STABLE SEQUENTIAL IDENTIFICATION OF CONTINUOUS NONLINEAR DYNAMICAL SYSTEMS BY GROWING RBF NETWORKS

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

Department of Automatic Control & Systems Engineering
University of Sheffield
PO Box 600, Mappin Street, Sheffield S1 4DU, UK
Email: guoping/visakan@acse.sheffield.ac.uk

Research Report No. **547**

Department of Automatic Control & Systems Engineering
The University of Sheffield

November 1994

STABLE SEQUENTIAL IDENTIFICATION OF CONTINUOUS NONLINEAR DYNAMICAL SYSTEMS BY GROWING RBF NETWORKS

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

Department of Automatic Control and Systems Engineering
University of Sheffield, UK.

Abstract

This paper presents a sequential identification scheme for continuous nonlinear dynamical systems using neural networks. The nonlinearities of the dynamical systems are assumed to be unknown. The identification model is a Gaussian radial basis function neural network that grows gradually to span the appropriate state-space and of sufficient complexity to provide an approximation to the dynamical system. The sequential identification algorithm for continuous dynamical nonlinear systems is developed in the continuous-time framework instead of in discrete-time. The approach, different from the conventional methods of optimizing a cost function, attempts to ensure stability of the overall system while the neural network learns the system dynamics. The stability and convergence of the overall identification scheme is guaranteed by parameter adjustment laws developed using the Lyapunov synthesis approach. To ensure the modelling error can be reduced arbitrarily, a one-to-one mapping is proposed so that the states and inputs of the system are transferred into compact sets. The operation of the sequential identification scheme is illustrated using simulated experimental results.

1 Introduction

The identification of nonlinear systems using neural networks has become a widely studied research area in recent years. System identification mainly consists of two steps: the first is to choose an appropriate identification model and the second is to adjust the parameters of the model according to some adaptive laws so that the response of the model to an input signal can approximate the response of the real system to the same input. Since neural networks have good approximation capabilities and inherent adaptivity features, they provide a powerful tool for identification of systems with unknown nonlinearities [1], [17].

The application of neural network architectures to nonlinear system identification has been demonstrated by several studies in discrete time (see, for example, [3], [6], [5], [13], [15], [21], [29], [36]) and in continuous time [26], [30], [31]. For the most part, much of the studies in discrete-time systems are based on first replacing unknown functions in the difference equation by static neural networks and then deriving update laws using optimization methods (*eg.*, gradient descent/ascent methods) for a cost function (quadratic in general), which has lead to various back-propagation-type algorithms [22], [34], [35]. Though such schemes perform well in many cases, in general, some problems arise, such as the stability of the overall identification scheme and convergence of the output error. Alternative approaches based on the model reference adaptive control scheme [20], [32] have been developed [26], [30], [31] where the stability of the overall scheme is taken into consideration.

Much of the neural network based identification schemes view the problem as deriving model parameter adaptive laws, having chosen a structure for the neural network. However, choosing this structure such as the number of basis functions (hidden units in a single hidden layer) in the model must be done *a priori*. This can often lead to an over-determined or under-determined network structure which in turn leads to an identification model that is not optimal. In discrete-time formulation, some approaches have been developed in determining the number of hidden units (or basis functions) using decision theory [2] and model comparison methods such as *minimum description length* [33] and *Bayesian methods* [16]. The problem with these methods are that they require all observations to be available together and hence are not suitable for on-line or sequential identification task.

Yet another line of approach, developed for discrete-time systems, is to begin with a larger network prune, as in [19], or begin with a smaller network grow as in [8], [24] until the optimal network complexity is found. Amongst these dynamic structure models, the *resource allocating network* (RAN) developed by Platt [24] is an on-line or sequential identification algorithm. The RAN is essentially a growing Gaussian radial basis function (GRBF) network whose growth criteria and parameter adaptation laws have been studied and extended further by Kadirkamanathan [9], [10] and applied to time-series analysis [12] and pattern classification [11]. The RAN and its extensions addressed the identification of only autoregressive systems with no external inputs and hence stability was not an issue.

Due to some desirable features such as local adjustment of the weights and mathematical tractability, radial basis function networks (RBF) have recently attracted considerable attention. Their importance has also greatly benefited from the work of Moody and Darken [18] and, Poggio and Girosi [25] who explore the relationship between regularization theory and radial basis function networks. The good approximation properties of the radial basis functions in interpolation have been well studied by Powell and his group [28].

In this paper, we present a sequential identification scheme for continuous-time nonlinear dynamical systems with unknown nonlinearities using a RAN [24] like growing Gaussian radial basis function (GRBF) network. The model starts with no hidden units and grows by allocating units on a regular grid, based on the novelty of 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 approach gradually the appropriate complexity of the network that is sufficient to provide an approximation to the system to be identified that is consistent with the observations being received. By allocating GRBF units on a regular grid, only the relevant state-space traversed by the dynamical system is spanned, resulting in a considerable savings on the final size of the network.

The parameters of the growing neural network based identification model are adjusted by adaptation laws developed using the Lyapunov synthesis approach. The sequential identification algorithm for continuous dynamical nonlinear systems is also developed in the continuous-time framework rather than in discrete-time. A similar scheme for adaptive control has been recently developed where growing GRBF network is used of for the control of feedback linearisable continuous-time nonlinear systems with laws derived using the Lyapunov synthesis techniques [7].

The paper is organised as follows: In Section 2, the modelling of single-input single-state (SISS) nonlinear dynamical systems by the GRBF network is discussed and a one-to-one mapping of the state-space to form compact network input space is also proposed. The basic identification algorithm with neural networks using the Lyapunov synthesis approach is developed in Section 3, where the stability of the overall identification scheme is guaranteed. Based on this basic identification algorithm and the RAN-like growing GRBF network, the sequential identification algorithm is developed in Section 4. In Section 5, the sequential identification for the single input single state (SISS) systems is extended to multi-input multi-state (MIMS) systems. The operation of the sequential identification scheme is demonstrated by a simulated example in Section 6.

2 Dynamical system modelling by neural networks

For the sake of simplicity, we first discuss the modelling of single-input single-state (SISS) continuous nonlinear dynamical systems. The multi-input multi-state (MIMS) case will be considered in Section 5. Consider the class of continuous-time dynamical systems with an input – state representation given by,

$$\dot{x}(t) = f(x(t), u(t)), \quad x(0) = x_0, \tag{1}$$

where $f(x, u)$ is an unknown nonlinear function that must be estimated, $u \in \Re^1$ is the input, $x \in \Re^1$ is the state and $t \in \Re^+$ is the temporal variable. By subtracting and adding ax, where a is some positive constant, the system (1) becomes,

$$\dot{x}(t) = -ax + g(x(t), u(t)), \qquad x(0) = x_0, \tag{2}$$

where

$$g(x, u) = f(x, u) + ax \tag{3}$$

is still a nonlinear function. Since neural networks provide an input – output mapping, we construct a model based on equation (2) by replacing the nonlinear part $g(x, u)$ by a neural network. Consider the model [14],

$$\dot{\hat{x}}(t) = -a\hat{x}(t) + \hat{g}(x(t), u(t); \mathbf{p}(t)), \qquad \hat{x}(0) = \hat{x}_0, \tag{4}$$

where $\hat{g}$ is the output of the neural network, $\hat{x}$ denotes the state of the identification model, while $\mathbf{p}(t)$ denotes the adjustable parameters of the network.

