Neural Networks for Signal Processing 2, Proc. of the 1992 IEEE Workshop, S.Y. Kung, F. Fallside, J. Aa. Sorenson, C.A. Kamm (Eds), IEEE Press, New York, pp. 444-453, 1992.
Copyright (c) 1992 by the Institute of Electrical and Electronics Engineers.

NONLINEAR SYSTEM IDENTIFICATION USING MULTILAYER PERCEPTRONS WITH LOCALLY RECURRENT SYNAPTIC STRUCTURE

Andrew D. Back†, Ah Chung Tsoi‡

† DSTO Information Technology Division, Salisbury, SA 5108. Australia.

‡ Department of Electrical Engineering, University of Queensland, St. Lucia, QLD 4072, Australia.

Abstract. Recurrent network structures are recognized as having significant capabilities in terms of modelling nonlinear dynamic systems, but few theoretical results have been obtained for such models. A global-feedforward local-recurrent structure was proposed recently having synapses constructed as adaptive IIR filters, and shown to be capable of better modelling performance than a global-feedforward local-feedforward structure, (that is, an MLP with FIR synapses). In this paper, theoretical proofs are given for the representational capabilities of networks with time delay synaptic structure, thereby providing an indication of the limitations of such network architectures.

INTRODUCTION

Modelling nonlinear systems with time-dependence is a topic of growing interest in neural network research. Network structures with time delay synapses have been demonstrated to have useful capabilities for nonlinear signal processing tasks [9]. A class of network intermediate between the usual time delay structures, and recurrent networks has been presented recently [1].

This network employs synapses which are constructed as adaptive filters, and combines principles of linear adaptive filtering and nonlinear modelling due to the MLP structure. In this paper, an analysis of the network architecture is presented, showing the types of nonlinear models that can be approximated by the IIR MLP.

The paper is organized as follows. A description of the network architecture is given in Section 2. The network is analysed in Section 3, and results are presented to prove its representational capabilities. Verification of the nonlinear system identification capabilities of the IIR MLP are given in simulations in Section 4.

THE IIR MULTILAYER PERCEPTRON

The IIR MLP overcomes limitations of previous networks with time-delay synapses, including the Time Delay Neural Network (TDNN) and FIR MLP [14]. Specifically, the IIR MLP does not have the problem of finite window size, and is able to model long sequences of data due to the locally-recurrent structure. Each of the synapses in the IIR MLP are constructed as adaptive IIR filters [8], and are described by

$$y(t) = c \left[\frac{\sum_{j=0}^{n_b} b_j(q^{-j})}{1 - \sum_{i=1}^{n_a} a_i(q^{-i})} z(t) \right] \tag{1}$$

where $y(t)$ is the synapse output, $z(t)$ is the synapse input, q^{-1} is the unit delay operator, c is the synaptic output gain, b_j are the feedforward weights, and a_i are the feedback weights. Note that in comparison to the network architecture introduced in [1], a synaptic gain term has been introduced. This allows the synapses to learn the pole-zero characteristics of the plant dynamics, while the synaptic output gain c is adjusted to give the overall static synapse gain. The equations defining an IIR MLP are given in the definition below.

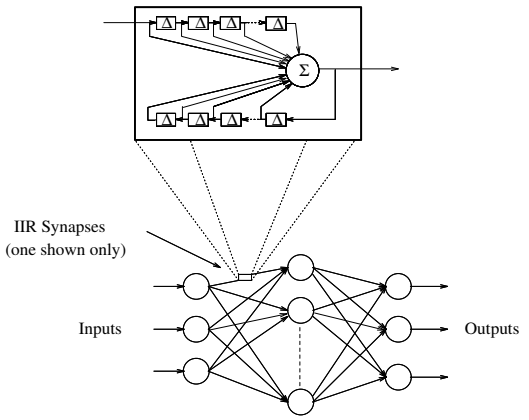

Figure 1: An IIR MLP with global-feedforward local-recurrent architecture.

