WAVELET NETWORK BASED NONLINEAR SYSTEM IDENTIFICATION

G.P. Liu, S.A. Billings and V. Kadirkamanathan

Department of Automatic Control & Systems Engineering
University of Sheffield, PO Box 600, Mappin Street, Sheffield S1 3JD, UK
Email: steve/guoping/visakan@acse.sheffield.ac.uk

Research Report No. **657**

Department of Automatic Control & Systems Engineering
The University of Sheffield

November 1996

Wavelet Network Based Nonlinear System Identification

G. P. Liu, S. A. Billings and V. Kadirkamanathan

Department of Automatic Control and Systems Engineering
University of Sheffield, PO Box 600, Mappin Street, Sheffield S1 3JD, U.K.
Tel: +44 1142825231 Fax: +44 1142731729
Email: guoping/steve/visakan@acse.sheffield.ac.uk

Abstract

Identification schemes using wavelet networks are presented for nonlinear dynamical systems. Based on fixed wavelet networks, parameter adaptation laws are developed using a Lyapunov synthesis approach. This guarantees the stability of the overall identification scheme and the convergence of both the parameters and the state errors, even in the presence of modelling errors. Using the decomposition and reconstruction techniques of multiresolution decompositions, variable wavelet networks are introduced to achieve a desired estimation accuracy and a suitable sized network, and to adapt to variations of the characteristics and operating points in nonlinear systems. B-spline wavelets are used to form the wavelet networks and the identification scheme is illustrated using a simulated example.

Keywords: Wavelets, networks, nonlinear system, B-splines, identification, multiresolution.

1 Introduction

Nonlinear system identification consists of model structure selection and parameter estimation. The first problem is concerned with selecting a class of mathematical operator as a model. The second is concerned with an estimation algorithm based on input-output data from the process, a class of models to be identified and a suitable identification criterion. A number of techniques have been developed in recent years for model selection and parameter estimation of nonlinear systems. Forward and backward regression algorithms were analyzed in [18]. Stepwise regression was used in [2] and a class of orthogonal estimators were discussed in [17]. Algorithms which save memory and allow fast computations have also been proposed in [6]. Methods to determine a priori structural identifiability of a model have also been studied [21]. A survey of existing techniques of nonlinear system identification prior to the 1980s is given in [1], a survey of the structure detection of input-output nonlinear systems is given in [13] and a recent survey of nonlinear black-box modelling in system identification can be found in [29].

The approximation of general continuous functions by nonlinear networks has been widely applied to system modelling and identification. Such approximation methods are particularly useful in the black-box identification of nonlinear systems where very little a priori knowledge is available. For example, neural networks have been established as a general approximation tool for fitting nonlinear models from input-output data on the basis of the universal approximation property of such networks (see, for example, [5] [19] [26]). There has also been considerable recent interest in identification of general nonlinear systems based on radial basis networks [27], fuzzy sets and rules [31], neural-fuzzy networks [4] [30] and hinging hyperplanes [3].

The recently introduced wavelet decomposition [7] [10] [12] [14] [22] [24] also emerges as a new powerful tool for approximation. In recent years, wavelets have become a very active subject in many scientific and engineering research areas. Wavelet decompositions provide a useful basis for localised approximation of functions with any degree of regularity at different scales and with a desired accuracy. Recent advances have also shown the existence of orthonormal wavelet bases, from which follows the variability of rates of convergence for approximation by wavelet based networks. Wavelets can therefore be viewed as a new basis for representing functions. Wavelet based networks (or simply wavelet networks) are inspired by both the feedforward neural networks and wavelet decompositions and have been introduced for the identification of nonlinear static systems [32]. But little attention has been paid to identification of nonlinear dynamical systems using wavelet networks [8] [9].

This paper presents a wavelet network based identification scheme for nonlinear dynamical systems. Two kinds of wavelet networks are studied: fixed and variable wavelet networks. The former is used for the case where the estimation accuracy is assumed to be achieved by a known resolution scale. But, in practice, this assumption is not realistic because the nonlinear function to be identified is unknown and the system operating point may change with time.

Thus, variable wavelet networks are introduced to deal with this problem. The basic principle of the variable wavelet network is that the number of wavelets in the network can either be increased or decreased over time according to a design strategy in an attempt to avoid overfitting or underfitting. In order to model unknown nonlinearities, the variable wavelet network starts with a lower resolution scale and then increases or reduces this according to the novelty of the observation. Since the novelty of the observation is tested, it is idealy suited for on-line identification problems. The objective behind the development is to gradually approach the appropriate network complexity that is sufficient to provide an approximation to the system nonlinearities and which is consistent with the observations received.

The parameters of the wavelet network are adjusted by adaptation laws developed using a Lyapunov synthesis approach. The identification algorithm is performed over the network parameters by taking advantage of the decomposition and reconstruction algorithms of a multiresolution decomposition when the resolution scale changes in the variable wavelet network. Combining the wavelet network and Lyapunov synthesis techniques, the identification algorithm developed for continuous dynamical nonlinear systems guarantees the stability of the whole identification scheme and the convergence of both the parameters and estimation errors. The wavelet network based identification scheme is realised using B-spline wavelets and it is shown how to calculate decomposition and reconstruction sequences needed for identification using variable wavelet networks. A simulated example shows the operation of the proposed identification scheme.

2 Wavelet Networks

Wavelets are a class of functions which have some interesting and special properties. Some basic concepts about orthonormal wavelet bases will be introduced initially. Then the wavelet series representation of one-dimensional and multi-dimensional functions will be considered. Finally, wavelet networks are introduced.

Throughout this paper, the following notations will be used. $\mathbb{Z}$ and $\mathbb{R}$ denote the set of integers and real numbers, respectively. $L_2(\mathbb{R})$ denotes the vector space of measurable, square-integrable one-dimensional functions $f(x)$. For $f, g \in L_2(\mathbb{R})$, the inner product and norm for the space $L_2(\mathbb{R})$ are written as

$$< f, g >:= \int_{-\infty}^{\infty} f(x)\overline{g(x)}dx \tag{1}$$

$$\| f \|_2 :=< f, f >^{1/2} \tag{2}$$

where $\overline{g(.)}$ is the conjugate of the function $g(.)$. $L_2(\mathbb{R}^n)$ is the vector space of measurable, square-integrable n-dimensional functions $f(x_1, x_2, ..., x_n)$. For $f, g \in L_2(\mathbb{R}^n)$, the inner product of $f(x_1, x_2, ..., x_n)$ with $g(x_1, x_2, ..., x_n)$ is written as

$$< f(x_1, x_2, ..., x_n), g(x_1, x_2, ..., x_n) >=$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,x_2,...,x_n)\overline{g(x_1,x_2,...,x_n)}dx_1dx_2...dx_n \qquad (3)$$

The original objective of the theory of wavelets is to construct orthogonal bases in $L_2(\mathbb{R})$. These bases are constituted by translations and dilations of the same function ψ. It is preferable to take ψ as localized and regular. The principle of wavelet construction is the following: (a) the function $\phi(x-k)$ are mutually orthogonal for k ranging over $\mathbb{Z}$; (b) ϕ is a scaling function and the family $\phi(2^jx-k)$ constitutes an orthogonal basis of $L_2(\mathbb{R})$; (c) the wavelet is defined as ψ and the family $\psi(2^jx-k)$ constitutes an orthogonal basis of $L_2(\mathbb{R})$. It can also be proved that the family $\{\phi(2^{j_0}x-k), \psi(2^jx-k), \text{for } j\geq j_0\}$ also forms an orthogonal basis of $L_2(\mathbb{R})$.

The wavelet subspaces W_j are defined as

$$W_j = \{\psi(2^jx-k), \quad k\in\mathbb{Z}\} \qquad (4)$$

which satisfy

$$W_j \bigcap W_i = \{\emptyset\}, \quad j\neq i \qquad (5)$$

Any wavelet generates a direct sum decomposition of $L_2(\mathbb{R})$. For each $j\in\mathbb{Z}$, let us consider the closed subspaces:

$$V_j = ... \oplus W_{j-2} \oplus W_{j-1} \qquad (6)$$

of $L_2(\mathbb{R})$, where $\oplus$ denotes the direct sum. These subspaces have the following properties:

(i) $... \subset V_{-1} \subset V_0 \subset V_1 \subset ...$;

(ii) $clos_{L_2}\left(\bigcup_{j\in\mathbb{Z}} V_j\right) = L_2(\mathbb{R})$;

(iii) $\bigcap_{j\in\mathbb{Z}} V_j = \{\emptyset\}$;

(iv) $V_{j+1} = V_j + W_j$, $j\in\mathbb{Z}$; and

(v) $f(x)\in V_j \Longleftrightarrow f(2x)\in V_{j+1}, j\in\mathbb{Z}$.

Hence, the sequence of subspaces V_j is nested, as described by property (i). Property (ii) shows that every function f in $L^2(\mathbb{R})$ can be approximated as closely as desirable by its projections, denoted by P_jf in V_j. But, by decreasing j, the projections P_jf could have arbitrarily small energy, as guaranteed by property (iii). The most important intrinsic property of these spaces is that more and more variations of P_jf are removed as $j\rightarrow-\infty$. In fact, these variations are peeled off, level by level in decreasing order of the rate of variations and stored in the

complementary subspaces W_j as in property (iv). This process can be made very efficient by an application of property (v).

If ϕ and ψ are compactly supported, they give a local description, at different scales j, of the considered function. The wavelet series representation of the one-dimensional function $f(x)$ is given by

$$f(x) = \sum_{k\in\mathbb{Z}} a_{j_0 k}\phi_{j_0 k}(x) + \sum_{j\geq j_0}\sum_{k\in\mathbb{Z}} b_{jk}\psi_{jk}(x) \tag{7}$$

where $\phi_{j_0 k}(x) = 2^{j_0/2}\phi(2^{j_0}x - k)$, $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$, and the wavelet coefficients $a_{j_0 k}$ and b_{jk} are

$$a_{j_0 k} =< f(x), \phi_{j_0 k}(x) > \tag{8}$$

$$b_{jk} =< f(x), \psi_{jk}(x) > \tag{9}$$

The wavelet series representation can easily be generalized to any dimension n. For the n-dimensional case $\mathbf{x} = [x_1, x_2, ..., x_n]$, we introduce the scaling function

$$\Phi(\mathbf{x}) = \phi(x_1)\phi(x_2)...\phi(x_n) \tag{10}$$

and the $2^n - 1$ mother wavelets $\Psi_i(\mathbf{x})$, $i = 1, 2, ..., 2^n - 1$, are obtained by substituting some $\phi(x_j)$s by $\psi(x_j)$ in (10). Then the following family is an orthonormal basis in $L_2(\mathbb{R}^n)$:

$$\{\Phi_{j_0\mathbf{k}}(\mathbf{x}), \Psi^{(1)}_{j\mathbf{k}}(\mathbf{x}), \Psi^{(2)}_{j\mathbf{k}}(\mathbf{x}), ..., \Psi^{(2^n-1)}_{j\mathbf{k}}(\mathbf{x})\} \tag{11}$$

for $j \geq j_0$, $j \in \mathbb{Z}$, $\mathbf{k} = [k_1, k_2, ..., k_n] \in \mathbb{Z}^n$, and

$$\Phi_{j\mathbf{k}}(\mathbf{x}) = 2^{jn/2}\Phi(2^j x_1 - k_1, 2^j x_2 - k_2, ..., 2^j x_n - k_n) \tag{12}$$

$$\Psi^{(i)}_{j\mathbf{k}}(\mathbf{x}) = 2^{jn/2}\Psi_i(2^j x_1 - k_1, 2^j x_2 - k_2, ..., 2^j x_n - k_n) \tag{13}$$