Radial basis functions were introduced to the neural network literature by Broomhead and Lowe [4] and have gained significance in the field due to several applications and theoretical results [18], [23], [25]. One of the commonly used radial basis function networks is the Gaussian radial basis function (GRBF) neural network, also called the localised receptive field network. The nonlinear function $g(x, u)$ is therefore approximated by the GRBF network, which is expressed by,

$$\hat{g}(x, u; \mathbf{p}) = \sum_{k=1}^{K} w_k \exp\left\{-\frac{1}{r_k^2}\left[(x - m_{k1})^2 + (u - m_{k2})^2\right]\right\} \tag{5}$$

where r_k is the width of the k^{th} basis function whose centre is $\mathbf{m}_k = [m_{k1}, m_{k2}]^T$, $\mathbf{p}$ is the parameter vector containing w_k, r_k and $\mathbf{m}_k$ $(k = 1, 2, \ldots, K)$.

It is well known that if the variables of a nonlinear function are in compact sets, the continuous function can be approximated arbitrarily well by the GRBF networks. If x and u are not in compact sets, we introduce the following one-to-one (1-1) mapping:

$$\bar{x}(t) = \frac{b_x x(t)}{|x(t)| + a_x}, \qquad \bar{u}(t) = \frac{b_u u(t)}{|u(t)| + a_u} \tag{6}$$

where a_x, b_x, a_u, b_u are positive constants, which can be chosen by the designer (*eg.*, a_x, b_x, a_u, b_u are 1). Thus, it is clear from equation (6) that $\bar{x}(t) \in [-b_x, b_x]$ and $\bar{u}(t) \in [-b_u, b_u]$ for $x, u \in (-\infty, +\infty)$. On the other hand, if x and u are already in compact sets, we only need to set $\bar{x} = x$ and $\bar{u} = u$. The above one-to-one mapping is illustrated in Figure 1, which shows that in two-dimensional space the entire area can be transferred into a rectangular one.

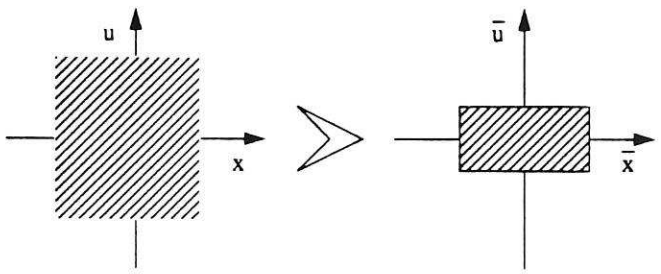

Figure 1: The one-to-one (1-1) mapping.

Replacing x and u by $\bar{x}$ and $\bar{u}$ in equation (5), the model of the system described by the GRBF network can be written as,

$$\hat{g}(\bar{x}, \bar{u}; \mathbf{p}) = \sum_{k=1}^{K} w_k \phi_k(\bar{x}, \bar{u}; \mathbf{m}_k, r_k) \tag{7}$$

where,

$$\phi_i(\bar{x}, \bar{u}; \mathbf{m}_k, r_k) = \exp\left\{-\frac{1}{r_k^2}\left[(\bar{x} - m_{k1})^2 + (\bar{u} - m_{k2})^2\right]\right\} \tag{8}$$

The problem then becomes that of estimating the function $\hat{g}(\bar{x}, \bar{u}, \mathbf{p})$ based on the variables $\bar{x}(t)$ and $\bar{u}(t)$, which are in compact sets. A schematic diagram of the identification framework is shown in Figure 2.

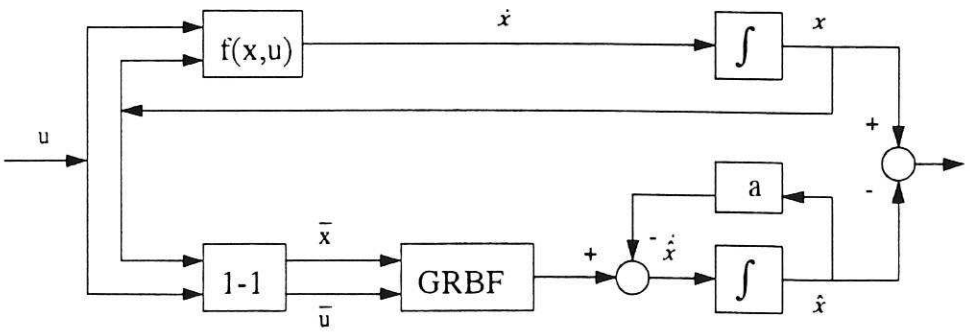

Figure 2: The configuration for identification framework.

3 Stable sequential identification

In system identification, the stability of the overall identification scheme is an important issue. Even when the real system to be identified is bounded-input bounded-state stable there is no ***a priori*** guarantee that the estimated state or the adjustable parameters of the model will remain bounded. The overall stability depends not only on the particular identification model that is chosen but also on the parameter adjustment rules that are used. This problem is solved here by developing a stable parameter adjustment rule based on Lyapunov stability techniques [14] for the GRBF network model discussed in Section 2.

We assume that the basis functions $\phi_k(\bar{x}, \bar{u}; m_k, r_k)$ for $k = 1, 2, \ldots, K$ are given. Section 4 will discuss how the basis functions of the network model are chosen. Thus, based on the above assumption, the GRBF network based system model can be rewritten in the form of,

$$\dot{x} = -ax + \sum_{k=1}^{K} w_k^* \phi_k(\bar{x}, \bar{u}; \mathbf{m}_k, r_k) + \varepsilon(t) \tag{9}$$

where w_k^* $(k = 1, 2, \ldots, K)$ is the optimal weight value and $\varepsilon(t)$ is the modelling error defined as,

$$\begin{aligned} \varepsilon(t) &= g(\bar{x}, \bar{u}) - \hat{g}(\bar{x}, \bar{u}, \mathbf{p}) \\ &= g(\bar{x}, \bar{u}) - \sum_{k=1}^{K} w_k^* \phi_k(\bar{x}, \bar{u}; \mathbf{m}_k, r_k) \end{aligned} \tag{10}$$

It is well known from approximation theory that the modelling error can be reduced arbitrarily by increasing the number K, *i.e.*, the number of the linear independent basis functions ϕ_k in the network model. Thus, it is reasonable to assume that the modelling error $\varepsilon(t)$ is bounded by a constant ε_K, which represents the accuracy of the model and this is defined as,

$$\varepsilon_K = \sup_{t \in \Re^+} |\varepsilon(t)| \tag{11}$$

Since $\bar{x}(t) \in [-b_x, b_x]$ and $\bar{u}(t) \in [-b_u, b_u]$ are bounded, the constant ε_K is finite. From equation (4) we also have the identification model

$$\dot{\hat{x}} = -a\hat{x} + \sum_{k=1}^{K} w_k \phi_k(\bar{x}, \bar{u}; \mathbf{m}_k, r_k) \tag{12}$$

where w_k $(k = 1, \ldots, K)$ is the estimate of w_k^*, while $\hat{x}$ is the estimated state of the model.

