Modeling and identification of nonlinear dynamic systems

Chen, Sheng

Sheng Chen, "Modeling and identification of nonlinear dynamic systems," Proc. SPIE 2296, Advanced Signal Processing: Algorithms, Architectures, and Implementations V, (28 October 1994); doi: 10.1117/12.190840

SPIE.

Event: SPIE's 1994 International Symposium on Optics, Imaging, and Instrumentation, 1994, San Diego, CA, United States

Modelling and Identification of Nonlinear Dynamic Systems

S. Chen

Department of Electrical and Electronic Engineering
University of Portsmouth, Anglesea Building
Portsmouth PO1 3DJ, U.K.
Email: schen@ee.port.ac.uk Tel.: +44 (0)705–842599

Abstract

The paper summarizes some results of nonlinear system modelling and identification. Connections with the dynamical systems theory and neural networks are emphasized. Two general modelling approaches are highlighted. Issues of identifiability and model validation are also discussed.

Keywords: nonlinear systems, modelling, identification.

1 Nonlinear system representation

Modelling and identification is one of the major areas of control engineering. Theory and practice of linear system identification are well established [1,2]. During the past decade, efforts have been focused on developing coherent and concise methods of nonlinear system modelling and identification.

For the class of discrete-time nonlinear dynamic systems shown in Fig.1, the general input-output relationship can be written as

$$y(k) = f_s(y(k-1), \cdots, y(k-n_y), u(k-1), \cdots, u(k-n_u), e(k-1), \cdots, e(k-n_e)) + e(k) \quad (1)$$

0-8194-1620-7/94/$6.00

where $f_s(\cdot)$ is some nonlinear function. This nonlinear system representation is known as the nonlinear autoregressive moving average with exogenous inputs (NARMAX) model[3,4]. If the system noise is additive, the NARMAX model may reduce to

$$y(k) = f_s(y(k-1), \cdots, y(k-n_y), u(k-1), \cdots, u(k-n_u)) + e(k) \tag{2}$$

For autonomous systems or time series, the model (2) is further reduced to

$$y(k) = f_s(y(k-1), \cdots, y(k-n_y)) + e(k) \tag{3}$$

A main assumption for the NARMAX model is that the "system state space" has a finite dimension. This agrees with a basic result of the dynamical systems theory, which states that if the attractor of a dynamical system is contained within a finite-dimensional manifold, then an embedding of the manifold can be constructed from time series of observations of the dynamics on the manifold. The dynamical system induced by the embedding is differentiably equivalent to the one being observed. In fact, the lag n_y in (3) corresponds to the embedding vector dimension.

The functional form $f_s(\cdot)$ for a real-world system is generally very complex and unknown. Any practical modelling must be based on a chosen model set of known functions. Obviously, this model set should be capable of approximating the underlying process to within an acceptable accuracy. Secondly, an efficient identification procedure must be developed for the selection of a parsimonious model structure because the dimension of a nonlinear model can easily become extremely large. Without efficient subset selection, the resulting model often has little practical value.

The relationship between model outputs and inputs must be nonlinear in a nonlinear model by definition. The relationship between model outputs and adjustable model parameters, however, can be either nonlinear or linear. From this viewpoint, various identification schemes can be classified into two categories, namely the nonlinear-in-the-parameters approach and the linear-in-the-parameters approach.

2 The nonlinear-in-the-parameters approach

Let Θ be the parameter vector of a nonlinear model. In general, the model output is nonlinear with respect to this adjustable parameter vector, and can be written as

$$\hat{y}(k) = f_m(\mathbf{x}(k); \Theta) \tag{4}$$

where $f_m(\cdot)$ is the nonlinear map realized by the model. The modelling error is defined by

$$\epsilon(k) = \epsilon(k; \Theta) = y(k) - \hat{y}(k) \tag{5}$$

The model input vector $\mathbf{x}(k)$ is given by

$$\mathbf{x}(k) = [y(k-1) \cdots y(k-n_y)\, u(k-1) \cdots u(k-n_u)\, \epsilon(k-1) \cdots \epsilon(k-n_e)]^T \tag{6}$$

or

$$\mathbf{x}(k) = [y(k-1) \cdots y(k-n_y)\, u(k-1) \cdots u(k-n_u)]^T \tag{7}$$

or

$$\mathbf{x}(k) = [y(k-1) \cdots y(k-n_y)]^T \tag{8}$$

depending on the system representation used. Examples of this class of nonlinear models are the nonlinear rational model [5] and the multilayer perceptron neural network model [6].

