ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

Nonlinear System Identification Using Maximum Likelihood Estimation

Siny Paul[1], Bindu Elias[2]

Associate Professor, Department of EEE, Mar Athanasius College of Engineering, Kothamangalam, Kerala, India[1,2]

Abstract: Different algorithms can be used to train the Neural Network Model for Nonlinear system identification. Here the 'Maximum Likelihood Estimation' is implemented for modeling nonlinear systems and the performance is evaluated. Maximum likelihood is a well-established procedure for statistical estimation. In this procedure first formulate a log likelihood function and then optimize it with respect to the parameter vector of the probabilistic model under consideration. Four nonlinear systems are used to validate the performance of the model. Results show that Neural Network with the algorithm of Maximum Likelihood Estimation is a good tool for system identification, when the inputs are not well defined.

Keywords: Neural Network, Nonlinear system, Mean square error, Modeling.

I. INTRODUCTION

This paper concentrates on modeling problem which arise when we can identify a certain quantity as a definite measurable output or effect but the causes are not well defined. This is called time series modeling, where inputs or causes are numerous and not quite known in addition to often being unobservable. This type modeling is also called stochastic modeling. In system identification we are concerned with the determination of the system models from records of system operation. The problem can be represented diagrammatically as below.

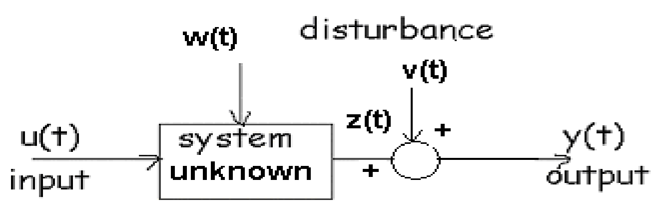

Fig.1 System Configuration [1]

where u(t) is the known input vector of dimension 'm'
z(t) is the output vector of dimension 'p'
w(t) is the input disturbance vector
n(t) is the observation noise vector
v(t) is the measured output vector of dimension 'p'

Thus the problem of system identification is the determination of the system model from records of u(t) and y(t).

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

An artificial neural network is a powerful tool for many complex applications such as function approximation, optimization, nonlinear system identification and pattern recognition. This is because of its attributes like massive parallelism, adaptability, robustness and the inherent capability to handle nonlinear system. It can extract information from heavy noisy corrupted signals.

System identification can be either state space model or input-output model [1].

II. INPUT-OUTPUT MODEL

An I/O model can be expressed as $y(t) = g(\phi(t,\theta)) + e(t)$, where, θ is the vector containing adjustable parameters which in the case of neural network are known as weights, g is the function realized by neural network and ϕ is the regression vector. Depends on the choice of regression vector different model structures emerge.
Using the same regressors as for the linear models, corresponding families of nonlinear models were obtained which are named NARX, NARMAX, etc. Different model structures in each model family can be obtained by making a different assumption about noise.

NNARX $\phi(t,\theta) = [y(t-1), y(t-2),.....y(t-n), u(t-1),.....u(t-m)]^T$ (1)

NNARMAX $\phi(t,\theta) = [y(t-1),.....y(t-n), u(t-1),.....u(t-m), e(t-1),.....e(t-k)]^T$ (2)

Where y(t) is the output, u(t),the input and e(t) is the error. For the implementation of the above system, Feed forward neural network can be used.

