Monograph:
Billings, S.A. (1983) Parameter Estimation, Structure Detection and Model Validity Tests for Nonlinear Systems. Research Report. ACSE Research Report 243 . Department of Control Engineering, University of Sheffield

PARAMETER ESTIMATION, STRUCTURE DETECTION AND MODEL VALIDITY TESTS FOR NONLINEAR SYSTEMS

by

S. A. Billings, BEng,PhD,CEng,MIEE,AFIMA,MInstMC

Department of Control Engineering
University of Sheffield
Mappin Street, Sheffield S1 3JD

Research Report No. 243

November 1983

To be presented at the IEE Colloquium on "Recent Advances in Identification and Signal Processing Part I", Savoy Place, February 15th, 1984.

PARAMETER ESTIMATION, STRUCTURE DETECTION AND MODEL VALIDITY TESTS FOR NONLINEAR SYSTEMS

S. A. Billings

Before nonlinear identification routines are applied the experimenter should attempt to determine if the process under test exhibits nonlinear characteristics which warrant a nonlinear model[1]. Whenever the input $u(t)+b$, $\overline{u(t)} = 0$, $b \neq 0$ is applied to a system, the system will be linear iff $\overline{y_b(t)} = \overline{y(t)}$ where $\overline{y_b(t)}$ and $\overline{y(t)}$ are the mean levels of the system output for the inputs b (i.e. $u(t) = 0$) and $u(t)+b$ respectively. Alternatively, if the third order moments of the input are zero and all even order moments exist (i.e., sine wave, gaussian, ternary sequence etc) then the process is linear iff[1]

$$\phi_{y'y'^2}(\sigma) = E\left[y'(t+\sigma)(y'(t))^2\right] = 0 \;\forall\sigma \tag{1}$$

The test will distinguish between additive noise corruption of the measurements and distortion due to nonlinear effects providing the input and noise are independent.

If the system is linear then it is finitely realizable and can be represented by the linear difference eqn model

$$y(t) = \sum_{i=1}^{n_y} (a_i y(t-i)) + \sum_{i=1}^{n_u} (b_i u(t-i)) \tag{2}$$

if the Hankel matrix of the system has finite rank. When the system is nonlinear a similar representation can be derived by considering the observability of nonlinear systems and utilizing results from automata theory to yield the nonlinear difference eqn model[2,3]

$$y(t) = F^*\left[y(t-1),\ldots y(t-n_y),u(t-1),\ldots u(t-n_u)\right] \tag{3}$$

where $F^*[\cdot]$ is some nonlinear function of $u(\cdot)$ and $y(\cdot)$. The extension to multivariable systems and conditions for the existence of such a model are rigorously defined elsewhere[3,4]. The Hammerstein, Wiener, bilinear, Volterra and other well known nonlinear models can be shown to be special cases of eqn (3).

A similar representation for nonlinear stochastic systems can be derived by considering input-output maps based on conditional probability density functions to yield the model[3,4]

Dr. Billings is with the Department of Control Engineering, University of Sheffield.

$$z(t) = F[z(t-1),\ldots z(t-n_z),u(t-1),\ldots u(t-n_u),$$
$$\varepsilon(t-1),\ldots\varepsilon(t-n_\varepsilon)] + \varepsilon(t) \qquad (4)$$

where $\varepsilon(t)$ is the prediction error. This model will be referred to as a Nonlinear AutoRegressive Moving Average model with eXogenous inputs or NARMAX model[3,4].

A NARMAX model with first order dynamics expanded as a second order polynomial nonlinearity would for example be represented as

$$y(t) = F_2[y(t-1),u(t-1)]$$
$$= C_1y(t-1) + C_2u(t-1) + C_{11}y^2(t-1) + C_{12}y(t-1)u(t-1)$$
$$+ C_{22}u^2(t-1) \qquad (5)$$

Assuming that the output measurements are corrupted by additive noise

$$z(t) = y(t) + e(t) \qquad (6)$$

gives the input-output model

$$z(t) = C_1z(t-1)+C_2u(t-1)+C_{11}z^2(t-1)+C_{12}z(t-1)u(t-1)$$
$$+C_{22}u^2(t-1)+e(t)-C_1e(t-1)-2C_{11}z(t-1)e(t-1)$$
$$+C_{11}e^2(t-1)-C_{12}e(t-1)u(t-1) \qquad (7)$$

Because the NARMAX model maps the past input and output into the present output multiplicative noise terms are induced in the model even though the noise was additive at the output. In general the noise may enter the system internally and because the system is nonlinear it will not always be possible to translate this to be additive at the output. This situation will again result in multiplicative noise terms in the NARMAX model with the added complication that the noise source and prediction error will not in general be equal. Since most of the parameter estimation techniques derived for linear systems assume that the noise is independent of the input, biased estimates result when they are applied to nonlinear systems eqn (4).