A class of parameter estimation algorithms widely used for nonlinear models is the prediction error algorithm [5,7]. This is a class of gradient based algorithms that minimize the performance function

$$J(\Theta) = \frac{1}{N} \sum_{k=1}^{N} \epsilon^2(k;\Theta) \tag{9}$$

The Hessian of $J(\Theta)$, $H(\Theta)$, also plays an important role in subset model selection [8]. A practical method for selecting parsimonious models is based on backward elimination, which eliminates some parameters of a large model according to information provided by $H(\Theta)$. This method is similar to the stepwise backward elimination scheme in statistical literature [9].

The importance of the parsimonious principle is widely recognized in the neural network community. Weight elimination has been suggested to reduce size of large network models, and the process is known as the pruning. Approaches adopted in pruning often have their root in more traditional methods for subset model selection. For example, the so-called optimal brain damage method [10] uses the diagonal elements of the Hessian $H(\Theta)$ in weight elimination.

3 The linear-in-the-parameters approach

An alternative approach to nonlinear modelling is to perform some fixed nonlinear functional transform or expansion of the inputs and to combine the resulting terms linearly. Specifically, a given functional expansion maps the input space onto a new space of increased dimension n,

$$\mathbf{x}(k) \longrightarrow [\phi_1(\mathbf{x}(k)) \cdots \phi_n(\mathbf{x}(k))]^T \tag{10}$$

The model output is obtained as a linear combination of the new bases $\phi_i(\mathbf{x}(k))$, $1 \leq i \leq n$,

$$\hat{y}(k) = \sum_{i=1}^{n} \theta_i \phi_i(\mathbf{x}(k)) \tag{11}$$

Generally, the value of a given basis function depends only on the input $\mathbf{x}(k)$, and $\phi_i(\mathbf{x}(k))$ contains no other adjustable parameters. An advantage of the model (11) is that the standard

least squares method can readily be applied to estimate the parameters θ_i. Examples of this class of nonlinear models include the Volterra series model [11], the fuzzy basis function network [12], and the general functional-link network [13]. When its hidden layer is fixed, a radial basis function (RBF) network [14] also has this linear-in-the-parameters structure.

In practice, the model dimension n can become excessively large. Consider, for example, the simplest and most well-known way of obtaining a linear-in-the-parameters model, namely Volterra expansion, which derives the set of model bases as the set of monomials of $\mathbf{x}(k)$. If the dimension of $\mathbf{x}(k)$ is 8, a degree-5 Volterra expansion will produce a model basis set of dimension $n = 1286$. If other numerous choices of model bases [15] are also considered, the problem of excessive model dimension can become even serious. Subset selection is therefore essential, and a very efficient subset selection procedure has been derived based on the orthogonal least squares (OLS) method [11,16]. Given the full set of n candidate bases, the algorithm selects significant model bases one by one in a forward regression manner until an adequate subset model is constructed. The selection procedure is made simple and efficient by exploiting an orthogonal property.

A technique for overcoming the overfitting problem in constructing large full-size neural network models is regularisation [17,18]. The regularisation method improves generalisation by adding a penalty function to the criterion $J(\Theta)$

$$J_R(\Theta, \lambda) = J(\Theta) + \lambda(\text{penalty function}) \tag{12}$$

where λ is a regularisation parameter. The simplest penalty function, known as the zero-order regularisation, is $\Theta^T\Theta$. The least squares criterion $J(\Theta)$ in certain circumstances is prone to overfitting. When the data are highly noisy and the model size is large, the problem can be serious. Instead of just relying on the regularisation mechanism, a better approach is to combine regularisation techniques with the parsimonious principle.

A subset selection procedure has been derived by incorporating the OLS algorithm with the zero-order regularisation technique [19]. This regularised OLS (ROLS) algorithm is capable of constructing parsimonious models which generalise well. Furthermore, it has the same computational requirement to that of the OLS algorithm and is, therefore, computationally very efficient. A simple example is used to demonstrate the advantage of the ROLS algorithm, where a RBF network with Gaussian basis function is used to approximate the scalar function $f(x) = \sin(2\pi x)$, $0 \leq x \leq 1$, buried under severe noise. Two network models, each having 15 centres, are constructed by the OLS and ROLS algorithms respectively. The network maps obtained by the two algorithms are plotted in Figs. 2 and 3 respectively.