Definition 2.1 *An IIR MLP of size* (L, n_b, n_a) *with* $N_0, N_1, ..., N_L$ *nodes per layer, is defined by*

$$z_k^{l+1}(t) = f\left(\hat{x}_k^{l+1}(t)\right) \tag{2}$$

$$\hat{x}_k^{l+1}(t) = \sum_{i=1}^{N_l} \hat{y}_{ik}^{l+1}(t) \tag{3}$$

$$\hat{y}_{ik}^{l+1}(t) = c_{ik}^{l+1}\left[\frac{B_{ik}^{l+1}(q^{-1})}{A_{ik}^{l+1}(q^{-1})} z_i^l(t)\right] \tag{4}$$

where

$$B_{ik}^l(q^{-1}) = \sum_{j=0}^{n_b} b_{ikj}^l (q^{-j}) \tag{5}$$

$$A_{ik}^l(q^{-1}) = 1 - \sum_{j=1}^{n_a} a_{ikj}^l (q^{-j}) \tag{6}$$

$$\mathrm{f}(\alpha) = \frac{e^{\alpha/2} - e^{-\alpha/2}}{e^{\alpha/2} + e^{-\alpha/2}} \tag{7}$$

and $k = 1, 2, ..., N_{l+1}$(output layer index), n_b = number of delayed feedforward inputs to a neuron, n_a = number of delayed feedback inputs to a neuron, and $z_k^l|_0$ = bias $(l = 0, 1, ..., L)$. The polynomials $B_{ik}^l(q^{-1})$, and $A_{ik}^l(q^{-1})$ are relatively prime with $n_a \geq n_b$. A diagram of an IIR MLP is shown in Figure 1. A reduced complexity network architecture was proposed in [2]. In this case only one IIR synapse is used for each input or output to a layer.

ANALYSIS OF THE IIR SYNAPSE MULTILAYER PERCEPTRON

A Decomposition Principle

The IIR MLP architecture models nonlinear dynamic systems by combining the known properties of multilayer perceptrons [7], and the established theory of linear systems (see for example Ljung and Soderstrom [11]). In previous work, it was shown that it is possible to

decompose the dynamic MLP architecture into a system comprising separable linear dynamics and static nonlinearities [2], [3]. The decomposition principle can be summarized as follows.

> *An MLP with dynamic synapses has in general two components: a forced component (the particular response), and a natural component (the homogeneous response). For a constant forced input, the particular solution of the synapses defines a static MLP with weights given by the static gains of the synaptic transfer functions. The homogeneous response is due to the pole-zero positions of the synapses, which occur independently of any input signals. This transient response is evoked at the start of any input to the synapses. The homogeneous response of an FIR MLP is zero. The particular response of an FIR MLP provides a finite transient output.*

The significance of the decomposition principle is that it allows the network to be interpreted as having a static nonlinear mapping (given by the static gains of the synapses), with the dynamics introduced as transient terms (assuming stable synaptic poles), having a component due to the inputs (the particular response), and a component independent of the input (the evoked homogeneous response), the latter component being present only in the IIR model.

Each synapse output $\hat{y}(t)$ can be decomposed into particular and homogeneous responses. The homogeneous response is due to the dynamics of the synapse, that is, the pole-zero positions of a^l_{ikj} and b^l_{ikj}. An underlying static nonlinear mapping exists in an FIR or IIR MLP which is determined by the weights

$$w^l_{ik} \quad = \quad c^l_{ik} \left[\frac{\sum_{j=0}^{n_b} b^l_{ikj}}{1 - \sum_{j=1}^{n_a} a^l_{ikj}} \right] \tag{8}$$

where w^l_{ik} are weights corresponding to an equivalent, static MLP in the usual sense.

Representational Capabilities

An important aspect of modelling is to determine the capabilities of a particular model. For the static MLP, it has been proven that the class of network with at least one hidden layer is a universal approximator [7]. In linear system modelling, the capabilities of models are immediately apparent from the model structure.

In the proofs that follow below, it is shown that classic *block-oriented* models [4], [12], such as the Wiener, Hammerstein and Cascade systems (defined below), can be represented by the IIR MLP. This gives an important understanding for the representational power of the IIR MLP, and also has an important bearing on the TDNN model.