Let us define the state error and the weight estimation error respectively as,

$$\begin{aligned} e_x &= x - \hat{x} \\ \xi_k &= w_k^* - w_k \end{aligned} \tag{13}$$

Hence, from equations (9), (12) and (13) the dynamical expression of the state error is,

$$\dot{e}_x = -ae_x + \sum_{k=1}^{K} \xi_k \phi_k(\bar{x}, \bar{u}; \mathbf{m}_k, r_k) + \varepsilon(t) \tag{14}$$

Consider the following Lyapunov function candidate,

$$V(e_x, \mathbf{z}) = \frac{1}{2}\left[e_x^2 + \frac{1}{\alpha}\sum_{k=1}^{K}\xi_k^2\right] \tag{15}$$

where $\mathbf{z} = [\xi_1, \ldots, \xi_K]^T$ and α is a positive constant which will appear in the sequential adaptation laws, also referred to as the learning or adaptation step size. Using equation (15), the time derivative of the Lyapunov function V is given by

$$\dot{V}(e_x, \mathbf{z}) = -ae_x^2 + \sum_{k=1}^{K} e_x\xi_k\phi_k(\bar{x}, \bar{u}; \mathbf{m}_k, r_k) + \frac{1}{\alpha}\sum_{k=1}^{K}\xi_k\dot{\xi}_k + e_x\varepsilon(t) \tag{16}$$

$$= -ae_x^2 + \frac{1}{\alpha}\sum_{k=1}^{K}(\alpha e_x\xi_k\phi_k(\bar{x}, \bar{u}; \mathbf{m}_k, r_k) + \xi_k\dot{\xi}_k) + e_x\varepsilon(t) \tag{17}$$

Since w_k^* is constant, we have that $\dot{w}_k = -\dot{\xi}_k$. Therefore it is clear from equation (17) that if the parameter estimates w_k are adapted according to the following laws:

$$\dot{w}_k = \alpha e_x\phi_k(\bar{x}, \bar{u}; \mathbf{m}_k, r_k), \quad \text{for} \quad k = 1, \ldots, K \tag{18}$$

then equation (17) becomes,

$$\begin{aligned}\dot{V}(e_x, \mathbf{z}) &= -ae_x^2 + e_x\varepsilon(t) \\ &\leq -ae_x^2 + |e_x|\varepsilon_K \\ &= -a|e_x|(|e_x| - \varepsilon_K/a)\end{aligned} \tag{19}$$

If there is no modelling error (*i.e.*, $\varepsilon_K = 0$), then from equation (19), $\dot{V}$ is negative semidefinite; hence the stability of the overall identification scheme is guaranteed. On the other hand, in the presence of a modelling error, if $|e_x| < \varepsilon_K/a$ then it is possible that $\dot{V} > 0$, which implies that the weights w_k may drift to infinity with time. In order to avoid this drift, we can set $\dot{w}_k = 0$ if $|e_x| < \varepsilon_K/a$ so that the state error will converge to the set $|e_x| \leq \varepsilon_K/a$. However, the upper bound ε_K is unknown. Thus, we set an upper bound $\sqrt{K}M$ on $\|\mathbf{w}\|$ (Euclidean norm of weight vector), where $\mathbf{w}(t) = [w_1(t), \ldots, w_K(t)]^T$, and an upper bound e_0 (required accuracy) on the state error e_x. Then the modified estimation law is,

$$\dot{w}_k = \begin{cases} \alpha_i e_x\phi_k(\bar{x}, \bar{u}; \mathbf{m}_k, r_k), & \text{if} \quad |e_x| \geq e_0 \quad \text{and} \quad \|\mathbf{w}\| \leq \sqrt{K}M \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

for $k = 1, \ldots, K$.

It can be seen from the modified weight adjustment laws above that if $|e_x| \geq e_0 \geq \varepsilon_K/a$, the first derivative of the Lyapunov function with respect to time t is always negative semidefinite. Although in the case where $e_0 \leq |e_x| \leq \varepsilon_K/a$, the weights may increase with time because it is possible that $\dot{V} > 0$, it is known from the estimation laws (20) that the weights are still limited by the bound $\sqrt{K}M$. If $|e_x| > e_{max}$ (the maximum tolerable accuracy) and $\|\mathbf{w}\| = \sqrt{K}M$, it means that more GRBF units are needed to approximate $\hat{g}$. Therefore, the overall identification scheme is still stable in the presence of modelling error. The Lyapunov function V depends also on the parameter error and the negative semi-definiteness then implies convergence of the algorithm.

4 Identification with growing GRBF network

The control of real-time systems with unknown structure and parameter information can be based on carrying out on-line or sequential identification using non-parametric techniques such as neural networks. The sequential identification problem for continuous dynamic systems can be stated as follows: Given the required modelling error, the prior identification model structure and the on-line or sequential continuous observation, how are these combined to obtain the model parameter adaptive laws or the required neural network approximation?

The identification problem for the dynamical system of equation (1) can be viewed as the estimation of the nonlinear function $g(\bar{x}, \bar{u}; \mathbf{p})$ as shown in Section 2. If the modelling error is greater than required, according to approximation theory more basis functions should be added to the network model to get a better approximation. In this case, denote the prior identification structure of the function at time t as $\hat{g}^{(t)}(\bar{x}, \bar{u}; \mathbf{p})$ and the structure immediately after the addition of a basis function as $\hat{g}^{(t+)}(\bar{x}, \bar{u}; p)$. Based on the structure of the function $\hat{g}(\bar{x}, \bar{u}; p)$ in equation (5), the identification structure now becomes,

$$\hat{g}^{(t+)}(\bar{x}, \bar{u}; \mathbf{p}) = \hat{g}^{(t)}(\bar{x}, \bar{u}; \mathbf{p}) + w_{K+1}\phi_{K+1}(\bar{x}, \bar{u}; \mathbf{m}_{K+1}, r_{K+1}) \tag{21}$$

where w_{K+1} is the weight of the new $(K+1)^{th}$ Gaussian radial basis function ϕ_{K+1}. The sequential identification scheme using neural network for the nonlinear function $g(\bar{x}, \bar{u}; \mathbf{p})$ is shown in Figure 3.

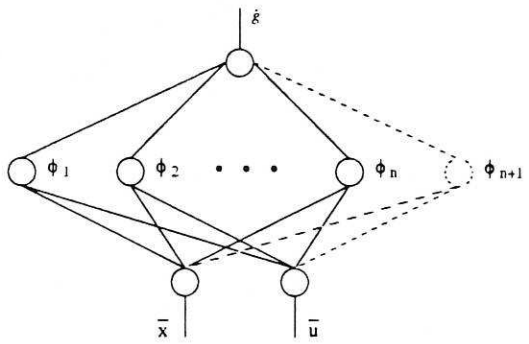

Figure 3: The sequential identification scheme using the GRBF network.

It is also known that the k^{th} Gaussian radial basis function has a localisation property that the influence area of this function is governed by the centre $\mathbf{m}_k$ and width r_k. In other words, once the centre $\mathbf{m}_k$ and the width r_k are fixed, the influence area of the k^{th} Gaussian radial basis function ϕ_k is limited in the state-space to the neighbourhood of $\mathbf{m}_k$.