4 Identifiability

Experiment design for linear system identification is well established [1]. Basically, the input signal chosen for an identification experiment should be persistently exciting, which means

that the input should excite all the frequencies of interests in the system. For nonlinear system identification, however, persistent excitation requires an additional condition that the input should also excite the system over the whole amplitude range of operation. The design of inputs for nonlinear system identification is a very complex problem and some useful results have been given in [20].

A simple example is given to illustrate the relationship between persistent excitation and identifiability. A nonlinear digital communication channel can be represented by

$$y(k) = f_s(u(k), \cdots, u(k - n_u)) + e(k) \tag{13}$$

where the input $u(k)$ is a white sequence taking values from the set $\{\pm 1\}$. Since $u(k)$ is white, it contains all frequency components, and is an ideal input signal for identifying a linear model. For nonlinear identification, the input should cover a sufficient range of amplitudes. The binary nature of $u(k)$ therefore represents a worst scenario and, as a consequence, parameters in some nonlinear model may not be identifiable. For example, consider the following channel model

$$\hat{y}(k) = \sum_{i=0}^{2} h_i u(k-i) + \sum_{i=0}^{2}\sum_{j=i}^{2} h_{ij} u(k-i)u(k-j) + \sum_{i=0}^{2}\sum_{j=i}^{2}\sum_{l=j}^{2} h_{ijl} u(k-i)u(k-j)u(k-l) \tag{14}$$

The rank of the 19×19 autocorrelation matrix of the estimator input vector is only 8. It is therefore impossible to identify all the 19 parameters in (14).

The requirement of covering a sufficient range of amplitudes is difficult to meet in practice. Normal operation of an industrial plant is often concerned with controlling the plant close to some operating points. Perturbing signals that the experimenter injects into the plant can only have a small amplitude in order not to cause large disturbances to the operation of the plant. If normal operation of the plant includes several operating levels, several sets of small perturbing data records can be obtained without violating the amplitude constraints for normal operation. These data records together may cover a sufficient range of amplitudes. A global-model fitting procedure [15] can then be applied to obtain a nonlinear model. Experiment design for nonlinear system identification is an area that more research efforts are required.

5 Model validation

For linear system identification, if the model structure and parameter values are correct, $\epsilon(k)$ will be uncorrelated with past inputs and outputs. Therefore, an identified linear model is regarded as adequate if the autocorrelation function of $\epsilon(k)$ and the cross-correlation function of $\epsilon(k)$ and $u(k)$ satisfy

$$\left.\begin{array}{l} \mathcal{R}_{\epsilon\epsilon}(\tau) = 0, \quad \tau \neq 0 \\ \mathcal{R}_{\epsilon u}(\tau) = 0, \quad \text{for all } \tau \end{array}\right\} \tag{15}$$

For validating a nonlinear model, (15) is clearly insufficient. Three additional tests were suggested [21,5]

$$\left.\begin{array}{ll} \mathcal{R}_{\epsilon(\epsilon u)}(\tau) = 0, & \tau \geq 0 \\ \mathcal{R}_{u^{2'}\epsilon}(\tau) = 0, & \text{for all } \tau \\ \mathcal{R}_{u^{2'}\epsilon^2}(\tau) = 0, & \text{for all } \tau \end{array}\right\} \tag{16}$$

where $\epsilon u(k) = \epsilon(k+1)u(k+1)$, $u^{2'}(k) = u^2(k) - \overline{u^2(k)}$ and the bar indicates the time average.

The tests (16) are higher-order statistics. If a nonlinear model is correct, $\epsilon(k)$ will be uncorrelated with all the linear and nonlinear combinations of past inputs and outputs. Further research is required to develop coherent and easy-to-use nonlinear model validation methods. Recent advance in higher-order statistics methods may provide some useful results.

6 Conclusions

Theory and practice of nonlinear system identification has advanced considerably during the past decade. Future research will involve multi-discipline approaches, including traditional control engineering, nonlinear dynamical systems theory, neural networks and higher-order statistics.

7 References

1. G.C. Goodwin and R.L. Payne, *Dynamic System Identification: Experiment Design and Data Analysis*. New York: Academic Press, 1977.

