process and measurement co variances are also evaluated and the suitable values are found out by running the simulations at different values (trial and error method).

C. Performance Analysis of Radial Basis Function Network

RBF is an alternative of the MLP network for performing nonlinear mapping. As a result the RBF network can immediately be employed to find the system identification operator [2][4]. Analysis is also done for studying the performance of the RBF network. This network consists of three layers (Fig: 3.5). The input layer has neurons with a linear function that simply feed the input signals to the hidden layer. Moreover, the connections between the input and hidden layer are not weighted. The hidden neurons are processing units that perform the radial basis function. The

output neuron is a summing unit to produce the output as a weighted sum of the hidden layer outputs.
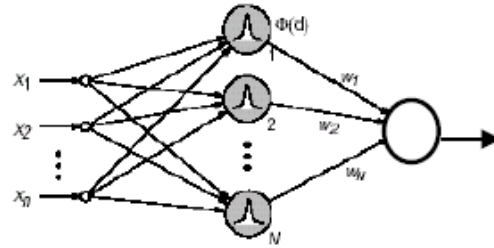

Fig 3.5: The structure of RBF network

In the training of the RBF network, the centre vector is updated using a clustering algorithm [2][4] which is described below.

The training is carried out in a systematic way for the RBF network. The input time series is assumed to be available as blocks of data given by,

$$X = [\ y_1\ ,\ y_2,\ \dots\ y_n]^T \tag{4}$$

The training is carried out with the following steps:
- Initialize the centre vectors C_j s as a random subset of the input vector space X_i
- Every cluster centre C_j is updated, each time an input vector X_i is applied to the network.
- The cluster nearest to X_i has its position updated using,

$$C_j\ (new) = C_j\ (old) + \alpha[X_i - C_j\ (old)] \tag{5}$$

Where α is the learning rate; α is taken as 0.025 in this problem.

Notice that the cluster centre C_j is moved closer to X_i because this above equation minimizes the error vector $(X_i - C_j)$.

After this adaptation, the output vector of the hidden layer is calculated to be,

$$h_j = \exp\left(\frac{\|Xi - Cj\|^2}{2\sigma_j^{\ 2}}\right) \tag{6}$$

for all the hidden layer neurons. A multivariate Gaussian function has been selected as the activation function, as given above.

The output element of a hidden neuron, h_j has a significant value if the Euclidean distance is the minimum and thus at a time only one hidden neuron output is significant.

The output weight matrix W_o is now obtained using the standard back propagation algorithm. The results were more encouraging in this as the weights are also adapted with an intention to reduce overall error. It was possible to minimize the error to a very small value in most of the nonlinearities [4].
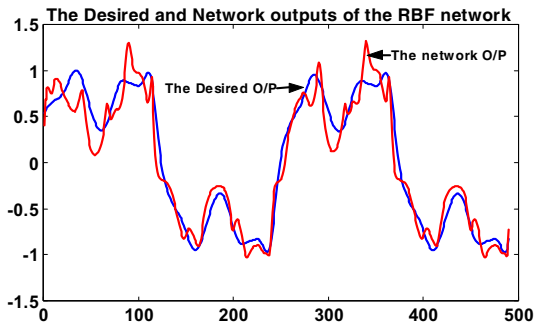
Fig 3.6: Comparison of the given time series and the modeled output in the RBF network.
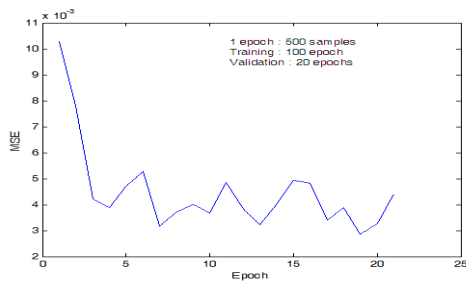

Fig 3.7 The mean square error (MSE) in RBF Network

IV. PERFORMANCE COMPARISON

A NARX model using neural network is selected for the system identification and different algorithms are used to train the model to make the model behaviour equivalent to that of actual system. The performance of the model can be evaluated in terms of the mean square error. The objective of this paper is to compare the performance of different algorithms for non-linear system identification. Here the algorithms Back Propagation (BPA), Extended Kalman Filter (EKF) for both feed neural networks, Expectation Maximization (EM) and Maximum Likelihood Estimation (MLE) are implemented and their performance is compared.