Definition 3.2 *A Wiener system is defined by*

$$y(t) = g(H(q^{-1})x(t)) \tag{9}$$

where $x(t)$ is the input to the system, $y(t)$ is the output, $H(q^{-1}) = \frac{\hat{D}(q^{-1})}{\hat{C}(q^{-1})}$ is the linear component of the system, and $g(\cdot)$ is a nonlinear function.

Definition 3.3 *A Hammerstein system is defined by*

$$y(t) = H(q^{-1})g(x(t)) \tag{10}$$

where $x(t)$, $y(t)$, $H(q^{-1})$ and $g(\cdot)$ are as in Definition 3.2.

Definition 3.4 *A Uryson system is defined by*

$$y(t) = \sum_{i=1}^{M_u} H_i(q^{-1})g_i(x(t)) \tag{11}$$

where $x(t)$, $y(t)$, $H(q^{-1})$ and $g(\cdot)$ are as in Definition 3.2.

The following lemmas are used in the main proofs.

Lemma 3.1 *Any layer of an IIR MLP defined by*

$$z_k^l(t) = f\left(\hat{x}_k^l(t)\right) \tag{12}$$

$$\hat{x}_k^l(t) = \hat{y}_{ik}^l(t) \qquad (i=1) \tag{13}$$

$$\hat{y}_{ik}^l(t) = c_{ik}^l(t)\left[\frac{B_{ik}^l(q^{-1})}{A_{ik}^l(q^{-1})} z_i^{l-1}(t)\right] \qquad k=1,2,...,N_l \tag{14}$$

which has a common set of dominant poles and zeros between each IIR synapse in that layer, can be modelled by two layers of reduced complexity, consisting of a single IIR synapse followed by N_l static feedforward weights (no time delays), defined by

$$\hat{y}_{ik}^l(t) = w_{ik}^l\left[\hat{c}_i^l\left[\frac{\hat{B}_i^l(q^{-1})}{\hat{A}_i^l(q^{-1})} z_i^{l-1}(t)\right]\right] \quad i=1,2,...,N_{l-1} \tag{15}$$

where $\hat{c}_i^l\left[\frac{\hat{B}_i^l(q^{-1})}{\hat{A}_i^l(q^{-1})}\right]$ is a single IIR synapse containing the dominant poles and zeros, and w_{ik}^l are the static feedforward weights. The new static weights are defined by

$$w_{ik}^l = c_{ik}^l\left[\frac{\sum_{j=0}^{n_b} b_{ikj}^l}{1-\sum_{j=1}^{n_a} a_{ikj}^l}\right] \tag{16}$$

The synaptic gain $\hat{c}_i^l(t)$ is defined by

$$\hat{c}_i^l = \frac{1-\sum_{j=1}^{\hat{n_a}} \hat{a}_{ij}^l}{\sum_{j=0}^{\hat{n_b}} \hat{b}_{ij}^l} \tag{17}$$

Proof. Each of the synapses has a common set of dominant poles and zeros, therefore let these be retained in the new IIR synapse. The synaptic output gains c_{ik}^l are normally adjusted to give the desired static mapping of the network. The original layer of IIR synapses can therefore be replaced by a reduced complexity layer consisting of the new IIR synapse defined by

$$\hat{c}_i^l\left[\frac{\hat{B}_i^l(q^{-1})}{\hat{A}_i^l(q^{-1})}\right] = \hat{c}_i^l\left[\frac{\sum_{j=0}^{\hat{n_b}} \hat{b}_{ij}^l(q^{-j})}{1-\sum_{j=1}^{\hat{n_a}} \hat{a}_{ij}^l(q^{-j})}\right] \tag{18}$$

where

$$\hat{c}_i^l = \frac{1-\sum_{j=1}^{\hat{n_a}} \hat{a}_{ij}^l}{\sum_{j=0}^{\hat{n_b}} \hat{b}_{ij}^l} \tag{19}$$