2. L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA.: MIT Press, 1983.

3. I.J. Leontaritis and S.A. Billings, "Input-output parametric models for non-linear systems – Part I: deterministic non-linear systems; Part II: stochastic non-linear systems," *Int. J. Control*, Vol.41, No.2, pp.303–344, 1985.

4. S. Chen and S.A. Billings, "Representation of non-linear systems: the NARMAX model," *Int. J. Control*, Vol.49, No.3, pp.1013–1032, 1989.

5. S.A. Billings and S. Chen, "Identification of non-linear rational systems using a prediction-error estimation algorithm," *Int. J. Systems Sci.*, Vol.20, No.3, pp.467–494, 1989.

6. S. Chen, S.A. Billings and P.M. Grant, "Non-linear systems identification using neural networks," *Int. J. Control*, Vol.51, No.6, pp.1191–1214, 1990.

7. S. Chen, C.F.N. Cowan, S.A. Billings and P.M. Grant, "Parallel recursive prediction error algorithm for training layered neural networks," *Int. J. Control*, Vol.51, No.6, pp.1215–1228, 1990.

8. I.J. Leontaritis and S.A. Billings, "Model selection and validation methods for non-linear systems," *Int. J. Control*, Vol.45, No.1, pp.311–341, 1987.

9. N.R. Draper and H. Smith, *Applied Regression Analysis*. New York: Wiley, 1981.

10. Y.le Cun, J.S. Denker and S.A. Solla, "Optimal brain damage," in D. Touretzky, ed., *Advances in Neural Information Processing Systems*, Vol.2, pp.598–605, 1990.

11. S. Chen, S.A. Billings and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, Vol.50, No.5, pp.1873–1896, 1989.

12. L.X. Wang and J.M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning," *IEEE Trans. Neural Networks*, Vol.3, No.5, pp.807–814, 1992.

13. Y.H. Pao, *Adaptive Pattern Recognition and Neural Networks*. Reading, MA.: Addition-Wesley, 1989.

14. D.S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Systems*, Vol.2, pp.321–355, 1988.

15. S.A. Billings and S. Chen, "Extended model set, global data and threshold model identification of severely non-linear systems," *Int. J. Control*, Vol.50, No.5, pp.1897–1923, 1989.

16. S. Chen, C.F.N. Cowan and P.M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Networks*, Vol.2, No.2, pp.302–309, 1991.

17. T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, Vol.78, No.9, pp.1481–1497, 1990.

18. C. Bishop, "Improving the generalization properties of radial basis function neural networks," *Neural Computation*, Vol.3, No.4, pp.579–588, 1991.

19. S. Chen, E.S. Chng and K. Alkadhimi, "Regularised orthogonal least squares algorithm for constructing radial basis function networks," submitted to *IEEE Trans. Neural Networks*, 1994.

20. I.J. Leontaritis and S.A. Billings, "Experiment design and identifiability for non-linear systems," *Int. J. Systems Sci.*, Vol.18, No.1, pp.189–202, 1987.

21. S.A. Billings and W.S.F. Voon, "Correlation based model validity tests for non-linear models," *Int. J. Control*, Vol.44, pp.235–244, 1986.

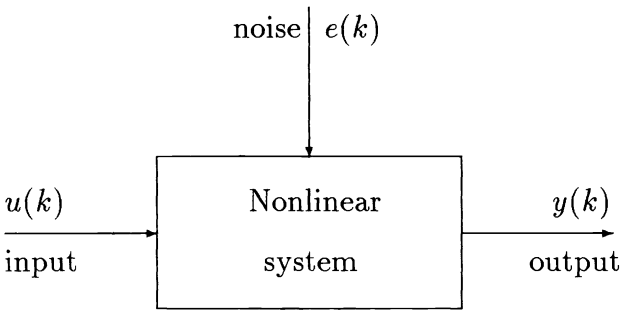

Figure 1: Block diagram of nonlinear system.