For $f(\mathbf{x}) \in L_2(\mathbb{R}^n)$, the n-dimensional wavelet series representation of the function $f(\mathbf{x})$ is

$$f(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^n} a_{j_0\mathbf{k}}\Phi_{j_0\mathbf{k}}(\mathbf{x}) + \sum_{j\geq j_0}\sum_{\mathbf{k}\in\mathbb{Z}^n}\sum_{i=1}^{2^n-1} b^{(i)}_{j\mathbf{k}}\Psi^{(i)}_{j\mathbf{k}}(\mathbf{x}) \tag{14}$$

where the wavelet coeffiecients are

$$a_{j_0\mathbf{k}} =< f(\mathbf{x}), \Phi_{j_0\mathbf{k}}(\mathbf{x}) > \tag{15}$$

$$b^{(i)}_{j\mathbf{k}} =< f(\mathbf{x}), \Psi^{(i)}_{j\mathbf{k}}(\mathbf{x}) > \tag{16}$$

For system identification, $f(\mathbf{x})$ is unknown. Then the wavelet coefficients $a_{j_0\mathbf{k}}$ and $b^{(i)}_{j\mathbf{k}}$ can not be calculated simply by (15) and (16). As (11) shows, constructing and storing orthonormal wavelet bases involves a prohibitive cost for large dimensions n. In addition, it is not realistic to use an infinite number of wavelets to represent the function $f(\mathbf{x})$. So, we consider the following wavelet representation of the function $f(\mathbf{x})$:

$$\hat{f}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathcal{A}_{j_0}} a_{j_0\mathbf{k}}\Phi_{j_0\mathbf{k}}(\mathbf{x}) + \sum_{j=j_0}^{N}\sum_{\mathbf{k}\in\mathcal{B}_j}\sum_{i=1}^{2^n-1} b^{(i)}_{j\mathbf{k}}\Psi^{(i)}_{j\mathbf{k}}(\mathbf{x}) \tag{17}$$

where $\mathcal{A}_{j_0}, \mathcal{B}_j \in \mathbb{Z}^n$ are the finite vector sets of integers and $N \in \mathbb{R}^1$ is a finite integer. Since the convergence of the series in (14) is in $L_2(\mathbb{R}^n)$,

$$\lim_{N\to\infty, \mathcal{A}_{j_0}, \mathcal{B}_{j_0}, ..., \mathcal{B}_N \to \mathbb{Z}^n} ||f(\mathbf{x}) - \hat{f}(\mathbf{x})||_2 = 0 \tag{18}$$

Hence, given $\varepsilon > 0$, there exists a number N^* and vector sets $\mathcal{A}^*_{j_0}, \mathcal{B}^*_{j_0}, \mathcal{B}^*_{j_0+1}, ..., \mathcal{B}^*_N$ such that for $N \geq N^*$, $\mathcal{A}_{j_0} \subseteq \mathcal{A}^*_{j_0}$ and $\mathcal{B}_{j_0} \subseteq \mathcal{B}^*_{j_0}$, $\mathcal{B}_{j_0+1} \subseteq \mathcal{B}^*_{j_0+1}$,..., $\mathcal{B}_N \subseteq \mathcal{B}^*_N$,

$$||f(\mathbf{x}) - \hat{f}(\mathbf{x})||_2 \leq \varepsilon \tag{19}$$

This shows that the required approximation accuracy of the function f by $\hat{f}$ can be guarantted by properly choosing the number N and the vector sets $\mathcal{A}_{j_0}, \mathcal{B}_{j_0}, ..., \mathcal{B}_N$. Following neural networks, the expression (17) is called a wavelet network. In this network, the parameters $a_{j_0\mathbf{k}}$ and $b^{(i)}_{j\mathbf{k}}$, and the number N and the vector sets $\mathcal{A}_{j_0}, \mathcal{B}_{j_0}, ..., \mathcal{B}_N$ will jointly be determined from the data, based on the scaling functions ϕ and the wavelets ψ.

3 Identification Using Fixed Wavelet Networks

Consider the multi-input multi-state (MIMS) continuous dynamical system described by,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}), \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{20}$$

where $\mathbf{u} \in \Re^{r\times 1}$ is the input vector, $\mathbf{x} \in \Re^{d\times 1}$ is the state vector and $\mathbf{f}(\cdot) = [f_1(\cdot), f_2(\cdot), ...f_d(\cdot)]^T \in L^d_2(\mathbb{R}^n)$ is an unknown nonlinear function vector. It is also assumed that $\mathbf{u}, \mathbf{x}$ are in compact sets.

Following the structure (17) of wavelet networks, at the resolution 2^{N+1}, the estimation $\hat{\mathbf{f}}(\mathbf{x}, \mathbf{u})$ of the function $\mathbf{f}(\mathbf{x}, \mathbf{u})$ is the output vector of the wavelet network which is expressed by

$$\hat{\mathbf{f}}(\mathbf{x}, \mathbf{u}) = \sum_{\mathbf{k}\in\mathcal{A}_{j_0}} \mathbf{A}_{j_0\mathbf{k}} \Phi_{j_0\mathbf{k}}(\mathbf{x}, \mathbf{u}) + \sum_{j=j_0}^{N} \sum_{\mathbf{k}\in\mathcal{B}_j} \sum_{i=1}^{2^n-1} \mathbf{B}^{(i)}_{j\mathbf{k}} \Psi^{(i)}_{j\mathbf{k}}(\mathbf{x}, \mathbf{u}) \tag{21}$$

where $n = r + d$, and $\mathbf{A}_{j_0\mathbf{k}}, \mathbf{B}^{(i)}_{j\mathbf{k}} \in \mathbb{R}^n$ are the wavelet coefficient vectors, the scaling function $\Phi_{j_0\mathbf{k}}(\mathbf{x}, \mathbf{u})$ and the wavelet functions $\Psi^{(i)}_{j\mathbf{k}}(\mathbf{x}, \mathbf{u})$ are similarly defined as $\Phi(\mathbf{x})$ and $\Psi_j(\mathbf{x})$, respectively, by replacing $\mathbf{x}$ with $(\mathbf{x}, \mathbf{u})$.

Here, it is assumed that the number N and the vector sets $\mathcal{A}_{j_0}, \mathcal{B}_j$ are given. So, the wavelet network (21) for the estimation of the nonlinear function $\mathbf{f}(\mathbf{x}, \mathbf{u})$ is called a fixed wavelet network. Based on the estimation $\hat{\mathbf{f}}(\mathbf{x}, \mathbf{u})$ by the fixed wavelet network, the nonlinear function $\mathbf{f}(\mathbf{x}, \mathbf{u})$ can be expressed by

$$\mathbf{f}(\mathbf{x}, \mathbf{u}) = \sum_{\mathbf{k}\in\mathcal{A}_{j_0}} \mathbf{A}^*_{j_0\mathbf{k}} \Phi_{j_0\mathbf{k}}(\mathbf{x}, \mathbf{u}) + \sum_{j=j_0}^{N} \sum_{\mathbf{k}\in\mathcal{B}_j} \sum_{i=1}^{2^n-1} \mathbf{B}^{(i)*}_{j\mathbf{k}} \Psi^{(i)}_{j\mathbf{k}}(\mathbf{x}, \mathbf{u}) + \varepsilon_N \tag{22}$$

where the optimal wavelet coefficient vectors $\mathbf{A}^*_{j_0\mathbf{k}}$ and $\mathbf{B}^{(i)*}_{j\mathbf{k}}$ are

$$\mathbf{A}^*_{j_0\mathbf{k}} = \begin{bmatrix} < f_1(\mathbf{x},\mathbf{u}), \Phi_{j_0\mathbf{k}}(\mathbf{x},\mathbf{u}) > \\ < f_2(\mathbf{x},\mathbf{u}), \Phi_{j_0\mathbf{k}}(\mathbf{x},\mathbf{u}) > \\ \vdots \\ < f_d(\mathbf{x},\mathbf{u}), \Phi_{j_0\mathbf{k}}(\mathbf{x},\mathbf{u}) > \end{bmatrix} \tag{23}$$

$$\mathbf{B}^{(i)*}_{jk} = \begin{bmatrix} < f_1(\mathbf{x},\mathbf{u}), \Psi^{(i)}_{j\mathbf{k}}(\mathbf{x},\mathbf{u}) > \\ < f_2(\mathbf{x},\mathbf{u}), \Psi^{(i)}_{j\mathbf{k}}(\mathbf{x},\mathbf{u}) > \\ \vdots \\ < f_d(\mathbf{x},\mathbf{u}), \Psi^{(i)}_{j\mathbf{k}}(\mathbf{x},\mathbf{u}) > \end{bmatrix} \tag{24}$$

$\varepsilon_N = [\varepsilon_{N1}, \varepsilon_{N2}, ..., \varepsilon_{Nd}]^T$ is the modelling error vector which is assumed to be bounded by

$$\bar{\varepsilon}_N = \max_{i=1,2,...,d} \sup_{t\in\mathbb{R}^+} \{|\varepsilon_{Ni}(t)|\} \tag{25}$$

Modelling the nonlinear function vector $\mathbf{f}(\mathbf{x},\mathbf{u})$ using wavelet networks gives the following identification model for the nonlinear dynamical system (20):

$$\dot{\hat{\mathbf{x}}} = \mathbf{A}(\hat{\mathbf{x}} - \mathbf{x}) + \hat{\mathbf{f}}(\mathbf{x},\mathbf{u}), \quad \hat{\mathbf{x}}(0) = \mathbf{x}_0 \tag{26}$$

where $\hat{\mathbf{x}}$ denotes the state vector of the network model and $\mathbf{A} \in \mathbb{R}^{d\times d}$ is a Hurwitz or stability matrix (i.e., all the eigenvalues are in the open left-half complex plane).

Define the state error vector and wavelet coefficient error vectors as

$$\mathbf{e}_x = \mathbf{x} - \hat{\mathbf{x}} \tag{27}$$

$$\tilde{\mathbf{A}}_{j_0\mathbf{k}} = \mathbf{A}^*_{j_0\mathbf{k}} - \mathbf{A}_{j_0\mathbf{k}} \tag{28}$$

$$\tilde{\mathbf{B}}^{(i)}_{j\mathbf{k}} = \mathbf{B}^{(i)*}_{jk} - \mathbf{B}^{(i)}_{j\mathbf{k}} \tag{29}$$

so that the dynamical expression of the state error is given by

$$\dot{\mathbf{e}}_x = \mathbf{A}\mathbf{e}_x + \sum_{\mathbf{k}\in\mathcal{A}_{j_0}} \tilde{\mathbf{A}}_{j_0\mathbf{k}}\Phi_{j_0\mathbf{k}}(\mathbf{x},\mathbf{u}) + \sum_{j=j_0}^{N}\sum_{\mathbf{k}\in\mathcal{B}_j}\sum_{i=1}^{2^n-1} \tilde{\mathbf{B}}^{(i)}_{j\mathbf{k}}\Psi^{(i)}_{j\mathbf{k}}(\mathbf{x},\mathbf{u}) + \varepsilon_N \tag{30}$$

Consider the Lyapunov function

$$V = \mathbf{e}_x^T\mathbf{P}\mathbf{e}_x + \frac{1}{\alpha_{j_0}}\sum_{\mathbf{k}\in\mathcal{A}_{j_0}} \tilde{\mathbf{A}}^T_{j_0\mathbf{k}}\tilde{\mathbf{A}}_{j_0\mathbf{k}} + \frac{1}{\beta_j}\sum_{j=j_0}^{N}\sum_{\mathbf{k}\in\mathcal{B}_j}\sum_{i=1}^{2^n-1} (\tilde{\mathbf{B}}^{(i)}_{j\mathbf{k}})^T\tilde{\mathbf{B}}^{(i)}_{j\mathbf{k}} \tag{31}$$

where $\mathbf{P} = \{p_{ij}\} \in \mathbb{R}^{d\times d}$ is chosen to be a positive definite matrix so that the matrix $\mathbf{Q} = -\mathbf{P}\mathbf{A} - \mathbf{A}^T\mathbf{P}$ is also a positive definite matrix, and α_{j_0} and β_j are positive constants which will appear in the parameter adaptation laws, also referred to as the adaptation rates.