Let us first consider how to limit the number of the centres and hence the size of the network. As shown in Figure 1, the observation pairs $(\bar{x}, \bar{u})$ are in a rectangular set. An $h_x \times h_u$ grid, where h_x and h_u are odd integers, can be produced by scaling the $\bar{x}$ and $\bar{u}$ axes

by $2b_x/(h_x - 1)$ and $2b_u/(h_u - 1)$, respectively, as shown in Figure 4. If the centres of the basis functions of the network model are located on some of the crosspoints of the grid it is clear that those centres will be equally distributed. For any point $(\bar{x}, \bar{u})$ in the rectangular set, the nearest crosspoint $(\bar{x}_m, \bar{u}_m)$ can be calculated by,

$$\bar{x}_m = \text{round}\left(\frac{\bar{x}}{\delta_x}\right)\delta_x, \qquad \bar{u}_m = \text{round}\left(\frac{\bar{u}}{\delta_u}\right)\delta_u \tag{22}$$

where $round(\cdot)$ is an operator for rounding the number $(\cdot)$ to the nearest integer, (for example, $round(2.51) = 3$), and

$$\delta_x = \frac{2b_x}{h_x - 1}, \quad \text{and} \quad \delta_u = \frac{2b_u}{h_u - 1}. \tag{23}$$

The main influence area D of the radial basis function with the centre $(\bar{x}_m, \bar{u}_m)$ is also shown in Figure 4.

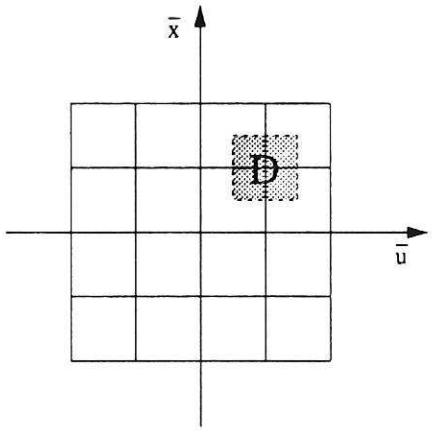

Figure 4: The two-dimensional grid.

Now, consider how the width r_k of the k^{th} basis function are chosen. The angle between the two GRBFs ϕ_i and ϕ_j with the same width $r_i = r_j = r_0$ is given by [9], [12],

$$\theta_{ij} = \cos^{-1}\left[\exp\left\{-\frac{1}{2r_0^2}\|\mathbf{m}_i - \mathbf{m}_j\|^2\right\}\right]. \tag{24}$$

In order to assign a new basis function ϕ_k that is nearly orthogonal to all existing basis functions, the angle between the GRBFs should be as large as possible. The width r_k should therefore be reduced. However, reducing r_k increases the curvature of ϕ_k which in turn gives a less smooth function and can lead to *overfitting* problems. Thus, to make a trade-off between the orthogonality and the smoothness, a good choice for the width r_k, which ensures the angles between GRBF units are approximately equal to some required angle θ_{min}, is [9], [12],

$$r_k = \kappa\|\mathbf{m}_k - \mathbf{m}_k^+\|, \tag{25}$$

where

$$\kappa = \left[\frac{1}{2\log(1/\cos^2\theta_{min})}\right]^{\frac{1}{2}} \tag{26}$$

with θ_{min} being the required minimum angle between Gaussian radial basis functions, and,

$$m_k^+ = \arg \min_{i=1,\ldots,K,\ \mathbf{m}_i \neq \mathbf{m}_k} \{\|\mathbf{m}_k - \mathbf{m}_i\|\} \tag{27}$$

is the nearest (in the Euclidean space) centre to the k^{th} centre. The above assignments are same as those for the resource allocating network (RAN) [24] for which the equations are arrived at from the consideration of observation novelty heuristics.

The growing network is initialised with no basis function units. As observations are received the network grows by adding new units. The decision to add a new unit depends on the observation novelty for which the following two conditions must be satisfied:

$$\text{(i)} \quad \begin{array}{c} \min_{k=1,\ldots,K} |\bar{x}(t) - m_{k1}| > \frac{\delta_x}{2} \\ \text{or} \\ \min_{k=1,\ldots,K} |\bar{u}(t) - m_{k2}| > \frac{\delta_u}{2} \end{array} \tag{28}$$

$$\text{(ii)} \quad |x(t) - \hat{x}(t)| > e_{max} \tag{29}$$

where δ_x and δ_u represent the scale of resolution in the input-state grid, and e_{max} is chosen to represent the desired maximum tolerable accuracy of the state estimation. Criterion (i) says that the current observation must be far from existing centers. Criterion (ii) means that the state error in the network must be significant.

When a new unit is added to the network at time t_1, the parameters associated with the GRBF units are adapted as follows:

$$\mathbf{m}_{K+1} = \left[\text{round}\left(\frac{\bar{x}(t_1)}{\delta_x}\right)\delta_x, \quad \text{round}\left(\frac{\bar{u}(t_1)}{\delta_u}\right)\delta_u\right]^T \tag{30}$$

$$\mathbf{m}_k^+ = \arg \min_{i=1,\ldots,K+1,\mathbf{m}_i \neq \mathbf{m}_k} \{\|\mathbf{m}_k - \mathbf{m}_i\|\} \tag{31}$$

$$r_k = \kappa \|m_k - m_k^+\| \tag{32}$$

$$\dot{w}_k = \begin{cases} \alpha e_x \phi_k(\bar{x}, \bar{u}, \mathbf{m}_k, r_k), & \text{if} \quad |e_x| \geq e_0 \quad \text{and} \quad \|\mathbf{w}\| \leq \sqrt{K+1}M \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

for $k = 1, \ldots, K+1$ and $w_{K+1}(t_1+) = 0$. If no new GRBF unit is added, only the weights are adapted by the law (33), for $k = 1, \ldots, K$.

It is known from approximation theory [27] that the approximation accuracy of a function by a set of basis functions, such as in neural networks, is proportional to the parameters δ_x and δ_u of the grid. In other words, the smaller the parameters δ_x and δ_u, the more accurate the neural model. If the tolerable accuracy of the state error is not reached, *ie.*, $|e_x| > e_{max}$, then the thresholds δ_x and δ_u on the criterion (i) should gradually be reduced by halving their values (*ie.*, $\delta_x/2$ and $\delta_u/2$) at each time step until the minimum allowed values are reached. In this way, the state error will be reduced and the existing centres of the basis functions of the network model are all still on the crosspoints of the new grid as shown in Figure 5.

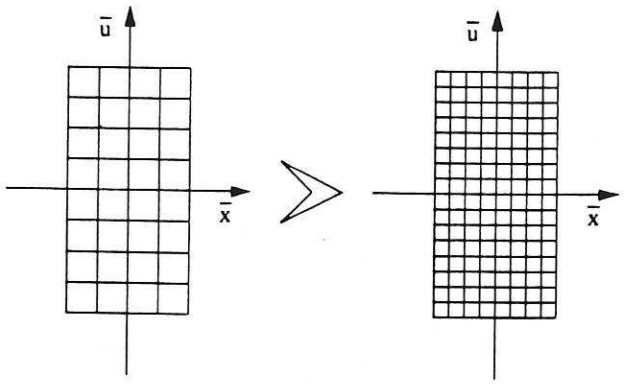

Figure 5: Modification of the two-dimension grid.

