IDENTIFICATION OF NARMAX AND RELATED MODELS

Stephen A. Billings and Daniel Coca
Department of Automatic Control and Systems Engineering, University of Sheffield, UK

Keywords: Activation function, Artificial neural network, Correlation tests, Differential equation, Difference equation, Discrete-time system, Dynamical system, Error reduction ratio, Estimation data set, Frequency response function, Least-squares, Model set, Model predicted output, Model Validation, Multilayer perceptron, Parameter estimation, Prediction error, Radial basis functions (RBF), RBF neural networks, Structure selection, System identification, Wavelet.

Contents

UNESCO – EOLSS SAMPLE CHAPTERS

Summary

Mathematical models are an essential part of most branches of science and engineering. A mathematical model of a system can for example be used to unveil fundamental properties of the system which are not apparent otherwise, leading to a better understanding of that system, or in the design of an automatic control system that can be used to regulate the behavior of certain system variables. A model of a system can be obtained directly from experimental input/output data by determining the model form and the numerical values of the unknown parameters. This process is known as system

identification. This chapter provides an overview of a nonlinear system identification methodology based on the NARMAX model. The NARMAX model is a general representation of a nonlinear dynamical system which takes the form of a nonlinear difference equation. The NARMAX methodology provides a unified solution to the problem of finding this equation based only on experimental data recorded from the system of interest. The process of identifying a NARMAX model involves determining the form or structure of the unknown nonlinear equation, estimating the parameters associated with this particular structure and finally checking or validating the resulting model in order to ensure that it describes accurately the real life system.

1. Introduction

Our present knowledge of nature and the universe is a result of a very slow and painful process of distilling the information gathered by observing physical, social, moral and economic phenomena. It was by thoroughly observing nature that our ancestors realized that our world has many features of regularity which could be recorded, analyzed and predicted. During the entire history of mankind humans have looked for patterns in their environment to improve decision making process. The phases of the moon, the tides, the succession of the seasons and even solar eclipses could be predicted by ancient civilizations of Egyptians, Chaldeans or Mayans using sophisticated calendars.

The development of mathematics and in parallel of more accurate measurement devices lead to the introduction of more accurate mathematical models of the natural processes. A crucial step in this direction was the introduction of differential calculus independently by Newton and Leibnitz. This allowed the derivation of accurate models of dynamical systems in all branches of science and engineering. Heat transfer, propagation of the electromagnetic field and gravity could all be described accurately in terms of differential equations. Mathematical models are fundamental for the analysis of system behavior, in controller design and many scientific and engineering studies. In every case an essential requirement of the model is an ability to reproduce the dynamical characteristics of the system as closely as possible. At the heart of the modeling process are practical experiments which are designed to probe the dynamical behavior of systems and to provide experimental data on which to base the modeling process. The final model would then be subject to a validation process in which the accuracy of the model predictions would be assessed in independent experiments.

The models of the fundamental processes in nature have become physical laws based on which models of most systems in the real world can be derived. This can be viewed as an axiomatic approach to modeling where the dynamics of a system are described in mathematical terms by analyzing its structure and applying elementary laws of physics or chemistry that govern its behavior. For many real systems however, this approach is not suitable either because the modeler does not have access to the internal structure of the system or because the system complexity is so high that the application of elementary laws to derive the system's equations is practically impossible.

In such case an experimental modeling approach is the only solution. During the past century this approach to modeling has evolved tremendously and has become a cornerstone of engineering practice, especially in the field of Automatic Control where

UNESCO – EOLSS SAMPLE CHAPTERS

it has been established as a well articulated theory of System Identification.

2. System Identification

In system identification the structure of the dynamical process under sturdy is often assumed to be unknown. Deriving a mathematical model of a dynamical system by this approach involves in the first stage applying a pre-designed sequence of stimuli, the input signal, to the system and recording the responses of the system to these stimuli, the output signal. These experiments attempt to capture all the effects the inputs have on the observed outputs which constitutes the input/output behavior of the system under study.

In the first stage, specific model selection and parameter estimation algorithms are used to derive, from the recorded input/output data, a mathematical model of the dynamical system which can reproduce the measured output signal when simulated using the same input sequence. A large class of dynamical systems can be represented with good approximation by linear expressions involving the variables which characterize the system. However, although this simple type of mathematical description has formed the basis for the development of various theories, such as mathematical control theory, ultimately, linear techniques are limited by the fact that real systems are more often than not nonlinear. It is true that local analysis and control design can in general be carried out satisfactorily using only such models, but there are many cases when a linear description of a process is not sufficient and a global, more accurate nonlinear model is required.