The first derivative of the Lyapunov function V with respect to time t is

$$\dot{V} = \dot{\mathbf{e}}_x^T \mathbf{P}\mathbf{e}_x + \mathbf{e}_x^T \mathbf{P}\dot{\mathbf{e}}_x + \frac{2}{\alpha_{j_0}} \sum_{\mathbf{k}\in\mathcal{A}_{j_0}} \dot{\tilde{\mathbf{A}}}_{j_0\mathbf{k}}^T \tilde{\mathbf{A}}_{j_0\mathbf{k}} + \frac{2}{\beta_j} \sum_{j=j_0}^{N} \sum_{\mathbf{k}\in\mathcal{B}_j} \sum_{i=1}^{2^n-1} (\dot{\tilde{\mathbf{B}}}_{j\mathbf{k}}^{(i)})^T \tilde{\mathbf{B}}_{j\mathbf{k}}^{(i)} \tag{32}$$

Substituting (30) into (32) gives

$$\begin{aligned}\dot{V} = -\mathbf{e}_x^T \mathbf{Q}\mathbf{e}_x &+ \frac{2}{\alpha_{j_0}} \sum_{\mathbf{k}\in\mathcal{A}_{j_0}} \left(\dot{\tilde{\mathbf{A}}}_{j_0\mathbf{k}}^T \tilde{\mathbf{A}}_{j_0\mathbf{k}} + \alpha_{j_0} \mathbf{e}_x^T \mathbf{P} \tilde{\mathbf{A}}_{j_0\mathbf{k}} \Phi_{j_0\mathbf{k}}(\mathbf{x},\mathbf{u}) \right) \\ &+ \frac{2}{\beta_j} \sum_{j=j_0}^{N} \sum_{\mathbf{k}\in\mathcal{B}_j} \sum_{i=1}^{2^n-1} \left((\dot{\tilde{\mathbf{B}}}_{j\mathbf{k}}^{(i)})^T \tilde{\mathbf{B}}_{j\mathbf{k}}^{(i)} + \beta_j \mathbf{e}_x^T \mathbf{P} \tilde{\mathbf{B}}_{j\mathbf{k}}^{(i)} \Psi_{j\mathbf{k}}^{(i)}(\mathbf{x},\mathbf{u}) \right) + 2\mathbf{e}_x^T \mathbf{P} \varepsilon_N \end{aligned} \tag{33}$$

Since $\mathbf{A}_{j_0\mathbf{k}}^*$ and $\mathbf{B}_{j\mathbf{k}}^{(i)*}$ are constant vectors, $\dot{\tilde{\mathbf{A}}}_{j_0\mathbf{k}} = -\dot{\mathbf{A}}_{j_0\mathbf{k}}$ and $\dot{\tilde{\mathbf{B}}}_{j\mathbf{k}}^{(i)} = -\dot{\mathbf{B}}_{j\mathbf{k}}^{(i)}$. If there is no modelling error, i.e., $\varepsilon_N = 0$, $\mathbf{A}_{j_0\mathbf{k}}$ and $\mathbf{B}_{j\mathbf{k}}^{(i)}$ can simply be estimated by the following adaptation laws:

$$\dot{\mathbf{A}}_{j_0\mathbf{k}} = \alpha_{j_0} \mathbf{e}_x^T \mathbf{P} \dot{\Phi}_{j_0\mathbf{k}}(\mathbf{x},\mathbf{u}) \tag{34}$$

$$\dot{\mathbf{B}}_{j\mathbf{k}}^{(i)} = \beta_j \mathbf{e}_x^T \mathbf{P} \Psi_{j\mathbf{k}}^{(i)}(\mathbf{x},\mathbf{u}) \tag{35}$$

In the presence of a modelling error ε_N, several algorithms can be applied to ensure the stability of the whole identification scheme, e.g., the fixed or switching σ-modification [15], ε-modification [25] and the dead-zone methods [28].

Define the following sets:

$$\mathcal{F}^-(\Gamma, M) = \{\mathbf{z} : \|\mathbf{z}\| < M \quad \text{or} \quad (\|\mathbf{z}\| = M \quad \text{and} \quad \mathbf{e}_x^T \mathbf{P}\mathbf{z}\Gamma \geq 0)\} \tag{36}$$

$$\mathcal{F}^+(\Gamma, M) = \{\mathbf{z} : \|\mathbf{z}\| = M \quad \text{and} \quad \mathbf{e}_x^T \mathbf{P}\mathbf{z}\Gamma < 0\} \tag{37}$$

where $\mathbf{z} \in \mathbb{R}^d$, Γ is a function and M is a positive constant. Here, in order to avoid parameter drift in the presence of modelling error, the application of the projection algorithm [11] gives the following adaptive laws for the parameter estimates $\mathbf{A}_{j_0\mathbf{k}}$ and $\mathbf{B}_{j\mathbf{k}}^{(i)}$:

$$\dot{\mathbf{A}}_{j_0\mathbf{k}} = \begin{cases} \alpha_{j_0} \mathbf{e}_x^T \mathbf{P} \Phi_{j_0\mathbf{k}} & \text{if} \quad \mathbf{A}_{j_0\mathbf{k}} \in \mathcal{F}^-(\Phi_{j_0\mathbf{k}}, M_{j_0\mathbf{k}}) \\ \alpha_{j_0} \mathbf{e}_x^T \mathbf{P} \Phi_{j_0\mathbf{k}} + \alpha_{j_0} M_{j_0\mathbf{k}}^{-2} \mathbf{e}_x^T \mathbf{P} \mathbf{A}_{j_0\mathbf{k}} \Phi_{j_0\mathbf{k}} \mathbf{A}_{j_0\mathbf{k}} & \text{if} \quad \mathbf{A}_{j_0\mathbf{k}} \in \mathcal{F}^+(\Phi_{j_0\mathbf{k}}, M_{j_0\mathbf{k}}) \end{cases} \tag{38}$$

$$\dot{\mathbf{B}}_{j\mathbf{k}}^{(i)} = \begin{cases} \beta_j \mathbf{e}_x^T \mathbf{P} \Psi_{j\mathbf{k}}^{(i)} & \text{if} \quad \mathbf{B}_{j\mathbf{k}}^{(i)} \in \mathcal{F}^-(\Psi_{j\mathbf{k}}^{(i)}, M_{j\mathbf{k}}^{(i)}) \\ \beta_j \mathbf{e}_x^T \mathbf{P} \Psi_{j\mathbf{k}}^{(i)} + \beta_j (M_{j\mathbf{k}}^{(i)})^{-2} \mathbf{e}_x^T \mathbf{P} \mathbf{B}_{j\mathbf{k}}^{(i)} \Psi_{j\mathbf{k}}^{(i)} \mathbf{B}_{j\mathbf{k}}^{(i)} & \text{if} \quad \mathbf{B}_{j\mathbf{k}}^{(i)} \in \mathcal{F}^+(\Psi_{j\mathbf{k}}^{(i)}, M_{j\mathbf{k}}^{(i)}) \end{cases} \tag{39}$$

where $M_{j_0\mathbf{k}}$, $M_{j\mathbf{k}}^{(i)}$ are the allowed largest values of $\|\mathbf{A}_{j_0\mathbf{k}}\|$ and $\|\mathbf{B}_{j\mathbf{k}}^{(i)}\|$, respectively. It is clear that if the initial parameter vectors are chosen such that $\mathbf{A}_{j_0\mathbf{k}}(0) \in \mathcal{F}^-(\Phi_{j_0\mathbf{k}}, M_{j_0\mathbf{k}}) \bigcup \mathcal{F}^+(\Phi_{j_0\mathbf{k}}, M_{j_0\mathbf{k}})$ and $\mathbf{B}_{j\mathbf{k}}^{(i)}(0) \in \mathcal{F}^-(\Psi_{j\mathbf{k}}^{(i)}, M_{j\mathbf{k}}^{(i)}) \bigcup \mathcal{F}^+(\Psi_{j\mathbf{k}}^{(i)}, M_{j\mathbf{k}}^{(i)})$, then the vectors $\mathbf{A}_{j_0\mathbf{k}}$ and

$\mathbf{B}_{j\mathbf{k}}^{(i)}$ are confined to the sets $\mathcal{F}^-(\Phi_{j_0\mathbf{k}}, M_{j_0\mathbf{k}}) \bigcup \mathcal{F}^+(\Phi_{j_0\mathbf{k}}, M_{j_0\mathbf{k}})$ and $\mathcal{F}^-(\Psi_{j\mathbf{k}}^{(i)}, M_{j\mathbf{k}}^{(i)}) \bigcup \mathcal{F}^+(\Psi_{j\mathbf{k}}^{(i)}, M_{j\mathbf{k}}^{(i)})$, respectively. Using the adaptive laws (38) and (39), (33) becomes

$$\begin{aligned} \dot{V} &\leq -\mathbf{e}_x^T \mathbf{Q} \mathbf{e}_x + 2 \sum_{i=1}^{d} \sum_{j=1}^{d} |p_{ij}||e_{xi}||\varepsilon_{Nj}| \\ &\leq -\mathbf{e}_x^T \mathbf{Q} \mathbf{e}_x + 2 \sum_{i=1}^{d} \sum_{j=1}^{d} |p_{ij}||e_{xi}|\bar{\varepsilon}_N \end{aligned} \tag{40}$$

where $\mathbf{e}_x = [e_{x1}, e_{x2}, ..., e_{xd}]^T$.

For the sake of simplicity, the positive definite matrix $\mathbf{Q}$ is assumed to be diagonal, i.e., $\mathbf{Q} = \text{diag}[q_1, q_2, ..., q_d]$. Also define

$$\Theta(\zeta) = \left\{ \mathbf{e}_x : \sum_{i=1}^{d} q_i \left(|e_{xi}| - \sum_{j=1}^{d} \frac{|p_{ij}|}{q_i} \zeta \right)^2 \leq \sum_{i=1}^{d} \left(\sum_{j=1}^{d} p_{ij} \right)^2 \frac{\zeta^2}{q_i} \right\} \tag{41}$$

where ζ is a positive variable, i.e., $\zeta \geq 0$.

If there is no modelling error (*i.e.*, $\bar{\varepsilon}_N = 0$), it is clear from Eq.(40) that $\dot{V}$ is negative semidefinite. Hence the stability of the overall identification scheme is guaranteed and $\mathbf{e}_x \to 0$, $\mathbf{A}_{j_0\mathbf{k}} \to 0$, $\mathbf{B}_{j\mathbf{k}}^{(i)} \to 0$. In the presence of modelling errors, if $\mathbf{e}_x \notin \Theta(\bar{\varepsilon}_N)$, it is easy to show from (40) that $\dot{V}$ is still negative and the state error $\mathbf{e}_x$ will converge to the set $\Theta(\bar{\varepsilon}_N)$. But, if $\mathbf{e}_x \in \Theta(\bar{\varepsilon}_N)$, it is possible that $\dot{V} > 0$, which implies that the weight vectors $\mathbf{A}_{j_0\mathbf{k}}$ and $\mathbf{B}_{j\mathbf{k}}^{(i)}$ may drift to infinity over time. The adaptive laws (38) and (39) avoid this drift by limiting the upper bounds of the parameters. Thus the state error $\mathbf{e}_x$ always converges to the set $\Theta(\bar{\varepsilon}_N)$ and the whole identification scheme will remain stable in the case of modelling errors.

4 Identification Using Variable Wavelet Networks

For nonlinear systems, the system operation can change with time. This will result in an estimation error for the fixed wavelet network that is beyond the required error. In order to improve the identification performance, both the structure and the parameters of the wavelet network model needs to be modified in response to variations of the plant characteristics. This section takes into account the modification of the wavelet network structure and the adaptation of the parameters.