With the increase of the number of the GRBFs and the cross-points of the grid, the approximation of a function by a GRBF network will be increasingly more accurate, *ie.*, $\varepsilon_{K+1} < \varepsilon_K$. According to the approximation theory there exists a number K^* such that $\varepsilon_{K^*} < \epsilon_0$. It has also been shown in section 3 that the overall identification scheme is stable and that the model parameters converge to within some bound of the optimal values. Therefore, the algorithm developed in this section guarantees the stability and convergence of the overall identification.

5 Identification of Multivariable Systems

In this section the sequential identification developed for single-input single-state systems is extended to multivariable systems. Consider the multi-input multi-state (MIMS) continuous dynamical system described by,

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{34}$$

where $\mathbf{u} \in \Re^{r\times 1}$ is the input vector, $\mathbf{x} \in \Re^{n\times 1}$ is the state vector and $\mathbf{f}(\cdot) \in \Re^{n\times 1}$ is a nonlinear function vector. Following the same line of analysis as for the single-input single-state case, the identification model for the system system (34) can be expressed by,

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x} + \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{35}$$

where

$$\mathbf{g}(\mathbf{x}, \mathbf{u}) = \mathbf{f}(\mathbf{x}, \mathbf{u}) - \mathbf{A}\mathbf{x} \tag{36}$$

and $\mathbf{A} \in \Re^{n\times n}$ is a Hurwitz or stability matrix (*ie.*, all the eigenvalues are in the open left-half complex plane). Modelling the nonlinear function vector $\mathbf{g}(\mathbf{x}, \mathbf{u})$ using GRBF neural networks gives the following identification model:

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{A}\hat{\mathbf{x}}(t) + \hat{\mathbf{g}}(\mathbf{x}(t), \mathbf{u}(t); \mathbf{p}(t)), \quad \hat{\mathbf{x}}(0) = \mathbf{x}_0 \tag{37}$$

where $\hat{\mathbf{x}}$ denotes the state vector of the network model and $\hat{\mathbf{g}}$ is the output vector of the GRBF neural network. Define the following one-to-one mappings for the inputs and states:

$$\bar{x}_i(t) = \frac{b_{xi}x_i(t)}{|x_i(t)| + a_{xi}} \quad \text{for} \quad i = 1, 2, \ldots, n, \qquad \bar{u}_i(t) = \frac{b_{ui}u_i(t)}{|u_i(t)| + a_{ui}} \quad \text{for} \quad i = 1, 2, \ldots, r \tag{38}$$

where $a_{xi}, b_{xi}, a_{ui}, b_{ui}$ are positive constants. These mappings ensure that the elements of the vectors $\bar{x}$ and $\bar{u}$ are all in compact sets. The estimate of the function $\hat{\mathbf{g}}$ then is written as,

$$\hat{\mathbf{g}}(\mathbf{x}(t), \mathbf{u}(t); \mathbf{p}(t)) = \hat{\mathbf{g}}(\bar{\mathbf{x}}, \bar{\mathbf{u}}; \mathbf{p}) = \mathbf{W}_K \Phi_K(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \tag{39}$$

where,

$$\begin{aligned} \Phi_K(\bar{\mathbf{x}}, \bar{\mathbf{u}}) &= [\phi_1(\bar{\mathbf{x}}, \bar{\mathbf{u}}), \phi_2(\bar{\mathbf{x}}, \bar{\mathbf{u}}), ..., \phi_K(\bar{\mathbf{x}}, \bar{\mathbf{u}})]^T \\ \phi_k(\bar{\mathbf{x}}, \bar{\mathbf{u}}) &= \exp\left\{ -\tfrac{1}{r_k^2} \left\| \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{u}} \end{bmatrix} - \mathbf{m}_k \right\|^2 \right\} \end{aligned} \tag{40}$$

$\mathbf{W}_K = \{w_{ik}\} \in \Re^{n \times K}$ is the weight matrix of the network with K GRBF units, $\phi_k(\cdot) \in \Re^1$ is the k^{th} Gaussian radial basis function, and $\mathbf{p}(t)$ denotes the adjustable parameter vector of the network consisting of the weight matrix $\mathbf{W}_K$, the centres $\mathbf{m}_k = [m_{k1}, m_{k2}, \ldots, m_{k(n+r)}]^T$ and the width r_k for $k = 1, 2, \ldots, K$.

Assuming that the basis function vector $\Phi_K(\bar{\mathbf{x}}, \bar{\mathbf{u}})$ is given the real system can be modelled by the GRBF network model as,

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x} + \mathbf{W}_K^* \Phi_K(\bar{\mathbf{x}}, \bar{\mathbf{u}}) + \mathbf{e}(t) \tag{41}$$

where $\mathbf{W}_K^*$ is the optimal weight matrix and $\mathbf{e}(t) = [e_1(t), e_2(t), ..., e_n(t)]^T$ is the modelling error vector which is assumed to be bounded by,

$$\varepsilon_K = \max_{i=1,\ldots,n} \sup_{t \in \Re^+} \{|e_i(t)|\} \tag{42}$$

Define the state error vector and weight error matrix as,

$$\begin{aligned} \mathbf{e}_x &= \mathbf{x} - \hat{\mathbf{x}} \\ \Gamma_K &= \mathbf{W}_K^* - \mathbf{W}_K \end{aligned} \tag{43}$$

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

$$\dot{\mathbf{e}}_x = \mathbf{A}\mathbf{e}_x + \Gamma_K \Phi_K(\bar{\mathbf{x}}, \bar{\mathbf{u}}) + \mathbf{e}(t) \tag{44}$$

Consider the Lyapunov function,

$$V(\mathbf{e}_x, \Gamma_K) = \frac{1}{2}[\mathbf{e}_x^T \mathbf{e}_x + \frac{1}{\alpha} tr(\Gamma_K \Gamma_K^T)] \tag{45}$$

where $tr(\cdot)$ denotes the trace of the matrix $(\cdot)$. The first derivative of the Lyapunov function V with respect to time t is,

$$\dot{V}(\mathbf{e}_x, \Gamma_K) = \mathbf{e}_x^T A \mathbf{e}_x + \mathbf{e}_x^T \Gamma_K \Phi_K(\bar{\mathbf{x}}, \bar{\mathbf{u}}) + \frac{1}{\alpha} tr(\Gamma_K \dot{\Gamma}_K^T) + \mathbf{e}_x^T \mathbf{e}(t) \tag{46}$$