Thus, the new synapse contains the previous dominant poles and zeros, while maintaining a static gain of unity. The single IIR synapse is followed by a layer of static weights which each have a static gain equal to that of the original IIR synapse in its place. By linear superposition, the gain between the input $z_i^{l-1}(t)$ and any synaptic output in the original setting, will be equivalent to the new formulation, due to the same static gain in each case, and the dominant poles and zeros being maintained. The new weights w_{ik}^l provide the same static nonlinear mapping as the original structure. Thus,

$$w_{ik}^l = c_{ik}^l\left[\frac{\sum_{j=0}^{n_b} b_{ikj}^l}{1-\sum_{j=1}^{n_a} a_{ikj}^l}\right] \tag{20}$$

Note that the new IIR synapse is of order $(\hat{n_b}, \hat{n_a})$, compared with (n_b, n_a) for each IIR synapse originally. For the new structure we have

$$\hat{y}_{ik}^l(t) = w_{ik}^l(t)\,\hat{z}_i^{l-1}(t) \tag{21}$$

$$\hat{z}_i^{l-1}(t) = \hat{c}_i^l(t)\left[\frac{\hat{B}_i^l(q^{-1})}{\hat{A}_i^l(q^{-1})} z_i^{l-1}(t)\right] \qquad (i=1) \tag{22}$$

where $\hat{y}_{ik}^l(t)$ is the synapse output, $\hat{B}_i^l(q^{-1})$ and $\hat{A}_i^l(q^{-1})$ are as defined above, z^{l-1} is the input to the layer as before, and $\hat{z}^{l-1}$ is the output of the new IIR synapse. The same results will clearly hold for an FIR MLP also. □

Lemma 3.2 *Any layer of an IIR MLP defined by*

$$z_k^l(t) = f\left[x_k^l(t)\right] \tag{23}$$

$$x_k^l(t) = \sum_{i=1}^{N_{l-1}} c_{ik}^l(t)\left[\frac{B_{ik}^l(q^{-1})}{A_{ik}^l(q^{-1})} z_i^{l-1}(t)\right] \tag{24}$$

which has a common set of dominant poles and zeros, can be modelled by two layers of reduced complexity, consisting of a single layer of static feedforward weights (no time delays) followed by a single IIR synapse, defined by

$$z_k^l(t) = f\left[\hat{c}_k^l\left[\frac{\hat{B}_k^l(q^{-1})}{\hat{A}_k^l(q^{-1})}\hat{z}_k^l(t)\right]\right] \tag{25}$$

$$\hat{z}_k^l(t) = \sum_{i=1}^{N_{l-1}} w_{ik}^l z_i^{l-1}(t) \tag{26}$$

where $\hat{c}_k^l\left[\frac{\hat{B}_k^l(q^{-1})}{\hat{A}_k^l(q^{-1})}\right]$ *is a single IIR synapse containing the dominant poles and zeros, and* w_{ik}^l *are the static feedforward weights. The new static weights are defined by*

$$w_{ik}^l = c_{ik}^l\left[\frac{\sum_{j=0}^{n_b} b_{ikj}^l}{1-\sum_{j=1}^{n_a} a_{ikj}^l}\right] \tag{27}$$

$$\hat{c}_k^l = \frac{1-\sum_{j=1}^{\hat{n_a}} \hat{a}_{kj}^l}{\sum_{j=0}^{\hat{n_b}} \hat{b}_{kj}^l} \tag{28}$$

Proof. The proof follows the same principles as Lemma 3.1. Let the static gain of each synapse in the original layer be

$$w_{ik}^l = c_{ik}^l\left[\frac{\sum_{j=0}^{n_b} b_{ikj}^l}{1-\sum_{j=1}^{n_a} a_{ikj}^l}\right] \tag{29}$$

Hence form a new layer of synapses defined by

$$\hat{z}_k^l(t) = \sum_{i=1}^{N_{l-1}} w_{ik}^l z_i^{l-1}(t) \tag{30}$$

Now form a new IIR synapse which retains the dominant poles and zeros, common to each of the previous synapses:

$$z_k^l(t) = f\left[\hat{c_k^l}\left[\frac{\hat{B}_k^l(q^{-1})}{\hat{A}_k^l(q^{-1})}\hat{z}_k^l(t)\right]\right] \tag{31}$$

where $\hat{c}_k^l\left[\frac{\hat{B}_k^l(q^{-1})}{\hat{A}_k^l(q^{-1})}\right]$ is the new IIR synapse, and $\hat{c}_k^l$ is the new synaptic gain which is adjusted to give an IIR synapse gain of one, ie

$$\hat{c}_k^l = \frac{1-\sum_{j=1}^{\hat{n_a}} \hat{a}_{kj}^l}{\sum_{j=0}^{\hat{n_b}} \hat{b}_{kj}^l} \tag{32}$$