It is known that the modelling error ε_N can be reduced arbitrarily by increasing the resolution of the wavelet network. But, generally when the resolution is increased beyond a certain value the modelling error ε_N will improve very little by any further increasing the resolution. This will also result in a large size network even for simple nonlinear problems and in practice, this is not realistic. In most cases, the required modelling error can be given by considering the design requirements and specifications of the system. Thus, the problem

now is to find a suitable sized network to achieve the required modelling error. Following variable neural networks [20], a variable wavelet network is introduced below.

Generally speaking, a variable wavelet network has the property that the number of wavelons in the network can be either increased or decreased over time according to a design strategy. For the problem of nonlinear modelling, the variable wavelet network is initialised with a small number of wavelons. As observations are received, the network grows by adding new wavelons or is pruned by removing old ones.

According to the multiresolution approximation theory, increasing the resolution of the network will improve the approximation. To improve the approximation accuracy, the growing network technique [16] [19] is applied. This means that the wavelets at a higher resolution need to be added to the network. Here it is assumed that at the resolution 2^N the approximation of the function $\mathbf{f}$ by the wavelet network is denoted as $\hat{\mathbf{f}}^{(N)}$. Based on the growing network technique and the structure of the function $\hat{\mathbf{f}}$ in (21), the adding operation is defined as

$$\hat{\mathbf{f}}(\mathbf{x},\mathbf{u}) = \hat{\mathbf{f}}^{(N)}(\mathbf{x},\mathbf{u}) \dot{\oplus} \sum_{\mathbf{k}\in\mathcal{B}_N} \sum_{i=1}^{2^n-1} \mathbf{B}_{N\mathbf{k}}^{(i)} \Psi_{N\mathbf{k}}^{(i)}(\mathbf{x},\mathbf{u}) \tag{42}$$

where $\dot{\oplus}$ denotes the adding operation. Eq.(42) means that wavelets at the resolution 2^{N+1} are added to the network. To add new wavelons to the network the following two conditions must be satisfied: a) The modelling error must be greater than the required accuracy. b) The period between the two adding operations must be greater than the minimum response time of the adding operation.

The removing operation is defined as

$$\hat{\mathbf{f}}(\mathbf{x},\mathbf{u}) = \hat{\mathbf{f}}^{(N)}(\mathbf{x},\mathbf{u}) \dot{\ominus} \sum_{\mathbf{k}\in\mathcal{B}_{N-1}} \sum_{i=1}^{2^n-1} \mathbf{B}_{(N-1)\mathbf{k}}^{(i)} \Psi_{(N-1)\mathbf{k}}^{(i)}(\mathbf{x},\mathbf{u}) \tag{43}$$

where $\dot{\ominus}$ denotes the removing operation. Eq.(43) implies that wavelets at the resolution 2^N are removed from the network. Similarly, to remove some old wavelons from the network, the following two conditions must be satisfied: a) The modelling error must be less than the required accuracy. b) The period between the two removing operations must be greater than the minimum response time of the removing operation.

In both the adding and the removing operations, condition a) means that the change of the modelling error in the network must be significant. Condition b) says the minimum response time of each operation must be considered.

From the set $\Theta(\bar{\varepsilon}_N)$ which gives a relationship between the state error $\mathbf{e}_x$ and the modeling errors ε_N, it can be shown that the state error depends on the modelling error. If the upper bound $\bar{\varepsilon}_N$ of the modelling error is known, then the set $\Theta(\bar{\varepsilon}_N)$ to which the state error will converge is also known. However, in most cases the upper bound $\bar{\varepsilon}_N$ is unknown.

In practice, systems are usually required to keep the state errors within prescribed bounds, that is,

$$|e_{xi}| \leq \bar{e}_i, \quad \text{for} \quad i = 1, 2, ..., n, \quad t \in \mathbb{R}^+ \tag{44}$$

where $\bar{e}_i$ is the required accuracy. At the beginning, it is very difficult to know how many wavelons are needed to achieve the above identification requirements. In order to find a suitable sized network for this identification problem, lower and upper bounds are set for the state errors which are functions of time t. A variable network such that

$$|e_{xi}| \in [\eta_i^L(t), \quad \eta_i^U(t) + \bar{e}_i], \quad \text{for} \quad i = 1, 2, ..., n \tag{45}$$

is then tried, where $\eta_i^L(t), \eta_i^U(t)$ are monodecreasing functions of time t, respectively. For example,

$$\eta_i^U(t) = e^{-\beta_U t}\eta_i^U(0) \tag{46}$$

$$\eta_i^L(t) = e^{-\beta_L t}\eta_i^L(0) \tag{47}$$

where β_U, β_L are positive constants, $\eta_i^U(0), \eta_i^L(0)$ are the initial values. It is clear that $\eta_i^U(t), \eta_i^L(t)$ decrease with time t. As $t \to 0$, $\eta_i^U(t), \eta_i^L(t)$ approach 0. Thus, in this way the state errors reach the required accuracies given in (44).

From the relationship between the modelling error and the state error and given the lower and upper bounds $\eta_i^U(t), \eta_i^L(t) + \bar{e}_i$ of the state errors the corresponding modelling error should be

$$\bar{\varepsilon}_N(t) \in [\varepsilon_L(t), \quad \varepsilon_U(t)] \tag{48}$$

From (41) the area that the set $\Theta(\zeta)$ covers is a hyperellipsoid with the center

$$\left(\sum_{j=1}^{d} \frac{|p_{1j}|}{q_1}\zeta, \quad \sum_{j=1}^{d} \frac{|p_{2j}|}{q_2}\zeta, \quad ..., \quad \sum_{j=1}^{d} \frac{|p_{dj}|}{q_n}\zeta \right). \tag{49}$$

It can also be deduced from the set $\Theta(\bar{\varepsilon}_N(t))$ that the upper bound $\varepsilon_U(t)$ and the lower bound $\varepsilon_L(t)$ are given by

$$\varepsilon_L(t) = \min_{i=1,2,...,d} \left\{ \left(\sum_{j=1}^{d} \frac{|p_{ij}|}{q_i} + \left(\sum_{k=1}^{d} \left(\sum_{j=1}^{d} p_{kj} \right)^2 \frac{1}{q_i q_k} \right)^{1/2} \right)^{-1} \eta_i^L(t) \right\} \tag{50}$$

$$\varepsilon_U(t) = \max_{i=1,2,...,d} \left\{ \left(\sum_{j=1}^{d} \frac{|p_{ij}|}{q_i} + \left(\sum_{k=1}^{d} \left(\sum_{j=1}^{d} p_{kj} \right)^2 \frac{1}{q_i q_k} \right)^{1/2} \right)^{-1} (\eta_i^U(t) + \bar{e}_i) \right\} \tag{51}$$

Thus, given the upper and lower bounds of the state error, the corresponding values for the modelling error can be estimated by (50) and (51).

To smooth the identification performance when the adding and removing operations are used, the decomposition and reconstruction algorithms of a multiresolution decomposition are

applied to the initial calculation of the wavelet coefficients. Here two important relations are introduced. Firstly, since the family $\{\Phi_{\mathbf{k}}\}$ spans V_0, then $\{\Phi(2\mathbf{x}-\mathbf{k})\}$ spans the next filter scale $V_1 = V_0 \oplus W_0$. Both the scaling function and the wavelet function can be expressed in terms of the scaling function at the resolution $2^j = 2^1$, i.e.,

$$\Phi(\mathbf{x}) = \sqrt{2}\sum_{\mathbf{k}\in\mathbb{Z}} c_{\mathbf{k}}\Phi(2\mathbf{x}-\mathbf{k}) \tag{52}$$

$$\Psi_i(\mathbf{x}) = \sqrt{2}\sum_{\mathbf{k}\in\mathbb{Z}} d_{\mathbf{k}}^{(i)}\Phi(2\mathbf{x}-\mathbf{k}) \tag{53}$$

where $c_{\mathbf{k}}$ and $d_{\mathbf{k}}^{(i)}$ are known as the two scale reconstruction sequences. Secondly, any scaling function $\Phi(2\mathbf{x})$ in V_1 can alternatively be written using the scaling function $\Phi(\mathbf{x})$ in V_0 and wavelet function $\Psi(\mathbf{x})$ in W_0 as

$$\Phi(2\mathbf{x}-\mathbf{l}) = \sum_{\mathbf{k}\in\mathbb{Z}}\left(a_{\mathbf{l}-2\mathbf{k}}\Phi(\mathbf{x}-\mathbf{k}) + \sum_{i=1}^{2^n-1} b_{\mathbf{l}-2\mathbf{k}}^{(i)}\Psi_i(\mathbf{x}-\mathbf{k})\right) \tag{54}$$

where $a_{\mathbf{k}}$ and $b_{\mathbf{k}}^{(i)}$ are known as the decomposition sequences, and $\mathbf{l}\in\mathbb{Z}^n$.

In addition, in terms of multiresolution decompositions, the approximation of the function $\mathbf{f}(\mathbf{x},\mathbf{u})$ at the resolution 2^j can be written as

$$\hat{\mathbf{f}}^{(j)}(\mathbf{x},\mathbf{u}) = \sum_{\mathbf{k}\in\mathcal{A}_{j-1}} \mathbf{A}_{(j-1)\mathbf{k}}\Phi_{(j-1)\mathbf{k}}(\mathbf{x},\mathbf{u}) + \sum_{\mathbf{k}\in\mathcal{B}_{j-1}}\sum_{i=1}^{2^n-1} \mathbf{B}_{(j-1)\mathbf{k}}^{(i)}\Psi_{(j-1)\mathbf{k}}^{(i)}(\mathbf{x},\mathbf{u}) \tag{55}$$

where

$$\begin{aligned}\sum_{\mathbf{k}\in\mathcal{A}_{j-1}} \mathbf{A}_{(j-1)\mathbf{k}}\Phi_{(j-1)\mathbf{k}}(\mathbf{x},\mathbf{u}) \;=\; &\sum_{\mathbf{k}\in\mathcal{A}_{j-2}} \mathbf{A}_{(j-2)\mathbf{k}}\Phi_{(j-2)\mathbf{k}}(\mathbf{x},\mathbf{u}) \\ &+ \sum_{\mathbf{k}\in\mathcal{B}_{j-2}}\sum_{i=1}^{2^n-1} \mathbf{B}_{(j-2)\mathbf{k}}^{(i)}\Psi_{(j-2)\mathbf{k}}^{(i)}(\mathbf{x},\mathbf{u})\end{aligned} \tag{56}$$

Hence, if the state error $\mathbf{e}_x \notin \Theta(\varepsilon_U(t))$, the network needs more wavelets. Add the wavelets at the resolution 2^{N+1} into the network. Following the adding operation (42) and the expression (55) of $\hat{\mathbf{f}}(\mathbf{x},\mathbf{u})^{(j)}$ at $j = N$, the structure of the approximated function $\hat{\mathbf{f}}(\mathbf{x},\mathbf{u})$ is of the form:

$$\hat{\mathbf{f}}(\mathbf{x},\mathbf{u}) = \sum_{\mathbf{k}\in\mathcal{A}_N} \mathbf{A}_{N\mathbf{k}}\Phi_{N\mathbf{k}}(\mathbf{x},\mathbf{u}) + \sum_{\mathbf{k}\in\mathcal{B}_N}\sum_{i=1}^{2^n-1} \mathbf{B}_{N\mathbf{k}}^{(i)}\Psi_{N\mathbf{k}}^{(i)}(\mathbf{x},\mathbf{u}) \tag{57}$$

The parameter vectors $\mathbf{A}_{N\mathbf{k}}$ and $\mathbf{B}_{N\mathbf{k}}^{(i)}$ are adapted by the laws (38) and (39). Using the sequences $c_{\mathbf{k}}$ and $d_{\mathbf{k}}^{(i)}$, the initial values after the adding operation are then given by the reconstruction algorithm [23]:

$$\mathbf{A}_{N\mathbf{k}}(0) = \sum_{\mathbf{l}\in\mathcal{A}_N}\left(c_{\mathbf{l}-2\mathbf{k}}\mathbf{A}_{(N-1)\mathbf{l}} + \sum_{i=1}^{2^n-1} d_{\mathbf{l}-2\mathbf{k}}^{(i)}\mathbf{B}_{(N-1)\mathbf{l}}^{(i)}\right) \tag{58}$$

$$\mathbf{B}_{N\mathbf{k}}^{(i)}(0) = 0 \tag{59}$$

where $\mathbf{A}_{(N-1)\mathbf{k}}$ and $\mathbf{B}^{(i)}_{(N-1)\mathbf{k}}$ are the estimated values before the adding operation.