Depending on whether the variables in the model can be computed at any time instance or only at certain discrete instances of time, the models can be classified into continuous-time models, such as differential equations and discrete-time models described by difference equations. Usually the experimental data is measured/sampled at regular time intervals and hence the models used for identification are in most cases discrete-time. An advantage of this approach is that there is no need to use additional integration routines when simulating discrete-models as would be the case with most nonlinear differential equations. The main disadvantage is that the discrete-time models are only valid for the chosen sampling time. The last stage of the identification process involves performing various model validation tests to ensure that the identified model provides an accurate representation of the observed system.

3. Nonlinear vs. Linear Models

The most common misinterpretation of reality is when scientists assume that all systems can be modeled using linear techniques. In the early stages the theory of system identification was focused mainly on the identification of linear dynamical systems. The lack of techniques to deal with nonlinear interactions meant that linearization occurred from the moment the equations were set. The common assumption was that a linear combination of the variables that characterize a system is sufficient to explain the system behavior. This assumption presumes that all the other possible interactions of a variable with other variables in the system are sufficiently weak so that for some specified conditions the nonlinear interactions may be neglected.

UNESCO – EOLSS SAMPLE CHAPTERS

It was assumed therefore that physical reality could be segmented, understood piece by piece and re-assembled, using the superposition principle, to form a linear model of the original system. Any remaining interaction would then be treated as a perturbation of the linear solution assuming throughout that this perturbation is not going to significantly alter the properties of the system. As linear models cannot reproduce dynamical regimes which result from nonlinear interactions such as hysteresis, amplitude dependence, bifurcation or chaos, the linear concepts, which literally collapsed under the weight of uninterpretable data, had to be replaced in many cases by nonlinear ones. Nonlinear dynamics has subsequently been revealed in practically every field of science. This made the use of nonlinear representations to describe the observed behavior unavoidable and triggered significant advances in nonlinear system identification.

One of the most complete and powerful identification methodologies available today for the identification of both linear and nonlinear systems is the NARMAX approach. This method is based on a general input/output representation of dynamical systems called the NARMAX (Nonlinear Autoregressive Moving Average with eXogenous inputs) model which is a generalization of the earlier ARMAX (Autoregressive Moving Average with eXogenous inputs) representation of linear discrete-time systems. Essentially the NARMAX model is a nonlinear difference equation which relates the output of the system at a given time instance to values of the inputs, outputs and noise at previous time instances.

The NARMAX model can describe a wide range of nonlinear dynamical behaviors and includes many other nonlinear model types, such as the Volterra, Hammerstein and Wiener models, as special cases. In other words NARMAX provides a one stop solution to a wide range of identification problems involving nonlinear systems ranging from chaotic electronic circuits and chaotic reactor systems to water management systems, turbocharged diesel engines, vasomotion oscillations in the brain or solar plasma turbulence. All these examples are only a fraction of real-life system identification problems where NARMAX has been successfully applied. In practice the nonlinear equation that describes the underlying system is actually not known in advance. The NARMAX methodology provides a unified solution to the problem of finding this equation based only on experimental data recorded from the system of interest.

UNESCO – EOLSS SAMPLE CHAPTERS

4. The NARMAX Model

A dynamical system is a system whose state is characterized in terms of a finite number of variables, the state variables, which evolve in time according to a transition rule. The transition rule, which in practice may be the solution of a system of differential equations or a difference equation in the case of a discrete-time system, can be used to determine the state of the system at any future time instant given the current or past state of the system.

In many cases the state variables cannot be accessed directly, that is the state values cannot be measured by an external observer of the system. What an external observer can measure are the outputs of the system. The number of the outputs is not necessarily the same as the number of the state variables. The relationship between the state

variables and the outputs of the system is usually described by a function known as the measurement or observation map. Dynamical systems are often characterized only in terms of the input/output behavior which illustrates the effects of the system's inputs on the observed outputs. The assumption is that all the information about the state should in principle be recoverable from the output measurements. In many cases it is more advantageous to have a direct characterization of the output of the system as a function of the input only.