Thus, in the case where each of the N_{l-1} IIR synapses in layer l have a common set of poles and zeros, then they can be replaced by a layer of static weights followed by a single IIR synapse. The same results will also hold for FIR MLPs. □

These lemmas are used in proving the following results for the modelling capabilities of FIR and IIR MLPs.

Theorem 3.1 *There exists an SISO IIR MLP, which is an equivalent model to a Wiener system defined by*

$$y(t) = g(H(q^{-1})x(t)) \tag{33}$$

where $g(x)$ is a nonlinear function, and $H(q^{-1}) = \frac{\hat{D}(q^{-1})}{\hat{C}(q^{-1})}$.

Proof. Let $L \geq 2$, and allow only the first layer to have IIR synapses (all higher layers have static weights). From Lemma 3.1, if there exists a dominant set of poles and zeros across the first layer, the first layer can be described by the following:

$$z_k^1(t) = f\left[\hat{x}_k^1(t)\right] \tag{34}$$

$$\hat{x}_k^1(t) = \hat{y}_{1k}^1(t) \tag{35}$$

$$\hat{y}_{1k}^1(t) = c_{1k}^1(t)\left[\frac{B_{1k}^1(q^{-1})}{A_{1k}^1(q^{-1})} z_1^0(t)\right] \qquad k = 1, 2, ..., N_l \tag{36}$$

The higher layers are given by

$$z_m^{l+1}(t) = f\left[\hat{x}_m^{l+1}(t)\right] \qquad m = 1, 2, ..., N_{l+1} \tag{37}$$

$$\hat{x}_m^{l+1}(t) = \sum_{k=1}^{N_l} \hat{y}_{km}^{l+1}(t) \tag{38}$$

$$\hat{y}_{km}^{l+1}(t) = c_{km}^{l+1}(t) z_k^l(t) \qquad k = 1, 2, ..., N_l \tag{39}$$

Hence, the layers for which $l \geq 2$, form a static MLP which can approximate $g(x)$, the static nonlinearity in the Wiener system, provided a sufficient number of units exist in the hidden layers. The first layer consists of a linear IIR filter, which, if of sufficient order, can represent $H(q)$, the linear part of a Wiener system (as in (33)). Thus a SISO IIR MLP with $L \geq 2$, and overparameterized or exact order synapses in the first layer, can represent a Wiener system as defined in (33). □

Theorem 3.2 *There exists an SISO IIR MLP which is an equivalent model to a Hammerstein system defined by*

$$y(t) = H(q^{-1})g(x(t)) \tag{40}$$

where $g(x)$ is a nonlinear function, and $H(q^{-1}) = \frac{\hat{D}(q^{-1})}{\hat{C}(q^{-1})}$.

Proof. The proof follows from Lemma 3.2. As before let $L \geq 2$, and for $l \leq L$, let all the synapses consist only of static weights. That is,

$$\frac{B_k^l(q^{-1})}{A_k^l(q^{-1})} = 1.0 \qquad l = 1, 2, ..., L-1 \tag{41}$$

Thus, from Lemma 3.2, layer L can be described by

$$z_k^L(t) = f\left(c_k^L(t)\left[\frac{B_k^L(q^{-1})}{A_k^L(q^{-1})}\hat{z}_k^L(t)\right]\right) \tag{42}$$

$$\hat{z}_k^L(t) = \sum_{i=1}^{N_{L-1}} w_{ik}^L(t) z_i^{L-1}(t) \tag{43}$$

where $\hat{c}_k^L\left[\frac{\hat{B}_k^L(q^{-1})}{\hat{A}_k^L(q^{-1})}\right]$ is the new IIR synapse, $\hat{w}_k^L$ is the new static weights defined in Lemma 3.1. In this case the output is linear, thus, $f(\alpha) = \alpha$ for $l = L$.