If the state error $\mathbf{e}_x \in \Theta(\varepsilon_L(t))$, some wavelets need to be removed because the network may be overfitted. In this case remove the wavelets associated with the resolution 2^N. In terms of the removing operation (43) and the expression (55) of $\hat{\mathbf{f}}(\mathbf{x},\mathbf{u})^{(j)}$ at $j = N$, the structure of the approximated function $\hat{\mathbf{f}}(\mathbf{x},\mathbf{u})$ is of the following form:

$$\hat{\mathbf{f}}(\mathbf{x},\mathbf{u}) = \sum_{\mathbf{k}\in\mathcal{A}_{N-2}} \mathbf{A}_{(N-2)\mathbf{k}}\Phi_{(N-2)\mathbf{k}}(\mathbf{x},\mathbf{u}) + \sum_{\mathbf{k}\in\mathcal{B}_{N-2}} \sum_{i=1}^{2^n-1} \mathbf{B}^{(i)}_{(N-2)\mathbf{k}}\Psi^{(i)}_{(N-2)\mathbf{k}}(\mathbf{x},\mathbf{u}) \tag{60}$$

The adaptive laws for the parameters $\mathbf{A}_{(N-2)\mathbf{k}}$ and $\mathbf{B}^{(i)}_{(N-2)\mathbf{k}}$ are still given by (38) and (39). But, using the sequences $a_{\mathbf{k}}$ and $b^{(i)}_{\mathbf{k}}$, the initial values after the removing operation are then changed by the decomposition algorithm [23] as follows:

$$\mathbf{A}_{(N-2)\mathbf{k}}(0) = \sum_{\mathbf{l}\in\mathcal{A}_{N-2}} a_{\mathbf{l}-2\mathbf{k}}\mathbf{A}_{N\mathbf{l}} \tag{61}$$

$$\mathbf{B}^{(i)}_{(N-2)\mathbf{k}}(0) = \sum_{\mathbf{l}\in\mathcal{B}_{N-2}} b^{(i)}_{\mathbf{l}-2\mathbf{k}}\mathbf{A}_{N\mathbf{l}} \tag{62}$$

where $\mathbf{A}_{N\mathbf{k}}$ and $\mathbf{B}^{(i)}_{N\mathbf{k}}$ are the estimated values before the removing operation.

Clearly, in both the above cases, the adaptive laws of the parameters are still given in the form of (38) and (39), based on the above changed parameters. It also follows that the convergence area of the state error vector begins with $\Theta(\varepsilon_U(0)) - \Theta(\varepsilon_L(0))$ and ends with $\Theta(\bar{\varepsilon})$, where $\bar{\varepsilon} = \varepsilon_U(\infty)$.

The determination of the vector sets $\mathcal{A}_j$ and $\mathcal{B}_j$, for $j = j_0, ..., N$ is also important but simple. The basic rule for choosing these sets is to make sure $2^j\mathbf{x} - \mathbf{k}$, for $\mathbf{k} \in \mathcal{A}_j$ or $\mathcal{B}_j$, is not out of the valid range of the variables of the scaling function $\Phi(.)$ or $\Psi_i(.)$, respectively.

5 Identification Using B-Spline Wavelets

One of the main ingredients in wavelet network based identification is the structure of the scaling function and the formulation of the wavelet decomposition and reconstruction. For many applications, it will not be essential for the wavelets to be orthonormal. Relaxing the orthonormality condition results in nonorthogonal multiresolution approximations and provides a more flexible framework for function approximation. A typical choice of scaling functions would be B-splines. B-splines are piece-wise polynomial functions and have a good local properties. They are very simple to implement on the computer and can be made as smooth as desired. For these reasons, B-splines have been used widely in interpolation problems. In the present section therefore wavelets which use B-spline functions as scaling functions are discussed.

5.1 One-dimensional B-Spline Wavelets

For the sake of simplicity, the one-dimensional B-spline wavelets will be considered initially. The one-dimensional B-spline function of m-th order is defined by the following recursive algorithm [7] [30]

$$B_m(x) = \frac{x}{m-1} B_{m-1}(x) + \frac{m-x}{m-1} B_{m-1}(x-1) \tag{63}$$

for $m \in \mathbb{Z}^+ \setminus \{0,1\}$, where

$$B_1(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{otherwise} \end{cases} \tag{64}$$

Let the m-th B-spline function be the scaling function, that is,

$$\phi(x) = B_m(x) \tag{65}$$

Then both the scaling function and the wavelets can be expressed in terms of the scaling function at the resolution $2^j = 2^1$

$$\phi(x) = \sum_{k=0}^{m} c_k B_m(2x-k) \tag{66}$$

$$\psi(x) = \sum_{k=0}^{3m-2} d_k B_m(2x-k) \tag{67}$$

where the two scale reconstruction sequences c_k and d_k are given by [7]

$$c_k = 2^{1-m} \binom{m}{k} \tag{68}$$

$$d_k = (-1)^k 2^{1-m} \sum_{l=0}^{m} \binom{m}{l} B_{2m}(k+1-l) \tag{69}$$

Also, the relationship between the scaling functions $B_m(2x)$ and $B_m(x)$ and the wavelet $\psi(x)$ can be expressed as

$$B_m(2x-l) = \sum_k (a_{l-2k} B_m(x-k) + b_{l-2k}\psi(x-k)), \quad l \in \mathbb{Z} \tag{70}$$

where the decomposition sequences a_k and b_k are given by [7]

$$a_k = \frac{1}{2} g_{-k} \tag{71}$$

$$b_k = \frac{1}{2} h_{-k} \tag{72}$$

and g_k and h_k are determined from the rational functions

$$G(z) = \frac{1}{2} \sum_k g_k z^k = z^{-1} \left(\frac{1+z}{2}\right)^m \frac{E_{2m-1}(z)}{E_{2m-1}(z^2)} \tag{73}$$

$$H(z) = \frac{1}{2} \sum_k h_k z^k = -z^{-1} \left(\frac{1-z}{2}\right)^m \frac{(2m-1)!}{E_{2m-1}(z^2)} \tag{74}$$

and E_{2m-1} is the Euler-Frobenius polynomial of order $2m-1$.

It is clear from the above that based on the B-splines the scaling function $\phi(x)$ and the wavelet $\psi(x)$ are easily constructed using (65) and (67), and the reconstruction sequences c_k, d_k and the decomposition sequences a_k, b_k are readily calculated from (68), (69) (71) and (72), respectively.

5.2 n-dimensional B-Spline Wavelets

The one-dimensional B-spline wavelet case will now be extended to the n-dimensional B-spline wavelet case. The m-th B-spline function will still be the one-dimensional scaling function, i.e. $\phi(x) = B_m(x)$. Then from (10), the n-dimensional scaling function is given by

$$\Phi(\mathrm{x}) = \prod_{l=1}^{n} B_m(x_l) \tag{75}$$

Using the relations (65) and (66) gives

$$\Phi(\mathrm{x}) = \sum_{k_1=0}^{m} \sum_{k_2=0}^{m} \cdots \sum_{k_n=0}^{m} c_{k_1} c_{k_2} \ldots c_{k_n} \prod_{i=1}^{n} B_m(2x_i - k_i) \tag{76}$$

which leads to

$$\Phi(\mathrm{x}) = \sum_{\mathrm{k}} c_{\mathrm{k}} \Phi(2\mathrm{x} - \mathrm{k}) \tag{77}$$

where $c_{\mathrm{k}} = c_{k_1} c_{k_2} \ldots c_{k_n}$.

The wavelets $\Psi_i(\mathrm{x})$ are a combination of n functions from the function set $\{\phi(x_1), \phi(x_2), \ldots, \phi(x_n), \psi(x_1), \psi(x_2), \ldots, \psi(x_n)\}$. The wavelets $\Psi_i(\mathrm{x})$ can similarly be expressed as the n-dimensional scaling function $\Phi(\mathrm{x})$. For example, if $\Psi_2(\mathrm{x}) = \psi(x_1)\phi(x_2)\psi(x_3)\ldots\psi(x_n)$, then using (66) and (67) results in

$$\Psi_2(\mathrm{x}) = \sum_{k_1=0}^{3m-2} \sum_{k_2=0}^{m} \sum_{k_3=0}^{3m-2} \cdots \sum_{k_n=0}^{3m-2} d_{k_1} c_{k_2} d_{k_3} \ldots d_{k_n} \prod_{i=1}^{n} B_m(2x_i - k_i) \tag{78}$$

In this case,

$$\Psi_2(\mathrm{x}) = \sum_{\mathrm{k}} d_{\mathrm{k}}^{(2)} \Phi(2\mathrm{x} - \mathrm{k}) \tag{79}$$

where $d_{\mathrm{k}}^{(2)} = d_{k_1} c_{k_2} d_{k_3} \ldots d_{k_n}$. Thus, all sequences $\{d_{\mathrm{k}}^{(i)}\}$ can be calculated in the same way as $d_{\mathrm{k}}^{(2)}$.

With (70), the relationship between the scaling functions $\Phi(2\mathrm{x})$ and $\Phi(\mathrm{x})$ and wavelets $\Psi_i(\mathrm{x})$ can be expresed as

$$\Phi(2\mathrm{x} - \mathrm{l}) = \prod_{i=1}^{n} \sum_{k_i} (a_{l_i - 2k_i} B_m(x_i - k_i) + b_{l_i - 2k_i} \psi(x_i - k_i)), \quad \mathrm{l} \in \mathbb{Z}^n \tag{80}$$

which results in the following compact form

$$\Phi(2\mathrm{x}-\mathrm{l}) = \sum_{\mathrm{k}} \left(a_{\mathrm{l}-2\mathrm{k}} \Phi(\mathrm{x}-\mathrm{k}) + \sum_{i=1}^{2^n-1} b^{(i)}_{\mathrm{l}-2\mathrm{k}} \Psi_i(\mathrm{x}-\mathrm{k}) \right) \tag{81}$$

where a_{k} and b_{k} can simply be calculated using a_k and b_k.

It is clear from the above that the scaling function, the wavelet, the reconstruction sequences and the decomposition sequences for the n-dimensional case can be computed directly from the those obtained for the one-dimensional case. Therefore, the structure of the scaling function and the formulation of the wavelet decomposition and reconstruction for the wavelet network based identification are completed using B-splines.