The function that relates directly the inputs and the outputs of a system without including explicitly any of the state variables is known as the input/output equation of the system. In system identification the input/output behavior of a dynamical system is measured experimentally over a finite time interval resulting in a set of input and output data samples which are usually contaminated by noise. In this context the system identification problem consists in determining the unknown discrete-time input/output equation that relates explicitly the sampled outputs of the system $y(t), y(t-1),...$ to the sampled inputs $u(t), u(t-1),...$. Although it is still possible to identify continuous time models from this data set, the identification of discrete-time models is usually more convenient. Discrete-time models are widely used for simulation an analysis as well as in the design of digital control systems.

The NARMAX model is a natural extension of the input/output equations used for linear discrete-time systems namely the ARMAX model. The NARMAX model takes the form of a set of nonlinear difference equations

$$y(t) = f(y(t-1),...,y(t-n_y),u(t-d),...,u(t-n_u),e(t-1),...,e(t-n_e)) + e(t) \qquad (1)$$

which relate the inputs and outputs and which takes into account the combined effects of measurement noise, modeling errors and unmeasured disturbances represented by the variable $e(k)$. A rigorous derivation of the NARMAX model and conditions for the existence of this model are available in the literature. In the above equations $f(\cdot)$ is an unknown nonlinear mapping, $u(t)$ is the input vector, $y(t)$ is the output vector and n_y and n_u are the maximum output and input lags. The noise variable vector $e(t)$, which cannot be measured directly, is assumed to be bounded $|e(t)| < \delta$ and uncorrelated with the input and n_e is the maximum noise lag. The random variable $e(t)$ is also known as the prediction error or innovation at time t.

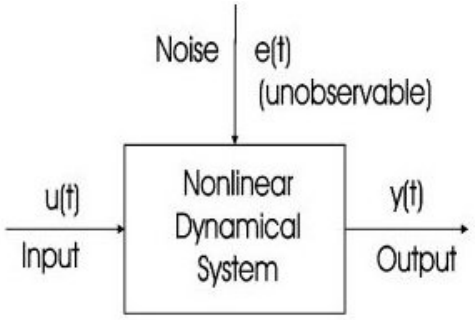

Figure 1: Block Diagram of a Nonlinear Dynamical System

UNESCO – EOLSS SAMPLE CHAPTERS

5. Practical Implementations of the NARMAX Model

The methodology of identifying a NARMAX representation involves estimating both the structure and the parameters of the unknown nonlinear system from the input/output data. This is quite a formidable task since the number of possible nonlinear implementations of $f(\cdot)$ is theoretically infinite. Because the form of $f(\cdot)$ for a real system is rarely known a priori the model is implemented practically using certain functions with good approximation properties. In this context finding the nonlinear function $f(\cdot)$ which best agrees with the experimental data according to some adequacy criterion becomes a nonlinear approximation problem where the choice of approximating functions is very important. For some types of approximating functions for example, the approximation is not guaranteed to converge to an arbitrary function. These can be used efficiently for certain families of nonlinear systems. There are a number of nonlinear representations that have received a great deal of attention because of their excellent approximation properties. Most commonly, NARMAX models can be implemented using polynomial and rational representations, neural networks or wavelets.

5.1. Polynomials and Rational Implementations

Polynomial models are possibly the most attractive of all nonlinear representations due to the inherent simplicity of the model structure and because such models reveal the dynamical properties of the underlying system is a very straightforward manner.

The nonlinear map $f(\cdot)$, in this case takes the form of a multivariable polynomial of finite degree in all variables

$$f(x) = \sum_{k}^{n} \theta_k g_k(\mathbf{x}) \qquad (2)$$

where θ_k are the coefficients of the polynomial, g_k represent multivariable polynomial terms up to a given order m and $\mathbf{x}$ is the vector

$$\mathbf{x} = [y(t-1),...,y(t-n_y),u(t-1),...,u(t-n_u),e(t-1),...,e(t-n_e)] \qquad (3)$$

of past outputs, inputs and noise. Using polynomial NARMAX models to represent nonlinear input/output behavior can be justified by results from approximation theory. However it is very important to determine the appropriate terms in a polynomial expansion. An important feature of the polynomial model is that it is linear in the parameters. Many linear identification results can easily be extended to the polynomial nonlinear model and several algorithms for model structure determination and parameter estimation exploit this property.

A more general representation which includes polynomial models as a special case is the rational representation where $f(\cdot)$ is represented as the ratio of two polynomial functions

UNESCO – EOLSS SAMPLE CHAPTERS

$$f(\mathbf{x}) = \frac{a(\mathbf{x})}{b(\mathbf{x})} \tag{4}$$

where $a(\cdot)$ and $b(\cdot)$ are polynomials of finite degree. Such nonlinear mappings have the advantage that they can represent certain types of singular or near-singular behavior which cannot be achieved with polynomials. The main disadvantage is that rational models are nonlinear in the parameters and it introduces a number of difficulties in the estimation of such representation. However, algorithms which overcome most of these difficulties have been developed and applied successfully in practice.