Since,

$$\begin{aligned} \mathbf{e}_x^T \Gamma_K \Phi_K(\bar{\mathbf{x}}, \bar{\mathbf{u}}) &= tr(\mathbf{e}_x^T \Gamma_K \Phi_K(\bar{\mathbf{x}}, \bar{\mathbf{u}})) \\ &= tr(\Gamma_K \Phi_K(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \mathbf{e}_x^T) \end{aligned} \tag{47}$$

equation (46) becomes,

$$\dot{V}(\mathbf{e}_x, \Gamma_K) = \mathbf{e}_x^T \mathbf{A} \mathbf{e}_x + tr(\Gamma_K \Phi_K(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \mathbf{e}_x^T) + \frac{1}{\alpha} tr(\Gamma_K \dot{\Gamma}_K^T) + \mathbf{e}_x^T \mathbf{e}(t) \tag{48}$$

and because $\mathbf{W}_K^*$ is a constant matrix $\dot{\mathbf{W}}_K = -\dot{\Gamma}_K$. Thus, if the estimation law for the weight matrix is given by,

$$\dot{\mathbf{W}}_K = \alpha \mathbf{e}_x \Phi_K^T(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \tag{49}$$

then equation (48) becomes,

$$\begin{aligned} \dot{V}(\mathbf{e}_x, \Gamma_K) &= \mathbf{e}_x^T \mathbf{A} \mathbf{e}_x + \mathbf{e}_x^T \mathbf{e}(t) \\ &\leq -|\lambda_{max}(\mathbf{A})| \mathbf{e}_x^T \mathbf{e}_x + \mathbf{e}_x^T \mathbf{e}(t) \\ &\leq -|\lambda_{max}(\mathbf{A})| \sum_{i=1}^{n} |e_{xi}| \left\{ |e_{xi}| - \frac{\varepsilon_K}{|\lambda_{max}(\mathbf{A})|} \right\} \end{aligned} \tag{50}$$

where $\lambda_{max}(\mathbf{A})$ is the maximum eigenvalue of the matrix $\mathbf{A}$ which could be negative because $\mathbf{A}$ is Hurwitz or stability matrix and e_{xi} is the i^{th} element of the state error vector $\mathbf{e}_x$. In the presence of a modelling error, if

$$\min_{i=1,\ldots,n} \{|e_{xi}|\} < \frac{\varepsilon_K}{|\lambda_{max}(\mathbf{A})|} \tag{51}$$

then it is possible that $\dot{V} > 0$, which implies that the weights w_{ik} may drift to infinity over time. Following the analysis for the single-input single-state case, this drift can be avoided by modifying the adaptation law as,

$$\dot{w}_{ik} = \begin{cases} \alpha e_{xi} \phi_K(\bar{\mathbf{x}}, \bar{\mathbf{u}}), & \text{if} \quad |e_{xi}| \geq e_0 \quad \text{and} \quad \|\mathbf{W}_K| \leq \sqrt{K} M \\ 0 & \text{otherwise} \end{cases} \tag{52}$$

for $i = 1, \ldots, n$ and $k = 1, \ldots, K$, where e_0 is the required accuracy of the state error $\mathbf{e}_x$ and $\sqrt{K} M$ are the upper bound on the Euclidean norm of the weight matrix $\mathbf{W}_K$. It is easy to show that the overall identification scheme with the modified identification laws is stable.

As observations are received the network grows by adding new units. The decision to add a new unit depends on the observation novelty for which the following must be satisfied:

$$\text{(i)} \qquad \min_{k=1,\ldots,K, i=1,\ldots,n} |\bar{x}_i(t) - m_{ki}| > \frac{\delta_{xi}}{2}$$

$$\text{or} \tag{53}$$

$$\min_{k=1,\ldots,K, j=1,\ldots,r} |\bar{u}_j(t) - m_{k(j+n)}| > \frac{\delta_{uj}}{2}$$

$$\text{(ii)} \qquad \max_{i=1,\ldots,n} \quad |x_i(t) - \hat{x}_i(t)| > e_{max} \tag{54}$$

where δ_{xi} and δ_{uj} represent the scale of resolution in the input-state grid and e_{max} is chosen to represent the desired maximum tolerable accuracy of the state estimation. When a new

unit is added to the network at time t_1 the parameters associated with the GRBF units are adapted as follows:

$$\mathbf{m}_{(K+1)i} = \text{round}\left(\frac{\bar{x}(t_1)}{\delta_{xi}}\right)\delta_{xi}, \quad \text{for} \quad i = 1,\ldots,n \tag{55}$$

$$\mathbf{m}_{(K+1)(j+n)} = \text{round}\left(\frac{\bar{u}(t_1)}{\delta_{uj}}\right)\delta_{uj}, \quad \text{for} \quad j = 1,\ldots,r \tag{56}$$

$$\mathbf{m}_k^+ = \arg \min_{i=1,\ldots,K+1, \mathbf{m}_i \neq \mathbf{m}_k} \{\|\mathbf{m}_k - \mathbf{m}_i\|\}, \quad \text{for} \quad k = 1,\ldots,K+1 \tag{57}$$

$$r_k = \kappa\|\mathbf{m}_k - \mathbf{m}_k^+\|, \quad \text{for} \quad k = 1,\ldots,K+1 \tag{58}$$

$$\dot{w}_{ik} = \begin{cases} \alpha e_{xi}\phi_k(\bar{\mathbf{x}},\bar{\mathbf{u}}), & \text{if} \quad |e_{xi}| \geq e_0 \quad \text{and} \quad \|\mathbf{W}_{K+1}| \leq \sqrt{K+1}M \\ 0 & \text{otherwise} \end{cases} \tag{59}$$

$$\text{for} \quad i = 1,\ldots,n \quad \text{and} \quad k = 1,\ldots,K+1$$

where $\mathbf{w}_{(K+1)}(t_1+) = \mathbf{0}$. If no new GRBF unit is added, only the weight matrix $\mathbf{W}_K$ is adapted using the law above in (59).

6 Simulation results

In this section a SISS dynamical system is used to demonstrate the operation of the sequential identification algorithm for a continuous nonlinear system. The following SISS continuous-time dynamical system was considered:

$$\dot{x} = 1.1(1 - x - 2xu + x^2)\exp(-0.5x^2 - u^2) \tag{60}$$

where the input u is assumed to be $cos(t)$ and the initial state $x(0) = 0$. The parameter values used in this example are as follows: $e_0 = 0.001, e_{max} = 0.005, \delta_x = \delta_u = 0.05, a = 0.5, M = 1.5, \alpha = 1, \kappa = 3.0, x_0 = 0$.

The simulation was begun with no GRBF units in the network model and the number of units increased with time, according to the growth criteria. The final results after an operation over a period of 10 seconds gave a GRBF network with 16 hidden units, for approximating the dynamical system. The performance of the sequential identification scheme using the GRBF network are shown in Figures 6 – 9, for a typical run of the algorithm observing that similar plots obtained under different operational conditions.

The actual and estimated states and the state error of the dynamical system against time t are shown in Figures 6 and 7, respectively. It can be seen from Figure 6 that for much of the operation, the state error is constrained within the maximum tolerable bound $e_{max} = 0.005$. The network parameters also converged to a set of values although they were oscillating around these values. A plot of the actual state x and the estimated state $\hat{x}$ against the input u is shown in Figure 8 which indicates the presence of the strong nonlinearity in the dynamical system. Figure 9 shows the relationship between the estimated state and its first derivative as they gradually approach the true set of values.