6 Simulation

Consider a nonlinear system described by

$$\dot{x} = (u - xu - x)e^{-0.5(x^2+u^2)} \tag{82}$$

where the input $u = 0.5(\cos(1.2t)\sin(1.7t) + \exp(-\sin(t^4)))$. Since $n = 2$, we will need 2-D B-spline wavelets for the wavelet network to identify this nonlinear dynamical system. The fourth order B-splines were used as the scaling function. Thus, the 2-D scaling function is given by

$$\Phi(x,u) = B_4(x)B_4(u) \tag{83}$$

where $B_4(.)$ is the 4th-order B-spline, which is a piecewise cubic function. For $n = 2$, there are three 2-D mother wavelets expressed by

$$\Psi_1(x,u) = B_4(x)\psi(u) \tag{84}$$

$$\Psi_2(x,u) = \psi(x)B_4(u) \tag{85}$$

$$\Psi_3(x,u) = \psi(x)\psi(u) \tag{86}$$

where the one-dimensional mother wavelet $\psi(x)$ is

$$\psi(x) = \sum_{k=0}^{10} \sum_{l=0}^{4} \frac{(-1)^k}{8} \binom{4}{l} B_8(k+1-l)B_4(2x-k) \tag{87}$$

The 2-D scaling function $\Phi(x,u)$ and the three 2-D wavelets $\Psi_1(x,u), \Psi_2(x,u), \Psi_3(x,u)$ are shown in Figs.(1)-(4). The state x and the nonlinear function $f(x,u)$ (or the state derivative $\dot{x}$) are shown in Figs.(5) and (6), respectively. Wavelet networks at the resolutions 2^j, for $j = 0,1,2,3$ were used for the identification with 16, 81, 146 and 278 wavelons, respectively. The state errors and the modelling errors for different resolutions are shown in Figs.(7)-(14). All figures denoted as (b) are larger scale versions of the figures denoted as (a).

As expected, at the beginning of the identification larger state errors and modelling errors exist. After a while, these errors become smaller and smaller, and finally they converge to certain ranges. It is clear from the simulation results that the whole identification scheme is stable from the beginning to the end. It has also been shown that the state error and the modelling error decrease with an increase in the resolution of the wavelet networks. But, the state error and the modelling error are improved only slightly when the resolution becomes adequate. Thus, for nonlinear dynamical system identification using wavelet networks, a proper resolution should be chosen so as to achieve the desired practical identification requirements.

7 Conclusions

A wavelet network based identification scheme has been proposed for nonlinear dynamical systems. Two kinds of wavelet networks, fixed and variable wavelet networks, were studied and parameter adaptation laws were derived which achieve the required estimation accuracy for a suitable sized network. The parameters of the wavelet network were adjusted using laws developed by the Lyapunov synthesis approach. By combining wavelet networks with Lyapunov synthesis techniques, adaptive parameter laws were developed which guarantee the stability of the whole identification scheme and the convergence of both the network parameters and the state errors. The wavelet network identification scheme was realised using B-spline wavelets and the calculation of the decomposition and reconstruction sequences using variable wavelet networks was given. A simulated example was used to demonstrate the operation of the identification scheme.

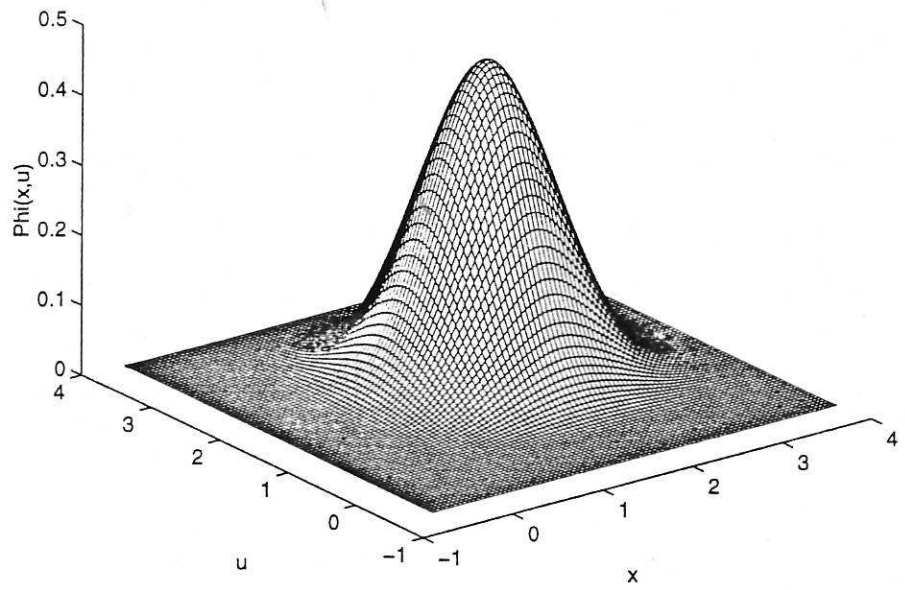

Figure 1: The scaling function $\Phi(x, u)$.

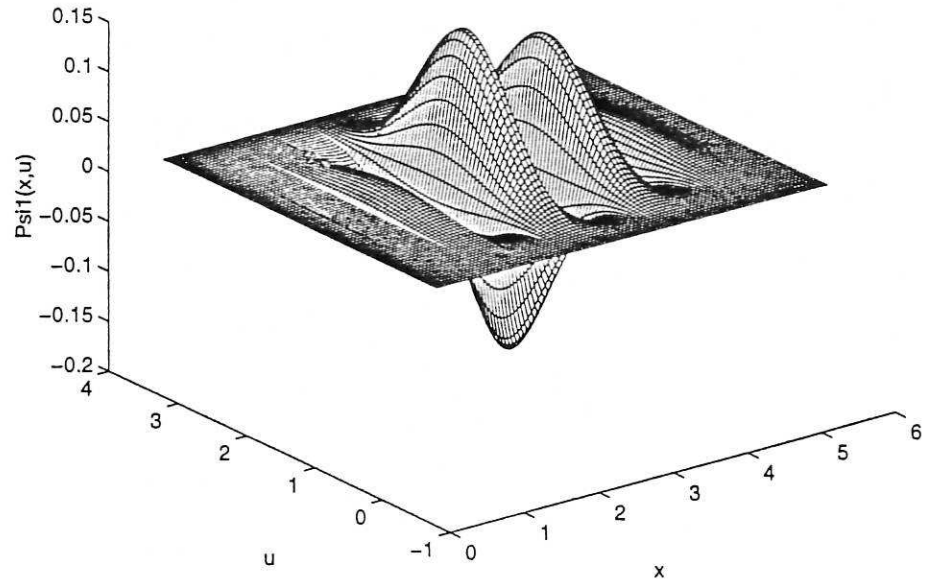

Figure 2: The wavelet function $\Psi_1(x, u)$.

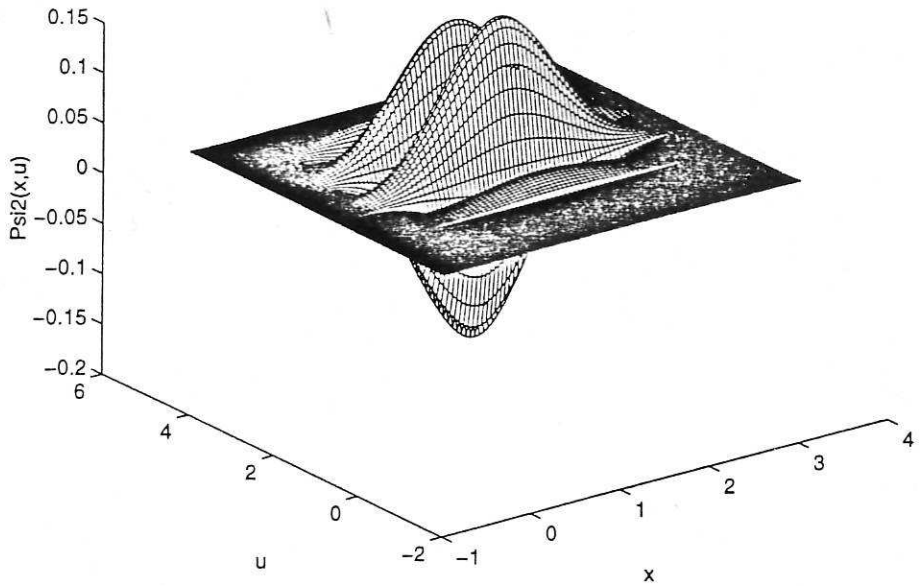

Figure 3: The wavelet function $\Psi_2(x, u)$.

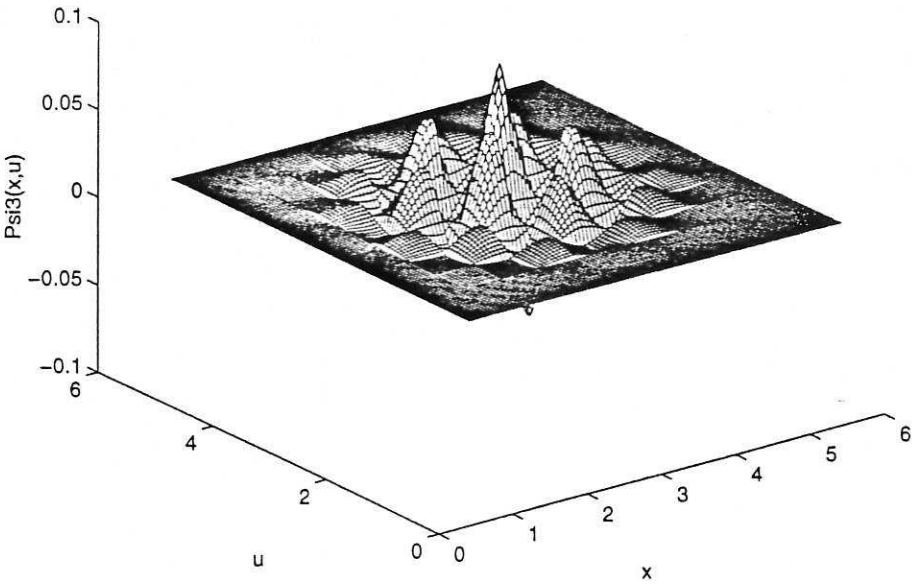

Figure 4: The wavelet function $\Psi_3(x, u)$.

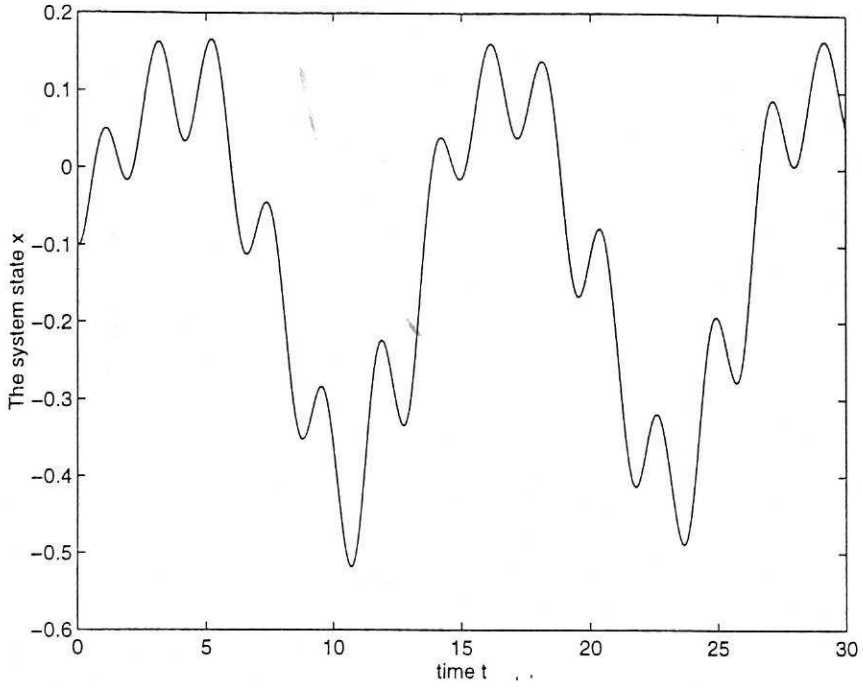

Figure 5: The system state $x(t)$.