Hence, the layers for $l \leq L$, form a static MLP capable of approximating the static nonlinearity in the Hammerstein system $g(x)$ due to the proofs in [6], [7]. Layer L contains a linear IIR transfer function. Given sufficient order, this synapse can model $H(q^{-1})$ in the Hammerstein system. Thus a SISO IIR MLP with $L \geq 2$ and synapses which are overparameterized or of exact order, can model a Hammerstein system as given by (10). □

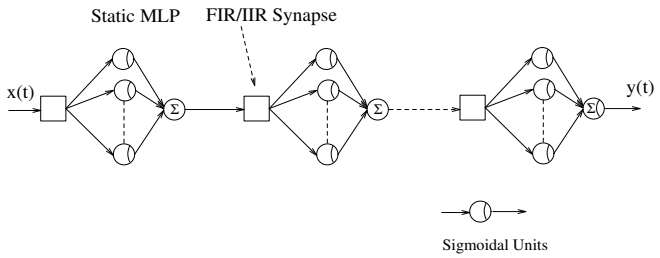

Figure 2: A Reduced Complexity IIR MLP modelling a cascade nonlinear system.

Theorem 3.3 *There exists an SISO IIR MLP which can represent a Cascade Wiener-Hammerstein system defined by either of the following forms*

$$y(t) = f_n(H_n(q^{-1})f_{n-1}(H_{n-1}(q^{-1})(\cdots f_1(H_1(q^{-1})x(t))))) \tag{44}$$

$$y(t) = H_n(q^{-1})f_n(H_{n-1}(q^{-1})f_{n-1}(\cdots f_1(H_1(q^{-1})x(t)))) \tag{45}$$

provided $L \geq 2n$.

Proof. The proof follows directly from Theorems 3.1, and 3.2, and is obtained by allowing successive layers to represent the components of the Cascade model. □

Figure 2 shows a reduced complexity network modelling a cascade system [4] given by

$$y_p(t) = g_1\left(\frac{B_1(q^{-1})}{A_1(q^{-1})}g_2\left(\frac{B_2(q^{-1})}{A_2(q^{-1})}\cdots g_n\left(\frac{B_n(q^{-1})}{A_n(q^{-1})}x(t)\right)\right)\right) \tag{46}$$

The computational complexity for the full IIR MLP is $O(n_a(n_a+n_b)N_s)$, where N_s is the total number of synapses. This is reduced to $O(n_a(n_a+n_b)+N_s)$ in the reduced complexity case. It can be noted that the network is similar to recent system identification models developed by Leithead [10].

An additional model structure of interest is the Uryson model [5] shown in Figure 3. In the following theorem it is proven that a standard IIR MLP can represent a Uryson model by using a network for each pathway, but cannot represent it directly. That is, the number of pathways needs to be known, and one IIR MLP is used for each.

Theorem 3.4 *A single IIR MLP as defined in Definition 2.1, is not sufficient to guarantee represention of a Uryson model defined by*

$$y(t) = \sum_{i=1}^{M_u} H_i(q^{-1})g_i(x(t)) \tag{47}$$

Proof. To model a Uryson system with an IIR MLP, we need to show that M_u independent pathways can exist within the network. Assume we have a multilayer network with input vector $X = [x_1, x_2, ..., x_{M_u}]^T$, and output vector $Y = [y_1, y_2, ..., y_{M_u}]^T$ (the outputs are summed). We require $y_i = g_i(x_i)$ (independent of x_j, $j \neq i$), where $g_i(.)$ is a nonlinear function.

Consider any output y_k which requires a number of hidden units (> 1) to form a particular nonlinear mapping $g_k(.)$. In using any hidden units other than the kth one, effects will be introduced inputs other than x_k. This contradicts the requirements that $y_i = g_i(x_i)$ (independent of x_j, $j \neq i$), and cannot occur unless the weights from those inputs are zero. This cannot be true however, since the same conditions must also apply to the other outputs.