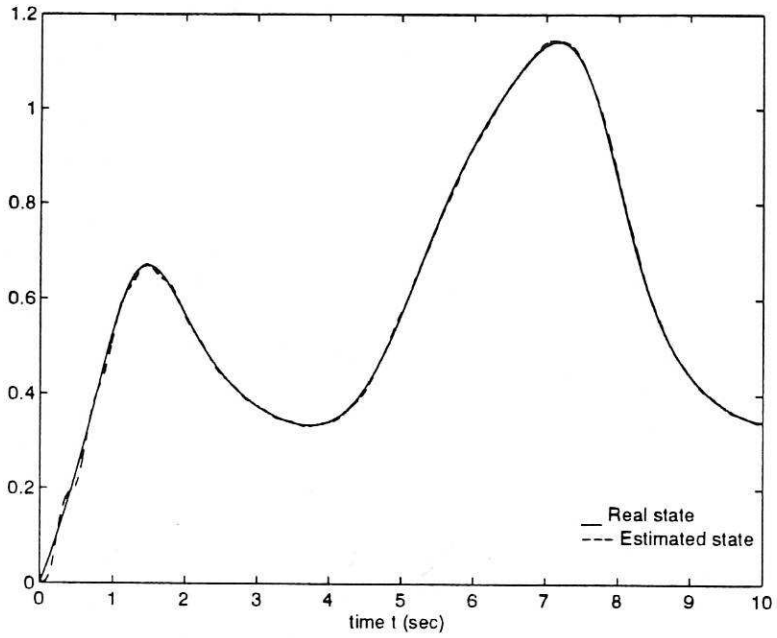

Figure 6: The real state x and the estimated state $\hat{x}$ over time.

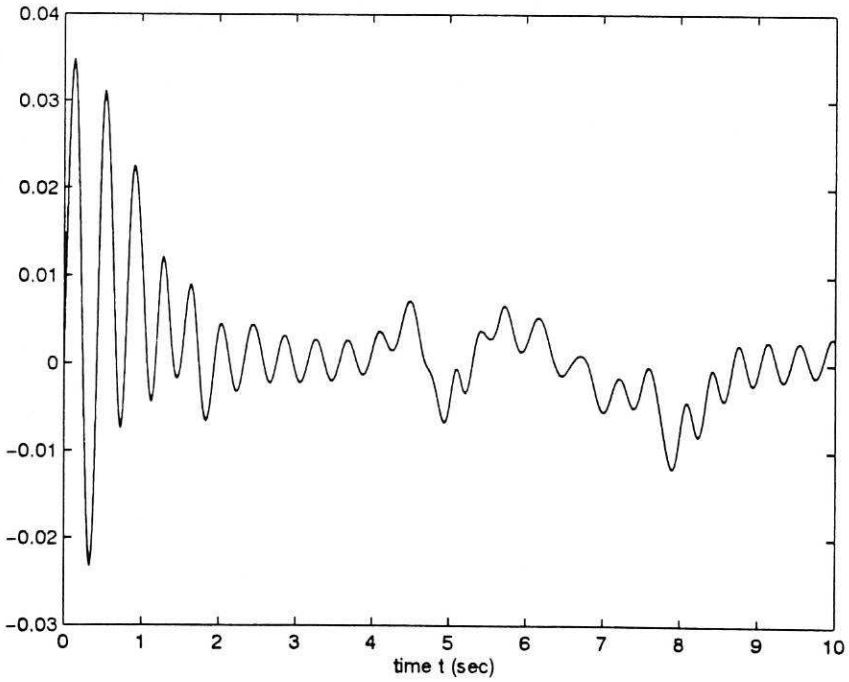

Figure 7: The state error e_x with time.

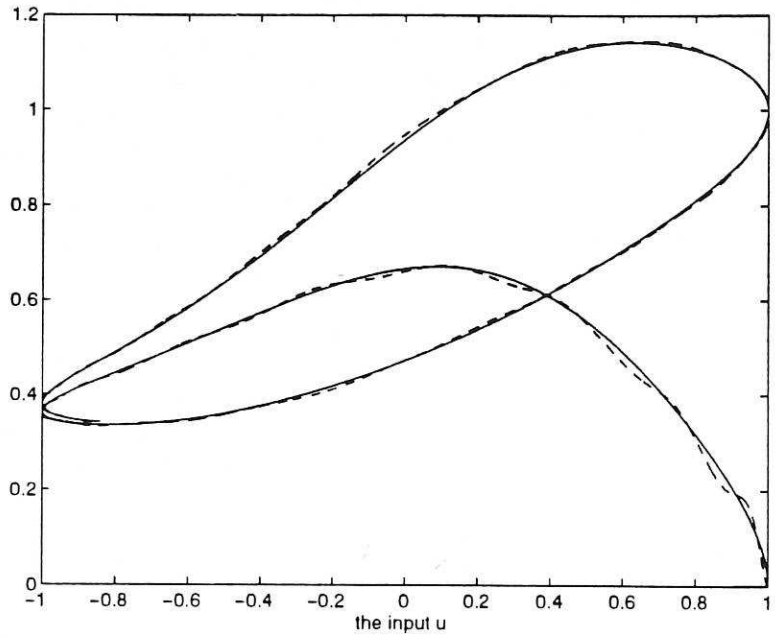

Figure 8: The actual state x (—) and the estimated state $\hat{x}$ (- - -) against the input u.

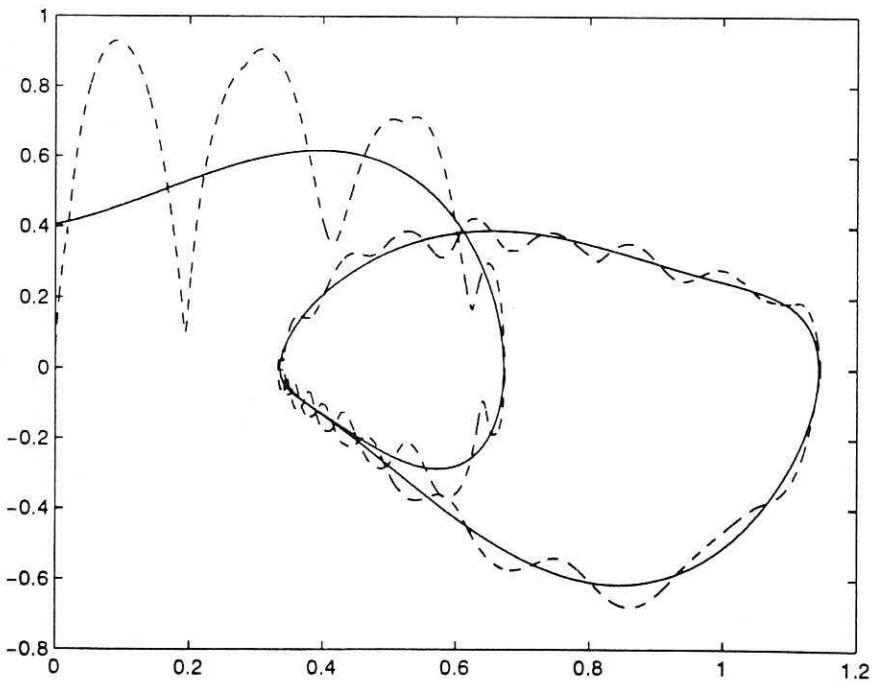

Figure 9: The actual derivative of the state $\dot{x}$ against the state x (—) and the estimated derivative of the state $\dot{\hat{x}}$ against the state $\hat{x}$ (- - -)

7 Conclusions

A sequential identification scheme for continuous nonlinear dynamical systems with unknown nonlinearities using neural networks has been developed. The main feature of this scheme is the combination of the growing Gaussian radial basis function network with that of Lyapunov synthesis techniques in developing the adaptive or estimation laws that guarantee the stability of the system. The idea of growing the network, similar to the resource allocating network (RAN) overcomes the problem of having to choose the neural network structure *a priori*, a difficult task which often results in an over-determined network. The network begins with no radial basis function units and with increasing time, the model grows gradually to approach the appropriate complexity of the network that is sufficient to provide the required approximation accuracy. The stability of the overall identification scheme and convergence of the model parameters are guaranteed by parameter adjustment laws developed using the Lyapunov synthesis approach. To ensure that the modelling error is reduced arbitrarily, a transformation is proposed so that the states and inputs of the system are mapped into compact sets. The operation of the sequential identification algorithm is demonstrated on a simulated experiment and the results conform to the theoretical expectations.