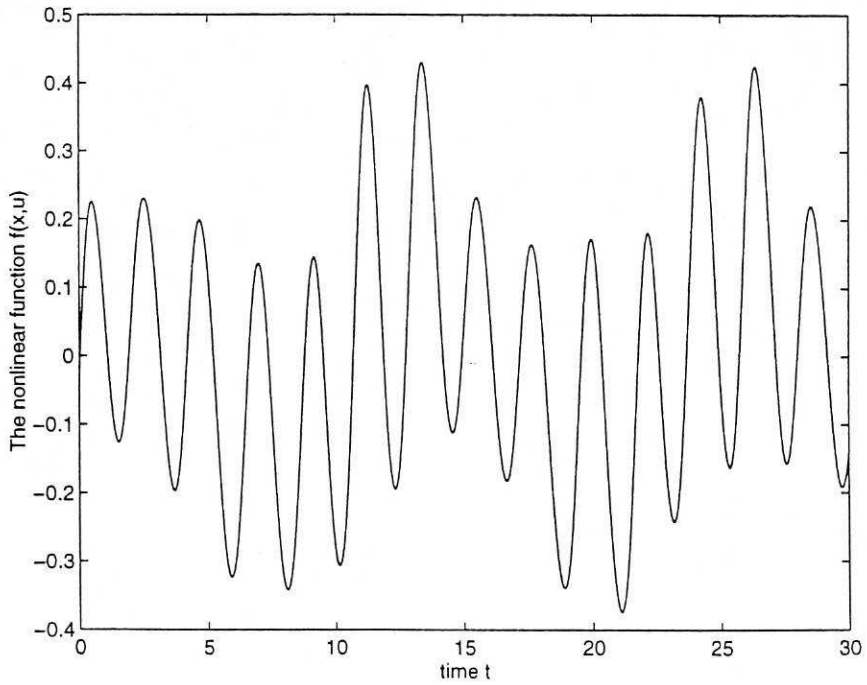

Figure 6: The nonlinear function $f(x, u)$ (or the state derivative $\dot{x}$).

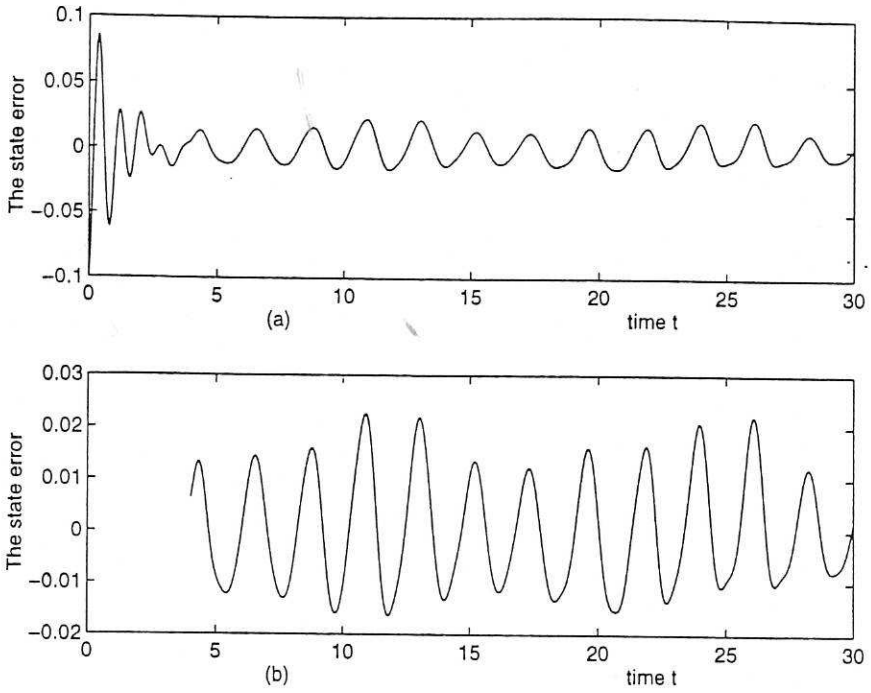

Figure 7: The state error $x(t) - \hat{x}(t)$ at the resolution 2^0.

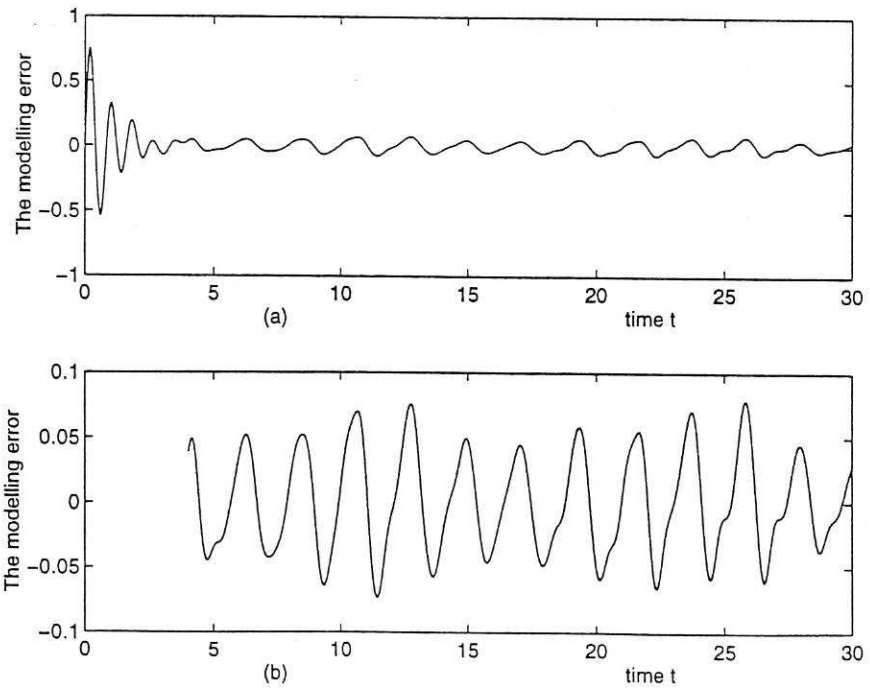

Figure 8: The modelling error $f(x,u) - \hat{f}(x,u)$ at the resolution 2^0.

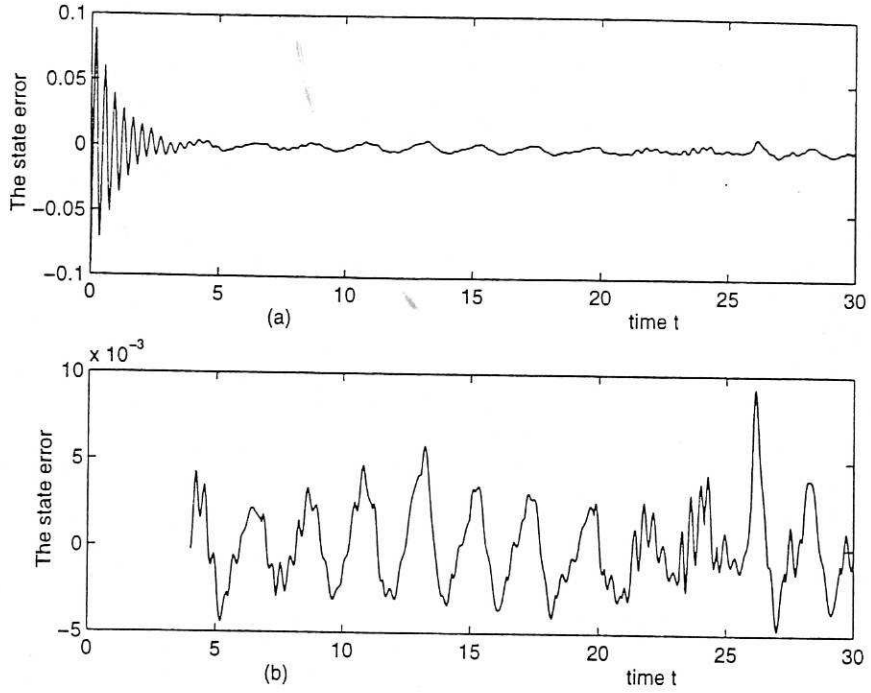

Figure 9: The state error $x(t) - \hat{x}(t)$ at the resolution 2^1.

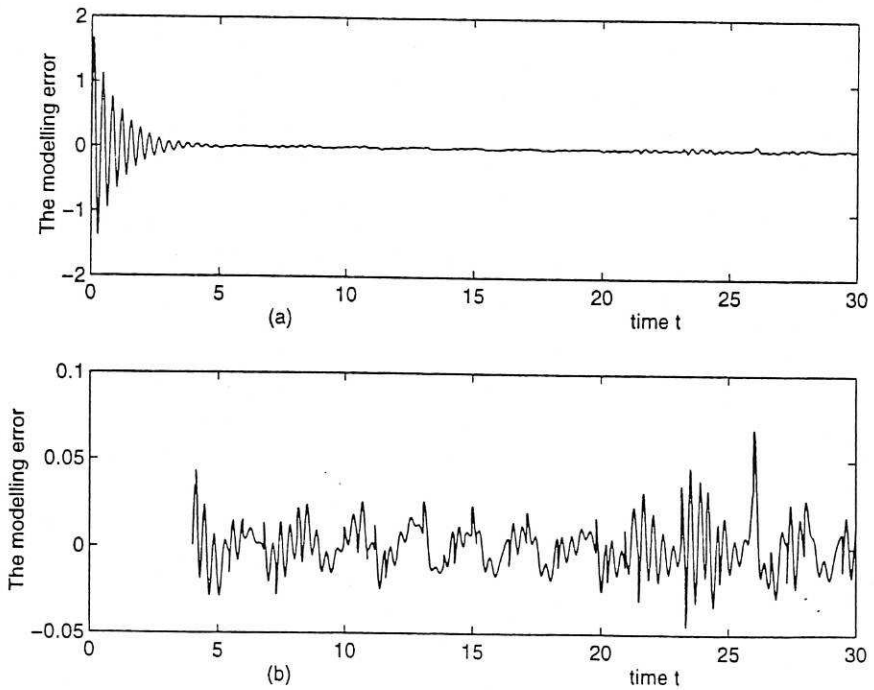

Figure 10: The modelling error $f(x,u) - \hat{f}(x,u)$ at the resolution 2^1.

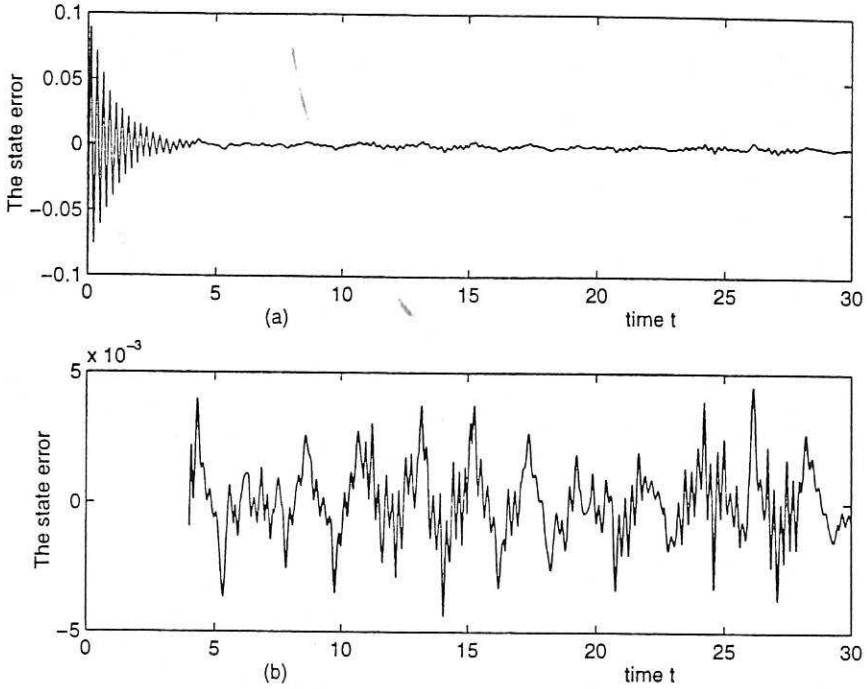

Figure 11: The state error $x(t) - \hat{x}(t)$ at the resolution 2^2.