Thus, the MLP architecture cannot be guaranteed to represent indepedent mappings $y_1 = g_1(x_1)$, $y_2 = g_2(x_2)$, ..., $y_{M_u} = g_{M_u}(x_{M_u})$ regardless of the type of synapse used. The

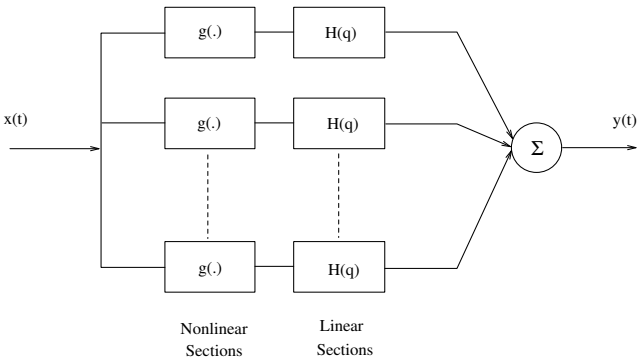

Figure 3: The Uryson nonlinear system.

only situation in which this *may* apply, is when the nonlinear mapping of each pathway can be represented by a single hidden unit (which is an unlikely occurrence). To approximate the Uryson model structure (equation (11)), each separate pathway therefore needs to be modelled by an IIR MLP (Theorem 3.2 is applicable to each network pathway). Note that a single IIR MLP *may* learn to model a specific Uryson system, but this cannot be guaranteed for all cases. □

Corollary 3.1 *A Uryson model of M_u pathways can be represented by a summation of M_u IIR MLPs.*

Proof. The proof follows directly from Theorems 3.2, and 3.4. □

NONLINEAR SYSTEM IDENTIFICATION: SIMULATION EXAMPLES

Hammerstein System Identification

To investigate the modelling properties of the network, simulations were carried out on Hammerstein, and Wiener systems. The Hammerstein system modelled was of the form

$$y(t) = \frac{0.0154 + 0.0462q^{-1} + 0.0462q^{-2} + 0.0154q^{-3}}{1 - 1.99q^{-1} + 1.572q^{-2} - 0.4583q^{-3}} \sin(x(t)) \tag{48}$$

An IIR MLP of 2 layers, 10 hidden units and synaptic orders of $(n_b, n_a) = (7, 6)$ (hidden layer), and $(n_b, n_a) = (0, 0)$ (output layer) was used to model the system. A reduced complexity network was used, with the learning algorithm presented in [1].

After convergence, the model was tested by applying a white noise input signal. It is observed that the model has learnt to approximate the plant with reasonable accuracy, though further improvements could be obtained using additional hidden units or slower learning rates.

In the next section, a more difficult identification problem is considered, in which a Wiener system is modelled by an IIR MLP.

Wiener System Identification

In the example considered here, the Wiener system is described by

$$y_p(t) = \sin\left(\frac{0.0154 + 0.0462q^{-1} + 0.0462q^{-2} + 0.0154q^{-3}}{1 - 1.99q^{-1} + 1.572q^{-2} - 0.4583q^{-3}} x(t)\right) \tag{49}$$

An IIR MLP of 2 layers, 10 hidden units and synaptic orders of $(n_b, n_a) = (6, 6)$ (hidden layer), and $(n_b, n_a) = (0, 0)$ (output layer) was used to model the system. A linear model of

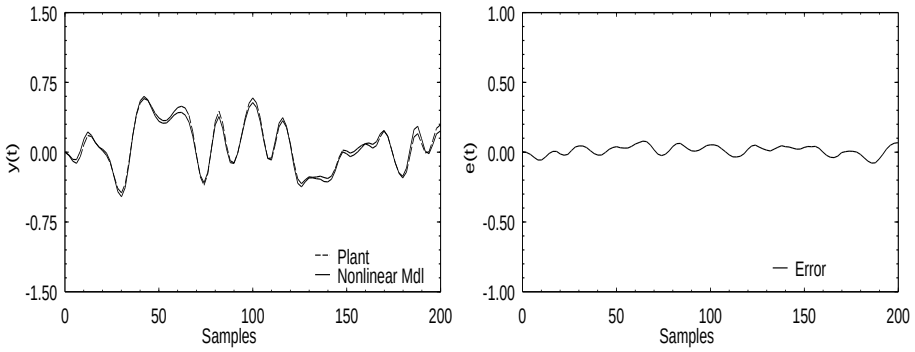

Figure 4: IIR MLP performance in modelling a Hammerstein system.