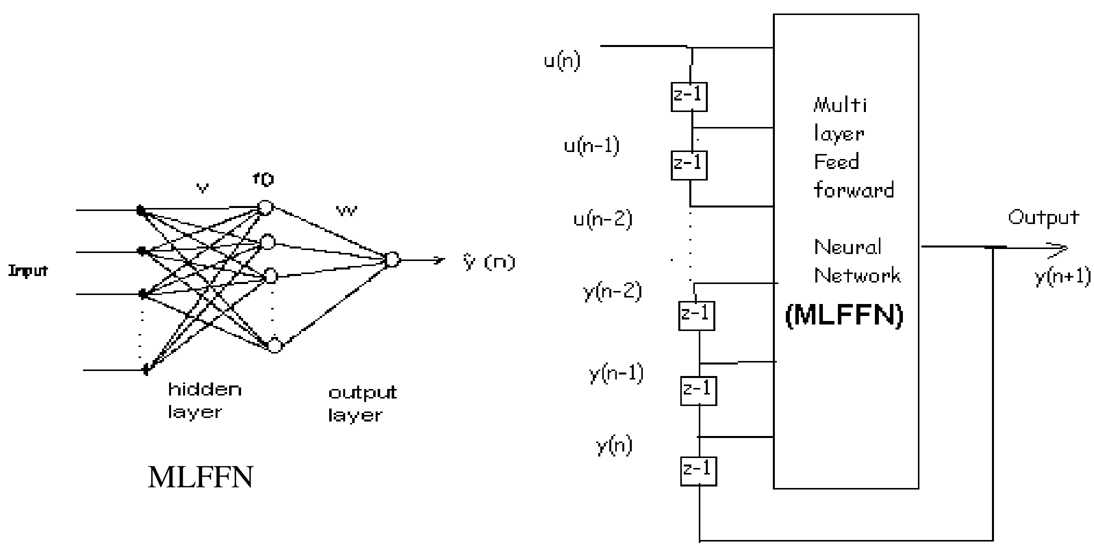

Fig. 2. NARX model [2]

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

NARX Model is well suited for Input-Output modeling of stochastic nonlinear systems [3] So in the proposed work, NARX model is chosen as the system model; in which the model structure is a Multi Layer Feed Forward Neural Network(MLFFN) as shown in Fig. 2 For all the models (using different algorithms).

III. MAXIMUM LIKELIHOOD ESTIMATION.

The term "maximum likelihood estimate" with the desired asymptotic properties usually refers to a root of the likelihood equation that globally maximizes the likelihood function [3]. In other words the ML estimate x_{ML} is that value of the parameter vector x for which the conditional probability density function P(z/x) is maximum [4].The maximum likelihood estimate x_{ML} of the target parameters x is the mode of the conditional probability density function(likelihood function):

$$p(z/x) = \frac{1}{(2\pi)^{N/2}\prod_{k=1}^{N}\sigma_k}\exp\left(-\tfrac{1}{2}\sum_{k=1}^{N} r_k^2\right) \quad (1)$$

The log likelihood function $$\log(p(z/x)) = -\tfrac{1}{2}\sum_{k=1}^{N} r_k^2 \quad (2)$$

Where r_k is the residual . $r_k = \frac{d_k - z_k}{\sigma_k}$ σ_k is the std. deviation, d_k=desired value,

z_k = measurement(estimated value).

Maximizing log likelihood function log(p(z/x)) is equivalent to minimizing the negative log likelihood function L(z,x).

By using the negative log-likelihood function L(z,x) the ML problem is reformulated as a nonlinear least square problem:

Minimize $$L(z_N, x) = \tfrac{1}{2}\sum_{k=1}^{N} r_k^2 \quad (3)$$

The ML estimate must satisfy the following optimality condition:

$$\nabla_x L(z, x_{ML}) = J(x_{ML})^T r(x_{ML}) = 0 \quad (4)$$

Where r(x) is the N dimensional residual vector and J(x) the N x n Jacobian matrix.

$$R(x) = [r_1(x) \ldots\ldots\ldots r_N(x)]^T \quad (5)$$

$$J(x)^T = \nabla_x r(x)^T \quad (6)$$

The operator ∇_x is defined as $\nabla_x = [\partial/\partial x_1 .. \partial/\partial x_2 .. \partial/\partial x_3 \ldots\ldots \partial/\partial x_N]^T$

Since the problem is nonlinear it is not possible to solve analytically. But certain optimization methods seem suitable to find the ML Estimate. Gauss-Newton method is one such optimization technique.

A. *System Modeling Using Gauss-Newton Method.*

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

A Feed forward Neural Network model similar to earlier cases is designed for the identification of the same nonlinear systems and trained using Gauss – Newton method [5].
The Gauss-Newton method is applicable to a cost function that is expressed as the sum of error squares.

$$\text{Let } E(x) = \tfrac{1}{2}\sum_{k=1}^{N} r(k)^2 \qquad (7)$$

The error signal r(k) is a function of adjustable state vector x. Given an operating point x(n), we linearize the dependence of rk on x by writing

$$r^{'}(k,x) = r(k) + [\partial r(k)/\partial x]^{T}{}_{x=x(n)}\left(x - x(n)\right), \quad k=1,2\ldots\ldots n \qquad (8)$$