5.2 Neural Network Representations

Artificial neural networks refer to a computational paradigm in which a large number of computational units or "neurons" interconnected to form a network perform complex computational tasks. This computational model was inspired by neurobiological systems that have the capability of learning from examples.

In systems identification the learning process can be associated with the parameter estimation and structure selection algorithms.

There are two main types of artificial neural networks which have been used extensively in system identification namely Multilayer Perceptron Networks and Radial Basis Networks.

5.2.1 Multilayer Perceptron Networks

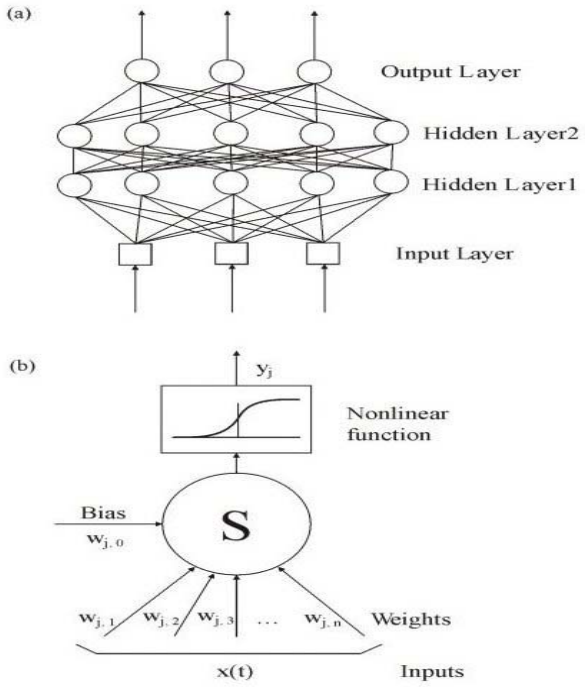

Figure 2: (a) Multilayer Perceptron Neural Network Architecture and (b) Model of a Neuron

Multilayer Perceptron Neural Networks (MLP) consist of one or more hidden layers of interconnected computing units or nodes sandwiched between the input and output layers. As seen in Figure 2a, the nodes in a layer are connected only with nodes in adjacent layers and not with nodes within the same layer.

The input/output relationship of a generic node is illustrated in Figure 2. The output of each node is obtained by calculating a weighted sum of the node inputs, adding a bias and passing the result through a nonlinear activation function

$$y_j = g\left(w_{j,0} + \sum_{i=1}^{n} w_{j,i} x_i \right) \tag{5}$$

where $w_{j,i}$ are the node weights and $w_{j,0}$ is the bias term. The best known choice of activation function is the sigmoid function

$$g(v) = \frac{1}{1+\exp(-v)} \tag{6}$$

The network is in fact just another general function approximation device. Given a set of input/output data the network can be trained to learn the underlying relationship by adjusting the weights/parameters for each node. Theoretical work has shown that actually one layer of nodes is sufficient to approximate any continuous function provided that there are sufficient nodes in this layer.

The NARMAX model can be implemented in a straightforward manner using MLP Neural Networks with a single hidden layer. In this case expression of the nonlinear function $f(\cdot)$ in (1) becomes

$$f(\mathbf{x}) = \sum_{j=1}^{m} \frac{w_j}{1+\exp[-(w_{j,0} + \sum_{i=1}^{n} w_{j,i} x_i)]} \tag{7}$$

where $\mathbf{x} = [x_1 ... x_n]$ is as before the vector of lagged outputs, inputs and noise and w_j are the output weights. In this case the function $f(\cdot)$ is nonlinear in the parameters $w_{j,i}$.

UNESCO – EOLSS SAMPLE CHAPTERS

-
-
-

TO ACCESS ALL THE **30 PAGES** OF THIS CHAPTER,
Click here

Bibliography

Billings, S. A. (1989a). Identification of nonlinear systems – A survey. *IEE Proceedings Part D* **127**,272-285. [A survey of traditional nonlinear identification procedures.]

Billings, S. A. and Chen, S. (1989b). Extended model set, global data and threshold model identification of severely nonlinear systems. *International Journal of Control*, **50**,1897-1923. [Shows how NARMAX can be used to model complex nonlinear systems.]