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] P. J. Antsaklis. (ed.), Special Issue on Neural networks in Control Systems, *IEEE Control Systems magazine*, Vol.10, No.3, 1990.

[2] E. Baum and D. Haussler, "What size net gives valid generalization", *Neural Computation*, Vol.1, No.1, 1989.

[3] S. A. Billings and S. Chen. "Neural Networks and System Identification", In K. Warwick *et al.*, (eds.) *Neural networks for systems and control*, pp.181-205, 1992.

[4] D. S. Broomhead and D. Lowe. "Multivariable functional interpolation and adaptive networks", *Complex Systems*, Vol.2, pp.321-355, 1988.

[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 S. A. Billings. "Neural networks for nonlinear dynamic system modelling and identification", *International Journal of Control*, Vol.56, No.2, pp.319-346, 1992.

[7] S. Fabri, and V. Kadirkamanathan. "Dynamic structure neural networks for stable adaptive control of nonlinear systems", Automatic Control and Systems Engineering Department Report No. 536, University of Sheffield, October, 1994.

[8] S. E. Fahlman and C. Lebiere. "The cascade-correlation architecture", In D. S. Touretzky (ed.), *Advances in Neural Information Processing Systems 2*, Morgan Kaufmann, CA: San Mateo, 1990.

[9] V.Kadirkamanathan. *Sequential Learning in Artificial Neural Networks*, Ph.D Thesis, University of Cambridge, UK, 1991.

[10] V. Kadirkamanathan. "A statistical inference based growth criteria for the RBF network," In *Proc. IEEE Workshop on Neural Networks for Signal Processing*, pp. 12-21, 1994.

[11] V. Kadirkamanathan and M. Niranjan. "Application of an architecturally dynamic network for speech pattern classification", In *Proceedings of the Institute of Acoustics*, Vol.14, Part 6, pp.343-350, 1992.

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

[13] J. G. Kuschewski, S. Hui and S. H. Zak. "Application of feedforward neural networks to dynamical system identification and control", *IEEE Trans. on Control Systems Technology*, Vol.1, No.1, pp.37-49, 1993.

[14] I. D. Landau. *Adaptive Control — The Model Reference Approach*, Marcel Dekker, Inc., New York, 1979.

[15] G. P. Liu and V. Kadirkamanathan. "Multiobjective criteria for nonlinear model selection and identification with neural networks", Automatic Control and Systems Engineering Department Report No. 508, University of Sheffield, March, 1994.

[16] D. J. C. MacKay. "Bayesian interpolation", *Neural Computation*, Vol.4, No.3, pp.415-447, 1992.

[17] T. M. Miller, R. S. Sutton III and P. J. Werbos (eds). *Neural Networks for Control*, MIT Press, Cambridge, MA, 1990.

[18] J. Moody and C. J. Darken. "Fast learning in networks of locally-tuned processing units", *Neural Computation*, Vol.1, pp.281-294, 1989.

[19] M. C. Mozer and P. Smolensky. "Skeletonization: A technique for trimming the fat from a network via relevance assignment", In D. S. Touretzky (ed.), *Advances in Neural Information Processing Systems 1*, Morgan Kaufmann, CA: San Mateo, 1989.

[20] K. S. Narendra and A. M. Annaswamy. *Stable adaptive systems*, Prentice-Hall, Engelwood Cliffs, New Jersey, 1989.

[21] 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.

[22] K. S. Narendra. and K. Parthasarathy. "Gradient methods for the optimization of dynamical systems containing neural networks", *IEEE Trans. on Neural Networks*, Vol.2, pp.252-262, 1991.

[23] M. Niranjan and F. Fallside. "Neural networks and radial basis functions for classifying static speech patterns," *Computer Speech and Language*, Vol.4, pp.275-289, 1990.

[24] J. Platt. "A resource allocating network for function interpolation", *Neural Computation*, Vol.4, No.2, pp.213-225, 1991.

[25] T. Poggio and F. Girosi. "Regularization algorithms for learning that are equivalent to multilayer networks", *Science*, Vol.247, pp.978-982, 1990.

[26] M. M. Polycarpou and P. A. Ioannou. "Identification and control of nonlinear systems using neural network models: design and stability analysis", Technical Report 91-09-01, Department of Electrical Engineering-Systems, University of Southern California, USA, 1991.

[27] M. J. D. Powell. *Approximation Theory and Methods*, Cambridge University Press, Cambridge, 1981.

[28] M. J. D. Powell. "Radial basis functions for multivariable interpolation: A review", In J. C. Mason and M. G. Cox, (eds), *Algorithms for Approximation*, pp.143-167, Oxford University Press, Oxford, 1987.

[29] S. Z. Qin, H. T. Su and T. J. McAvoy. "Comparison of four net learning methods for dynamic system identification", *IEEE Trans. on Neural Networks*, Vol.3, No.1, pp.122-130, 1992.

[30] N. Sadegh. "A perceptron network for functional identification and control of nonlinear systems", *IEEE Transactions on Neural Networks*, Vol.4, No.6, pp.982-988, 1993.

[31] R. M. Sanner and J. J. E. Slotine. "Gaussian networks for direct adaptive control", *IEEE Trans. on Neural Networks*, Vol.3, No.6, pp.837-863, 1992.

[32] J. J. E. Slotine and W. Li. *Applied Nonlinear Control.* Prentice-Hall International, Englewood Cliffs, New Jersey 1991.

[33] P. Smyth. "On stochastic complexity and admissible models for neural network classifiers", In R. P. Lippmann, J. Moody and D. S. Touretzky (eds.), *Advances in Neural Information Processing Systems 3*, Morgan Kaufmann, CA, San Mateo, 1991.

[34] R. J. Werbos. "backpropagation through time: what it does and how to do it", In *Proceedings of the IEEE*, Vol.78, pp.1550-1560, 1990.

[35] R. J. Williams and D. Zipser. "A learning algorithm for continuous running fully recurrent neural networks", *Neural Computation*, Vol.1, pp.270-280, 1989.

[36] M. J. Willis, G. A. Montague, C. Di. Massimo, M. T. Tham and A. J. Morris. "Artificial neural networks in process estimation and control", *Automatica*, Vol.28, No.6, pp.1181-1187, 1992.