Equivalently, by using matrix notation we may write

$$r'(k,x) = r(k) + J(n)\left(x - x(n)\right) \qquad (9)$$

The updated state vector x(n+1) is then defined by

$$x(n+1) = \arg.\min\left(\tfrac{1}{2} r^{'}(n,x)^2\right) \qquad (10)$$

squared Euclidean norm of r’(n,x),

$$\tfrac{1}{2} r^{'}(n,x)^2 = \tfrac{1}{2} r(n) + r(n)^T J(n)(x - x(n)) + \tfrac{1}{2}(x - x(n)^T J(n)^T J(n)(x - x(n)) \qquad (11)$$

Hence differentiating this expression with respect to x and setting the result equal to zero, we obtain

$$J(n)^T r(n) + J(n)^T J(n)(x - x(n)) = 0 \qquad (12)$$

Solving this equation for x,

$$x(n+1) = x(n) - (J(n)^T J(n))^{-1} J(n)^T r(n). \qquad (13)$$

which describes the pure form of the Gauss-Newton method.

Unlike Newton’s method that requires knowledge of the Hessian matrix of the cost function E(n), Gauss – Newton method only requires the Jacobian matrix of the error vector r(n). However, for the Gauss- Newton iteration to be computable, the matrix product $J(n)^T J(n)$ must be nonsingular. Unfortunately, there is no guarantee that this condition will always hold. To guard against the possibility that J(n) is rank deficient, the customary practice is to add the diagonal matrix δI to the matrix $J(n)^T J(n)$. The parameter δ is a small positive constant chosen to ensure that;

$J(n)^T J(n) + \delta I$: positive definite for all n.

On this basis, the Gauss-Newton method is implemented in slightly modified form:

$$x(n+1) = x(n) - (J(n)^T J(n) + \delta I)^{-1} J(n)^T r(n) \qquad (14)$$

Here **x** represents the ‘state vector’ of the system. In the Neural Network model **x** becomes the weight vector **w**.

Thus the update equation of the weight vector becomes;

$$w(n+1) = w(n) - (J(n)^T J(n) + \delta I)^{-1} J(n)^T r(n) \qquad (15)$$

where **J(n)** is the Jacobian matrix equal to $\nabla_x r(w)$ (ie derivative of the error w.r.t weights).

IV. RESULTS AND DISCUSSIONS.

The same four nonlinear systems are modeled using feed forward neural network. The NARX model with 14 inputs is used similar to the earlier cases. The performance analysis is done by plotting the mean square error in each case.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

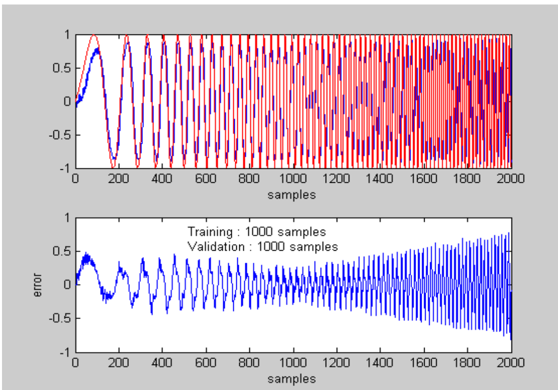

Fig 3. Superposition of model output and desired output Nonlinear System $y = \mathrm{Sin}(x^2 + x)$

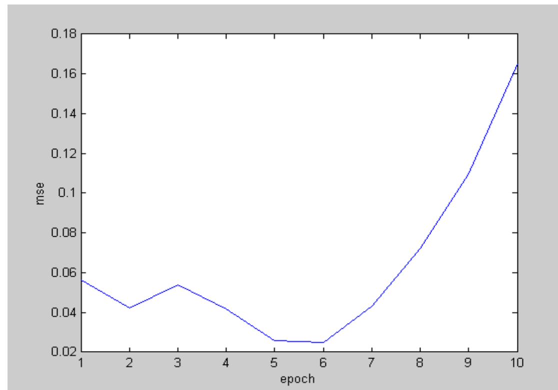

Fig. 4. MSE Vs data samples.