Results show that MSE is less for the EKF model for all the three nonlinear systems and so it performs well compared to BPA. Performance of the algorithms also depends on the type of non-linear system; the kind of nonlinearity involved etc. Three entirely different nonlinear systems are used, and it is seen that the superiority in the performance of EKF is consistent for all of them
.

TABLE I
COMPARISON OF PERFORMANCE AMONG VARIOUS METHODS

Nonlinearity	EKF	RBF	BPA
$y=\sin(x^2+x)$	2.8×10^{-3}	0.0863	0.0980
Nonlinear Source-A	0.0010	0.0065	0.0068
Nonlinear Source-B	0.0086	0.0187	0.0182

C. Comparison based on the MSE

The comparison of the performance of different approaches using the MSE is summarized in Table 4.1 below. From the table, it is seen that EKF algorithm converges faster and has better performance compared to the other Algorithms. It is also consistent for all the nonlinear systems modeled. The other algorithms also give reasonably good results and computationally efficient.

D. Comparison based on CRLB

The efficiency of an estimator can be checked by, establishing the Cramer Rao Lower Bound for the various estimators. According to it, the mean square error corresponding to the estimator of a parameter cannot be smaller than a certain quantity related to the likelihood function. If an estimator's variance is equal to the CRLB, then such estimator is called efficient. [5].

The Cramer Rao Lower Bound (CRLB) on the covariance matrix of the target parameter estimate **X** is (assuming this estimate to be unbiased),

$$E(x-\mathbf{X})(x-\mathbf{X})^T \geq FIM^{-1} \tag{7}$$

Where FIM is the Fisher Information Matrix,.. Following [5] we note that FIM can be written as,

$$-r^{-1}\sum_k \partial h(k,x)/\partial x\ \partial h(k,x)/\partial x\ |_{x=\mathbf{X}}, \tag{8}$$

Where $h(.)$ is the modeling function and r the variance of the measurement z(k) given by,

$$z(k) = h(k, x) \tag{9}$$

This follows from the assumption that the measurement noises are white, zero mean and with variance r. [5].

A necessary condition for an estimator to be consistent in the mean square sense is that there must be an increasing amount of information (in the sense of Fisher) about the parameter in the measurements. The Fisher information has to tend to infinity as $k \to \infty$. Then the CRLB converges to zero as $k \to \infty$ and thus the variance in the estimate can also converge to small values.

CRLB calculations are done for the models and thus checked the efficiency of the models. If the model satisfies CRLB, that is an efficient estimator. Model convergence is checked for 100 different values of initialisations of the state vector (keeping mean and variance same) and based on that CRLB is calculated. CRLB checking involves comparison of two matrices; one is the state covariance matrix and the other inverse of Fischer Information matrix. It is seen that these two matrices are always diagonally dominant. So the checking becomes easy by comparing the diagonal elements of the matrices. That is done as shown below. The comparison is also possible by subtracting one matrix from the other and

checking the positive semi definiteness of the resultant. (If A-B is positive semi definite, if A>B).
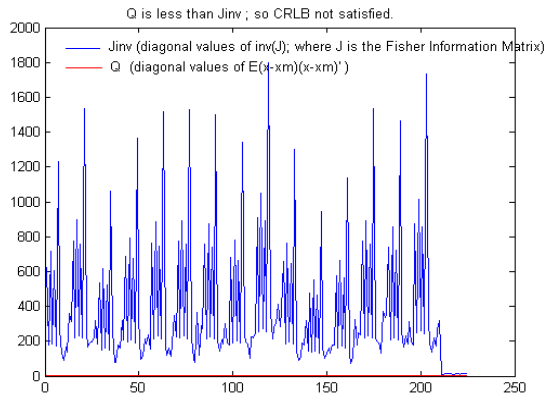

Fig 4.1 CRLB plot for the BPA trained network; here the variance (close to X axis) is much lower than the inverse of the uncertainty matrix.
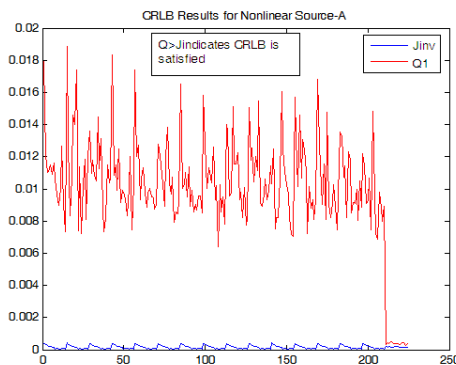

Fig 4.2 CRLB plot for the EKF trained network. Here the variance (above) is much above the inverse of the uncertainty matrix (below and close to X axis)

From the simulations it is seen that the BPA methods applied with MLP and RBF networks do not satisfy CRLB in most of the cases. BPA is a gradient based algorithm, so that chances are likely to settle in local minima. The CRLB is calculated for all the nonlinear systems modeled using BPA and got the similar results. In most of the example cases tried out, the EKF based weight estimate satisfied the CRLB conditions.