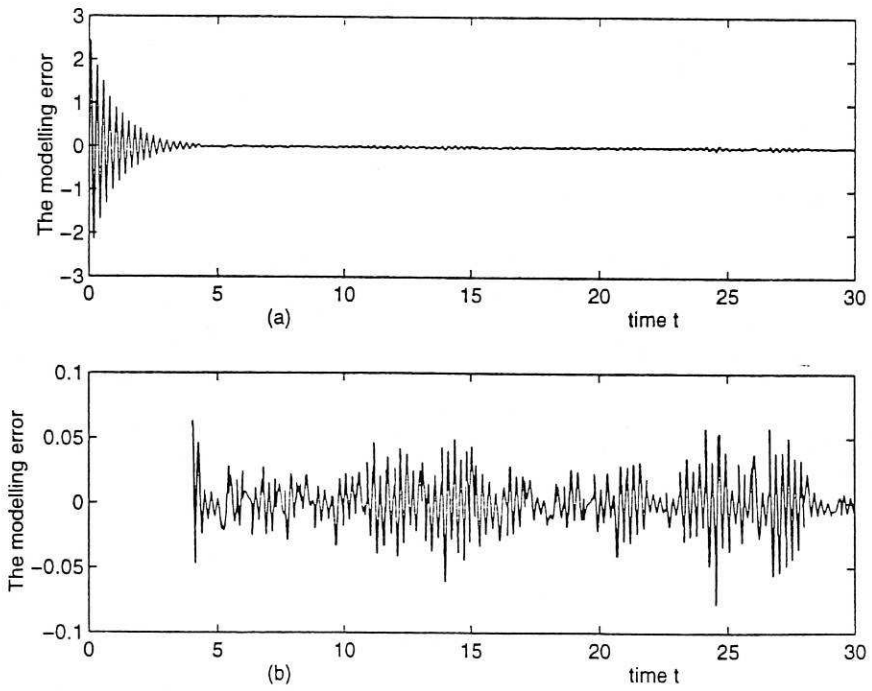

Figure 12: The modelling error $f(x,u) - \hat{f}(x,u)$ at the resolution 2^2.

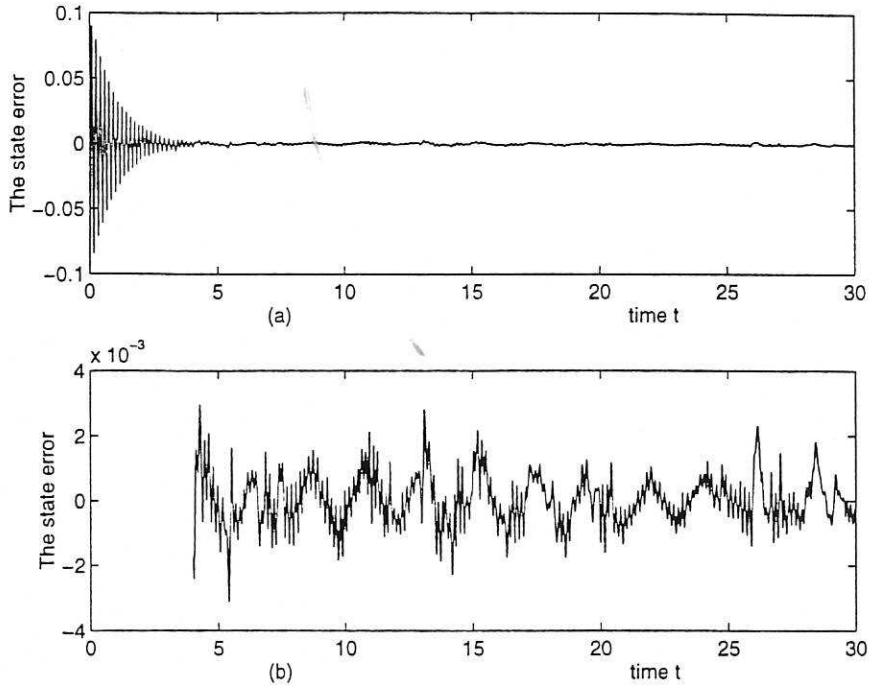

Figure 13: The state error $x(t) - \hat{x}(t)$ at the resolution 2^3.

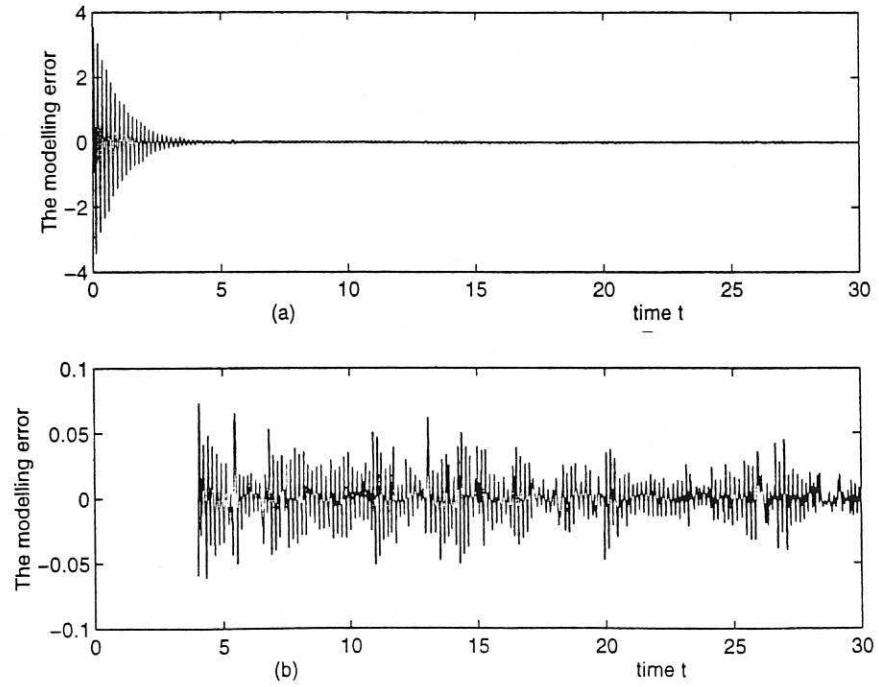

Figure 14: The modelling error $f(x,u) - \hat{f}(x,u)$ at the resolution 2^3.

Acknowledgements

The authors gratefully acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC) of UK under the contract GR/J46661.

References

[1] S.A. Billings, "Identification of nonlinear systems – a survey," *IEE Proceedings, Part D*, Vol. 127, pp.272-285, 1980.

[2] S.A. Billings and W.S.F. Voon, "A prediction-error and stepwise-regression estimation algorithm for non-linear systems," *International Journal of Control*, Vol.44, No.1, pp.803-822, 1986.

[3] L. Beriman, "Hinging hyperplanes for regression, classification and function approximation," IEEE Trans. Inf. Theory, vol. 39, pp. 999-1013, 1993.

[4] M. Brown and C.J Harris, Neurofuzzy Adaptive Modelling and Control, Prentice-Hall, New York, 1994.

[5] S. Chen, S. A. Billings and P. M. Grant, "Nonlinear system identification using neural networks", *International Journal of Control*, Vol.51, No.6, pp.1191-1214, 1990.

[6] S. Chen and J. Wigger, " A fast orthogonal least squares algorithm for efficient subset model selection," *IEEE Trans. on Signal Processing*, Vol. 43, No. 7, pp. 1713-1715, 1995.

[7] C.K. Chui, An Introduction to Wavelets, Academic Press, INC., 1992.

[8] D. Coca and S.A. Billings, "Continuous-time modelling of Chua's circuit from noisy measurements using wavelet decompositions," The Fourth International Workshopon Nonlinear Dynamics of electronic Systems, Seville, Spain, pp. 271-276, 1996.

[9] D. Coca and S.A. Billings, "Smoothing chaotic signals for system identification: a multiresolution wavelet decomposition approach," submitted for publication.

[10] I. Daubechies, "orthonormal bases of compactly supported wavelets," Commun. on Pure and Appl. Mathematics, vol.16, pp. 909-996, 1988.

[11] G. C. Goodwin and D. Q. Mayne, " A parameter estimation perspective of continuous time model reference adaptive control," *Automatica*, Vol. 23, No. 1, pp. 57-70, 1987.

[12] A. Grossmannand and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," SIAM J. Math. Anal. vol. 15, pp. 723-736, 1984.

[13] H. Haber and H. Unbehauen, "Structure identification of nonlinear dynamic systems – A survey on input/output approaches," *Automatica*, Vol.26, No.4, pp.651-677, 1990.

[14] Scanning the special issue on wavelets, The Proceedings of the IEEE, vol.84, no.4, 1996.

[15] P. A. Ioannou and P. V. Kokotovic, *Adaptive Systems with Reduced Models*, Springer-Verlag, New York, NY, 1983.

[16] V. Kadirkamanathan and M. Niranjan, "A function estimation approach to sequential learning with neural networks", *Neural Computation*, Vol.5, pp.954-957, 1993.

[17] M. Korenberg, S.A. Billings, Y.P. Liu and P.J. McIlroy, "Orthogonal parameter estimation algorithm for non-linear stochastic systems," *International Journal of Control*, Vol.48, No.1, pp.193-210, 1988.

[18] I.J. Leontaritis and S.A. Billings, "Model selection and validation methods for non-linear systems," *International Journal of Control*, Vol.45, No.3, pp.311-341, 1987.

[19] G. P. Liu, V. Kadirkamanathan and S.A. Billings, "Stable sequential identification of continuous nonlinear dynamical systems by growing RBF networks," International Journal of Control, vol. 65, no. 1, pp. 53-69, 1996.

[20] G. P. Liu, V. Kadirkamanathan and S.A. Billings, "Variable neural networks for nonlinear adaptive control," preprints of the 13th IFAC Congress, USA, vol. F, pp. 181-184, 1996.

[21] L. Ljung and T. Glad, Modelling of Dynamic Systems, Information and System Sciences Series, Prentice Hall, Englewood Cliffs, NJ., 1994.

[22] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.11, no.7, 1989.

[23] S.G. Mallat, "Multifrequency channel decompositions of images and wavelet models," IEEE Trans. Accoust., Speech, and signal process., vol. 37, pp.2091-2110, 1989.

[24] Y. Meyer, Wavelets and Operators, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1993.

[25] K. S. Narendra and A. M. Annaswamy, "A new adaptive law for robust adaptation with persistent excitation," *IEEE Trans. Aut. Control*, Vol. 32, No.2 pp. 134-145, 1987.

[26] K. S. Narendra and K. Parthasarathy. "Identification and control of dynamical systems using neural networks", *IEEE Trans. on Neural Networks*, Vol.1, No.1, pp.4-27, 1990.

[27] T. Poggio and F. Girosi, "Networks for approximation and learning", *Proceedings of the IEEE*, Vol. 78, No. 9, 1990.

[28] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence and Robustness*, Prentice-Hall, Engelwood Cliffs, 1989.

[29] J. Sjoberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson and A. Juditsky, "Nonlinear black-box modelling in system identification: a unified overview," Automatica, Vol.31, No.12, pp.1691-1724, 1995.

[30] H. Wang, G.P. Liu, C.J. Harris and M. Brown, Advanced Adaptive Control, *Pergamon Press Ltd*, Oxford, 1995.

[31] L. Zadeh, "Fuzzy logic, neural networks, and soft computing," Commun. ACM, vol. 37, pp.77-86, 1994.

[32] Q. Zhang and A. Benveniste, "Wavelet Networks," IEEE Trans. on Neural Networks, vol. 3, no. 6, pp. 889-898, 1992.