Billings, S. A. and Chen, S. (1998). The determination of multivariable nonlinear models for dynamic systems. In Leondes, C.T., editor, *Neural Network System Techniques and Applications; Control and Dynamic Systems Series 7* 231-278. [A comprehensive review of NARMAX modeling techniques for multivariable nonlinear dynamic systems.]

Billings, S. A., Chen, S. and Kronenberg, M. J. (1988). Identification of MIMO Non-linear Systems using a Forward-regression Orthonormal Estimator. *International Journal of Control*, **50,**2157-2189. [Discusses the identification of multi-input multi-output nonlinear systems using the orthogonal least squares (OLS) algorithm.]

Chen, S. and (1989) Billings, S. A. (1989). Representations of non-liner systems: the NARMAX model. *International Journal of Control*, **49**,1013-1032. [Describes the range of systems that NARMAX can be applied to.]

Haber, R. and Unbehauen, H. (1990). Structure Identification for Nonlinear Dynamical Systems. A survey on Input-Output Approaches. *Automatica*, **26,**651-677. [Reviews the identification of certain classes of nonlinear systems.]

Leotaritis, I. and Billings, S. A. (1985a). Input-Output Parametric Models for Non-linear Systems- Part I: Deterministic Non-linear Systems. *International Journal of Control*, **41,** 303-328; (1985b) Input-Output Parametric Models for Non-linear Systems –Part II: Stochastic Non-linear Systems. *International Journal of Control*, **41**, 329-344. [Formal introduction to the NARMAX model.]

Juditski, A., Hjalmarsson, H., Beneviste, B., Deylon, H., Ljung, L., Sjóberg, J. and Zhang, Q. (1995). Nonlinear Black-Box Models in System Identification: Mathematical Foundations. *Automatica*, **31**, 1725-1750. [Discusses the theoretical foundation of various nonlinear model implementations.]

Person, R. K. (1995). Nonlinear input/output modeling. *Journal of Process Control*, **5,**197-211. [A good abbreviated overview of empirical nonlinear modeling and identification.]

Person, R. K. and Ogunnaike, B.A. (1997). Nonlinear process identification. In Henson, M. A. and Seborg, D. E., editors, *Nonlinear Process Control*, chapter 2. Prentice Hall, Englewood Cliff. [Excellent review of nonlinear system identification.]

Swain, A. K., Billings, S. A., Stansby, P.K. and Baker, M. (1998). Accurate predictions of nonlinear wave forces; Part I: Fixed cylinder. *Mechanical Systems and Signal Processing*, **21**,449-485. [Shows how NARMAX can be used in practice to determine fundamental models.]

Biographical Sketches

Stephen A Billings received the BEng degree in Electrical Engineering with first class honours from the University of Liverpool in 1972, the degree of PhD in Control Systems Engineering from the University

of Sheffield in 1976, and the degree of DEng from the University of Liverpool in 1990. He is a Chartered Engineer [CEng] , Chartered Mathematician [CMath], Fellow of the IEE [UK] and Fellow of the Institute of Mathematics and its Applications.

He was appointed as Professor in the Department of Automatic Control and Systems Engineering, University of Sheffield, UK in 1990 and leads the Signal Processing and Complex Systems research group. His research interests include system identification and information processing for nonlinear systems, narmax methods, model validation, prediction, spectral analysis, adaptive systems, nonlinear systems analysis and design, neural networks, wavelets, fractals, machine vision, cellular automata, spatio-temporal systems, fMRI and optical imagery of the brain, metabolic systems engineering, systems biology and related fields. The Institute of Scientific Information in the USA has recently identified Professor Billings as one of the world's most highly cited researchers in all branches of engineering over the past 20 years.

Daniel Coca received the MEng degree in Electrical Engineering with first class honors from "Transilvania" University of Brasov in 1993 and the PhD degree in Control Systems Engineering from the University of Sheffield in 1997. He is a Chartered Engineer [CEng] and member of the IEE [UK].

He is currently a lecturer in the Department of Automatic Control and Systems Engineering, University of Sheffield, UK. His research interests include nonlinear and complex systems identification, control theory, distributed parameter systems, protein identification using MALDI-TOF mass spectrometry, brain activity modeling based on neuroimaging data, modeling and analysis of solar wind-magnetosphere interaction.

UNESCO – EOLSS SAMPLE CHAPTERS