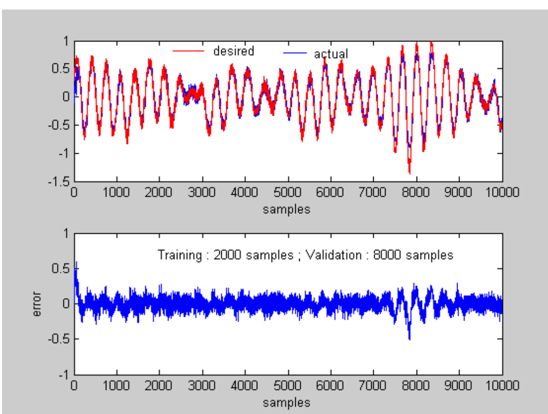

Fig 5. Superposition of model output and desired output of Ambient Noise.

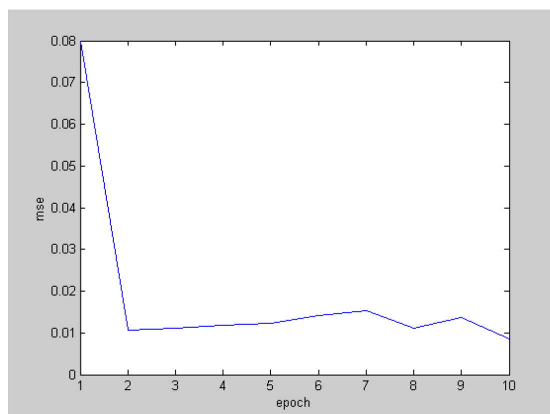

Fig 6. MSE Vs data samples.

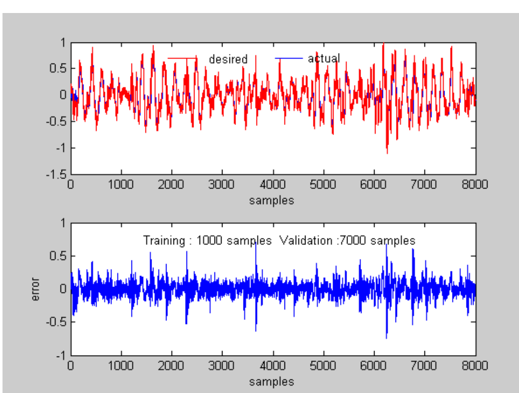

Fig. 7. Superposition of model output and desired output of Accoustic source A.

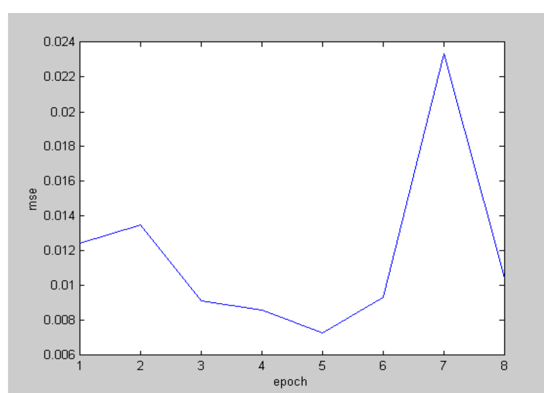

Fig. 8. MSE Vs data samples.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

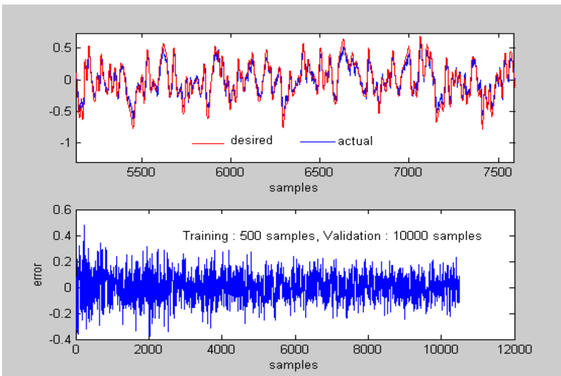

Fig 9. Superposition of model output and desired output of Accoustic source B.

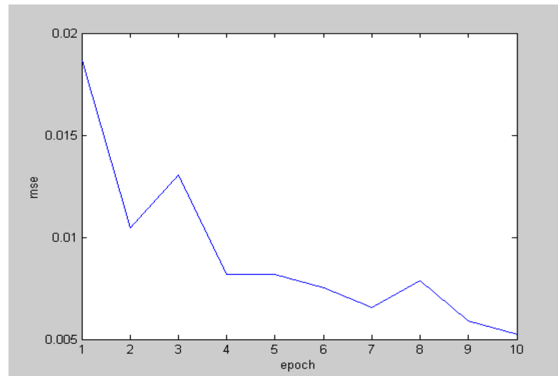

Fig 10. MSE Vs data samples.