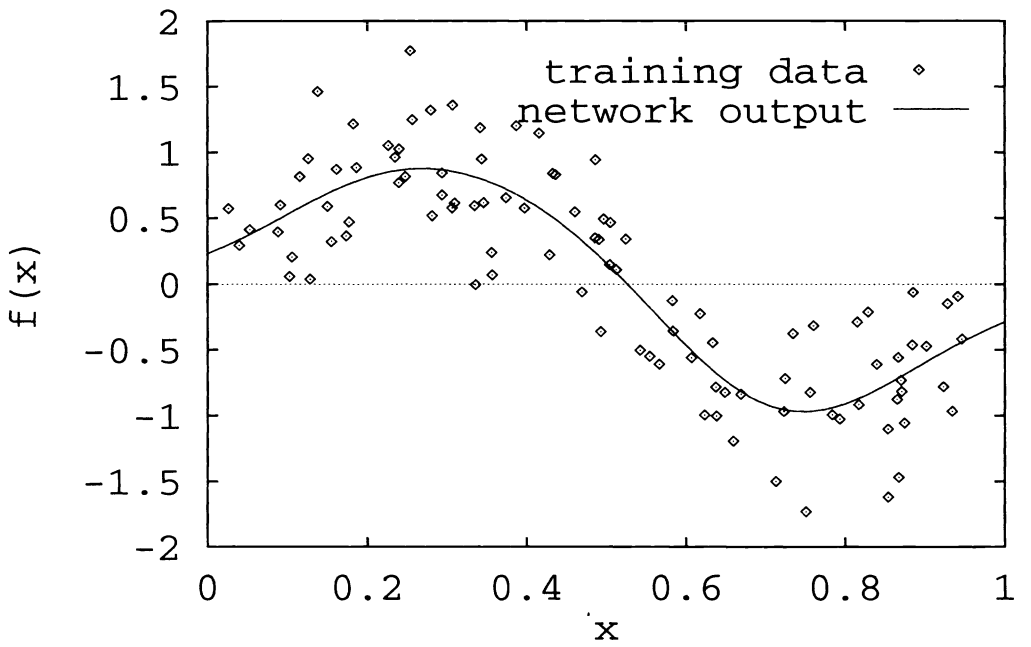

Figure 2: Network mapping constructed by the regularised orthogonal least squares algorithm.

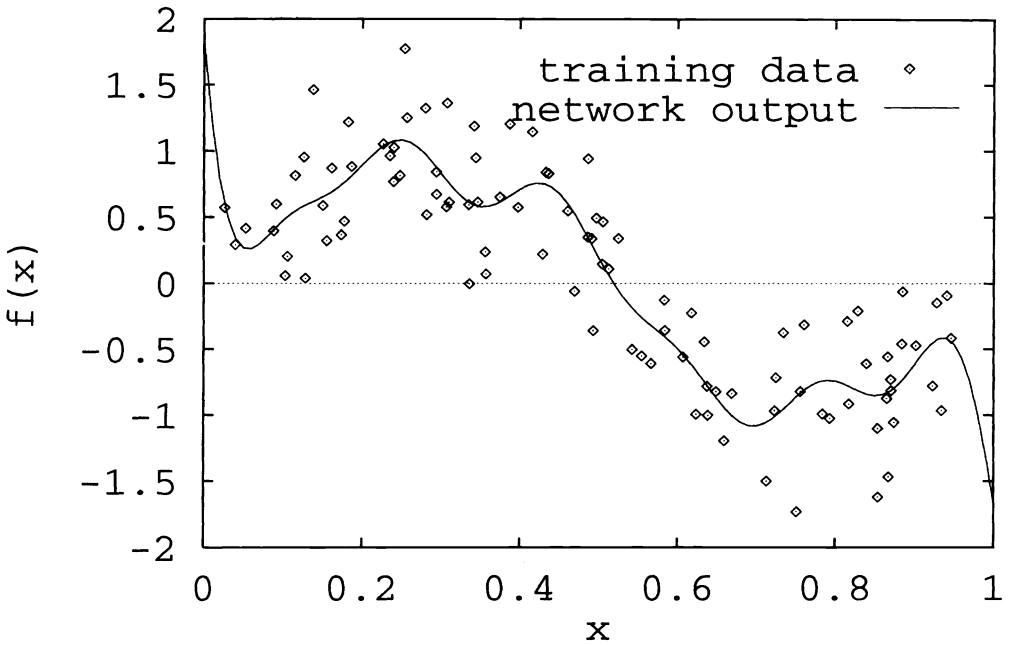

Figure 3: Network mapping constructed by the orthogonal least squares algorithm.