The recursive extended least squares (RELS) algorithm can however be readily adapted to the NARMAX model, by defining the following vectors

$$Q(t) = [z(t-1),u(t-1),z^2(t-1),z(t-1)u(t-1),u^2(t-1),\varepsilon(t-1),$$
$$\varepsilon(t-1)\ z(t-1),u(t-1)\varepsilon(t-1),\varepsilon^2(t-1)]^T$$
$$\hat{\theta} = [\hat{C}_1,\hat{C}_2\ldots\hat{C}_s]^T$$
$$\varepsilon(t+1) = z(t+1) - Q(t+1)^T\hat{\theta}(t) \qquad (8)$$

for the model of eqn (7) for example. With these definitions the standard RELS algorithm[5] can be applied to yield unbiased parameter estimates. The development of a recursive maximum likelihood algorithm (RML) is more involved and requires a complete derivation by working backwards from known conditions of convergence[5]. The major disadvantage of both these algorithms when applied to nonlinear systems is the need to include prediction error terms in the estimation vector. It can be shown that instrumental variables (RIV) will yield unbiased estimates providing the noise terms in the NARMAX model can be represented as a purely linear map[6]. This restriction can be widened slightly by employing a new suboptimal least squares (SOLS) routine[6] based on the model

$$z(t) = F'[\hat{y}(t-1),\ldots\hat{y}(t-n_y),u(t-1),\ldots u(t-n_u)]+\varepsilon(t) \qquad (9)$$

where $\hat{y}(t-1)$ represents the predicted output. The algorithm will yield unbiased estimates whenever the noise is additive at the output.

The direct application of a maximum likelihood algorithm is not possible because in general the prediction errors will not have a Gaussian distribution. However, by considering the loss function

$$J(\theta) = \frac{1}{2N} \log_e \det \sum_{t=1}^{N} \varepsilon(t;\theta)\varepsilon(t;\theta)^T \qquad (10)$$

it can be shown that the prediction error estimates obtained by minimising eqn (10) have very similar asymptotic properties to the maximum likelihood estimates even when $\varepsilon(t)$ is non-gaussian[5]. A prediction error algorithm has been developed for the NARMAX model based on this result. This together with the RELS, RML, RIV and SOLS routines have been augmented with a stepwise regression algorithm, a likelihood ratio test and Akaike tests to detect the correct model structure prior to final estimation.

Whichever model formulation or identification algorithm is mplemented it is important to test that the identified model does adequately describe the data set[1]. When the system is nonlinear the residuals $\xi(k)$ should be unpredictable from all linear and nonlinear combinations of past inputs and outputs and this condition will hold iff

$$\begin{aligned} \phi_{\xi\xi}(\tau) &= \delta(\tau) \\ \phi_{u\xi}(\tau) &= 0 \; \forall \tau \\ \phi_{\xi\xi u}(\tau) &= E[z(t)z(t-1-\tau)u(t-1-\tau)] = 0 \; \forall \tau \geq 0 \end{aligned} \qquad (11)$$

Notice that for nonlinear systems the traditional linear tests $\phi_{\xi\xi}(\cdot)$ and $\phi_{u\xi}(\cdot)$ are not sufficient.

If RIV and SOLS are used the residuals may be coloured and specific tests which determine if the process model is correct without testing the whiteness of the residuals are required. It can be shown that models estimated using RIV or SOLS will be unbiased iff[6]

$$\begin{aligned} \phi_{u\xi}(\tau) &= 0 \, \forall \tau \\ \phi_{u^2\xi^2}(\tau) &= 0 \, \forall \tau \end{aligned} \tag{12}$$

Another problem in the identification of nonlinear systems is the design of input signals. Whenever possible the input should be selected to excite all the modes and amplitudes of interest within the system. Pseudo-random sequences are not in general appropriate for nonlinear systems since they exhibit discontinuous probability density functions and will not yield a persistently exciting input over the full amplitude range of any input non-linearities. The design of inputs for nonlinear system identification is very complex but general rules can be derived from information theoretic arguments. These indicate that for a power or amplitude constraint on the input, the input should be an independent sequence. In addition the input should have a gaussian distribution for a power constraint, and a uniform distribution for an amplitude constraint.

Much work remains to be done to simplify the identification of nonlinear systems and develop algorithms which can be easily applied to practical situations. The ideas and algorithms presented above represent just some of the possible approaches which may lead to a solution to these problems.

Acknowledgements

The author gratefully acknowledges that this work is supported by SERC under grant GR/B/31163.

References

1. BILLINGS, S.A., VOON, W.S.F.: "Structure detection and model validity tests in the identification of nonlinear systems", Proc.IEE, Part D, 130, 1983, pp 193-199.

2. BILLINGS, S.A.: "Identification of nonlinear systems - a survey", Proc.IEE, Part D, 1980, 127, pp 272-285.

3. BILLINGS, S.A., LEONTARITIS, I.J.: "Parameter estimation techniques for nonlinear systems", 6th IFAC Symp. Ident. and Syst. Par. Est., Washington, June 1982, pp 427-432.

4. LEONTARITIS, I.J., BILLINGS, S.A.: "Input-output models for nonlinear systems", Res. Report, Univ. of Sheffield (in preparation).

5. GOODWIN, G.C., PAYNE, R.L.: "Dynamic System Identification: Experiment Design and Data Analysis", Academic Press, N.Y., 1977.

6. BILLINGS, S.A., VOON, W.S.F.: "Least squares parameter estimation algorithms for nonlinear systems", Int. J. Systems Science (to appear).