TABLE I.
MEAN SQUARE ERROR FOR DIFFERENT SYSTEMS.

System	Mean Square Error
y = sin (x^2+x)	0.0635
Ambient noise	0.0083
Acoustic source ‘A’	0.0118
Acoustic source ‘B’	0.0092

From the above results it is seen that the model is giving good performance for all the four nonlinear systems which proves that NARX Model is well suited for Input-Output modeling of stochastic nonlinear systems

V. CONCLUSION

In this paper, a comprehensive analysis for nonlinear system identification is done and its performance is compared by implementation of the same in a Neural Network NARX model using MATLAB programs. The adaptive feature revealed by feed forward and recurrent neural network as well as their ability to model nonlinear time varying process, provides a surplus value to the model based predictive control. When applied correctly, a neural or adaptive system may considerably outperform other methods. This is an attempt to provide guideline to the practitioners to choose the suitable method for their specific problem in the field of system identification especially in the stochastic modeling of nonlinear systems. It is proved that MLE can be reformulated as a minimization problem; The results show good performance of the models and it is proven that MLE is good for nonlinear system identification. Four different nonlinear systems are used to check the consistency of the performance of algorithm. The performance of MLE is good in terms of mean square error.

REFERENCES

[1] N.K Sinha and B Kuszta, *Modeling and Identification of Dynamic systems*, Vol.1, New York,Van Nostrand Reinhold Company, 1983.
[2] Simon Haykin, *Neural Networks a comprehensive Foundation*, 2nd ed.,India, Pearson Education, 1999.
[3]A.V. Balakrishnan, *Kalman Filtering Theory*, New York, Optimization Software Inc. Publications Division, ,1992.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 1, December 2013

[4] 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.

[5] Roy L Streit and Tod E Luginbubl, "Maximum Likelihood Training of Probabilistic Neural Networks", *IEEE Transactions on Neural Networks*, Vol.5, No.5, pp 764-783, September 1994.

[6] Tsung-Nan Lin, C.Lee Giles and Sun-Yaun Kung, "A Delay Damage Model Selection Algorithm for NARX Neural Networks", *IEEE Transactions on Signal Processing* Vol.45,No.11, pp 2719-2730,November 1997.

[7] Wen Yu and Xiaoou Li"Some New Results on System Identification with Dynamic Neural Network", *IEEE Transactions on Neural Network"*, Vol 12, No.2, pp 412-417, March 2001.

[8] X.M. Ren, A.B. Rad, P.T. Chan and Wai Lan Lo, "Identification and control of continuous time Nonlinear Systems via Dynamic Neural Networks" *IEEE Transactions on Industrial Electronics*", Vol 50,No.3, pp 478 – 486, June 2003.

[9] Songwu Lu and Tamer Basar , " Robust Nonlinear System Identification using Neural Network Models", *IEEE Transactions on Neural Networks*, Vol.9, No.3, pp 407 –429, May 1998.

[10] Mohamed Ibnkabla, "Statistical Analysis of Neural Network Modeling and Identification of Nonlinear Systems with Memory" , *IEEE Transactions on Signal Processing*, Vol. 50, No.6, pp 1508-1517, June 2002.

[11] Nader Sadegh, "A Perceptron Network for Functional Identification and Control of Nonlinear Systems", *IEEE Transactions on Neural Networks*, Vol.4, No.6, pp 982-989, November 1993.

[12]Octavian Stan and Edward Kamen, " A localized Least Squares Algorithm for Training Feed forward Neural Networks", *IEEE Transactions on Neural networks*, Vol.11, No.2, pp 487-495, march 2000.

[13] Si Zhao Qin, Hong Te Su and Thomas J McAvoy, "Comparison of four Neural Network Learning Methods for Dynamic System Identification", *IEEE Transactions on Neural Networks* , Vol. 3, No.1, pp 122-130, January 1992.

[14] 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,March 1994.

[15] 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

[16] 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.4, pp144-149, June 2001.