V. CONCLUSIONS

In this paper an evaluation of two approaches for non-linear system identification has been carried out. The performance is compared by implementation of the BPA and the EKF based algorithms in a Neural Network NARX model. The adaptive feature revealed by feed forward and recurrent neural network as well as their ability to model non-linear time varying process has resulted in developing efficient models.

It is expected that the results achieved in this paper can help control engineers to choose proper approach for system identification especially non-linear systems. Literature survey shows that many other approaches are also available for system identification. Our further efforts would be to explore such approaches too.

VI. ACKNOWLEDGMENT

REFERENCES

[1] N.K Sinha and B Kuszta, *Modeling and Identification of Dynamic systems*,Van Nostrand Reinhold Company, New York,1983.
[2] Simon Haykin, *Neural Networks a comprehensive Foundation*, Prentice Hall International Editions, 1999.
[3] A.V. Balakrishnan, *Kalman Filtering Theory*, Optimization Software Inc. Publications Division, Newyork,1992
[4] Fa-Long Luo and Rolf Unbehauen, *Applied Neural Networks for Signal Processing*, Cambridge University Press, 1997.
[5] Yaakov Bar-Shalaom and Xiao-Rong Li, *Estimation and Tracking: Principles, Techniques and Software*. Artech House, Boston, London, 1993.
[6] G.V. Puskorius and L.A Feldkamp, " Neuro Control of nonlinear dynamical systems with Kalman Filter trained recurrent networks" IEEE Transactions on neural networks, Vol 5, No.2, pp 279-297, 1994.
[7] K.S. Narendra and K Parthasarathy, "Identification and control of Dynamical systems using neural networks", IEEE Transactions on Neural Networks, Vol 1, No.2, pp 4-27, March 1990
[8] Shuhi Li, "Comparative Analysis of back propagation and Extended Kalman filter in Pattern and Batch forms for training Neural Networks", IEEE Transactions on Neural Network, Vol 2, No.1, pp144-149, 2001
[9] M.S. Grewal and A.P Andrews, *Kalman Filtering Theory and Practice*, Prentice Hall, Englewood Cliffs, 1993.
[10] Y. Linguni, H. Sakai, H. Tokumaru, "*A Real Time Learning Algorithm for Multilayered Neural Network based on the Extended Kalman Filter*", IEEE Transactions on Signal Processing, Vol 40, No.4, pp 959-966,1992.
[11] Joost H. de Vlieger and Robert H.J. Gmelig Meyling , "Maximum Likelihood Estimation for Long Range Target Tracking Using Passive Sonar Measurements", IEEE transactions on Signal Processing, Vol. 40, No.5, pp 1216-1225,May 1992

[12] Ivan Petrovic, Mato Baotic, Nedjeljko Peric "*Model structure selection for nonlinear system identification using feed forward neural networks*" ,Department of Control and Computer Engineering in Automation, Unska 3, HR – 10 000 Zagreb, Croatia.
[13] Simon Haykin, *Adaptive Filter Theory*, Prentice Hall International editions, 1986
[14] Anders Forsgren and Robert Kling "*Implementation of Recurrent Neural Networks for Prediction and control of Nonlinear Dynamic Systems*", Lulea University of Technology, Lulea, Sweden
[15] Ben James, Brian D.O, Anderson, and Robert . C. Williamson "Conditional Mean and Maximum Likelihood approaches to Multi harmonic frequency estimation.", IEEE Transactions on Signal Processing, Vol 42, No.6, pp 1366-1375, June 1994.
[16] Langford B White, "Robust Approximate likelihood ratio Tests for Nonlinear dynamic systems",   IEEE Transactions on Signal Processing, Vol 43, No.8, pp 2028-2031, August 1995.