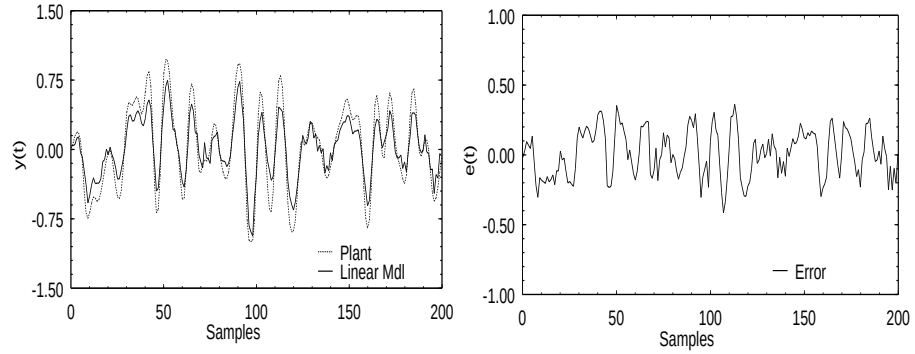

Figure 5: Identification performance for a Wiener system using a linear IIR model.

the same order was trained also to compare peformances. The linear system used the gradient descent rule described by White [15] (this is also known as Stearn's rule [16]).

After convergence, each system was validated by applying a white gaussian input signal. Referring to Figure 5, it is seen that the linear model seems to approximate the plant reasonably well, although the error plot shows large differences. The performance of the nonlinear model is significantly better. The small output error for the gaussian noise input, indicates the resulting model has good generalization properties (Figure 6).

These simulations indicate that IIR MLP is capable of accurately modelling both the linear and nonlinear components of systems such as the Wiener and Hammerstein models.

CONCLUSIONS

It has been proven in this paper that a multilayer perceptron with IIR synapses can represent a class of nonlinear block-oriented systems [4]. This includes the well known Wiener, Hammerstein and Cascade or Sandwich systems. Previous methods used to model these systems such as the Volterra series representation are known to be extremely inefficient [13], and so the IIR MLP presents an effective method of modelling block-oriented nonlinear systems. This was demonstrated by simulations on two models within the class.

The significance of the IIR MLP is that it demonstrates that a useful range of systems can be modelled by a network architecture based on the multilayer perceptron and adaptive linear filters.

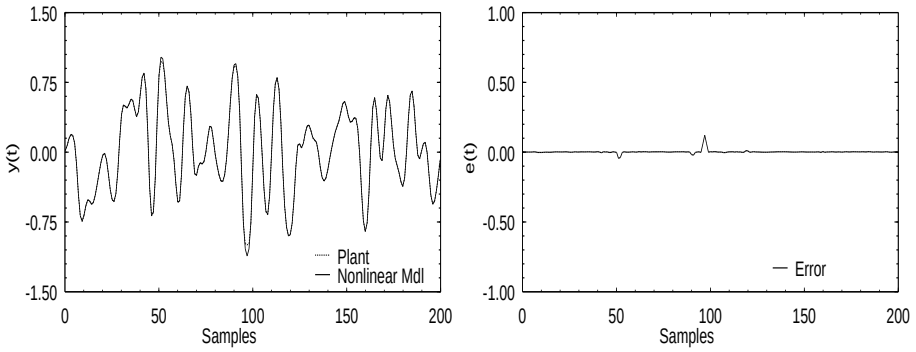

Figure 6: Identification performance for a Wiener system using an IIR MLP.

References

[1] A.D. Back, and A.C. Tsoi, "FIR and IIR Synapses, a New Neural Network Architecture for Time Series Modelling". *Neural Computation,* Vol. 3, No. 3, pp. 375-385, 1991.

[2] A.D. Back, and A.C. Tsoi, "Analysis of Hidden Layer weights in a Dynamic Locally Recurrent Network", *Artificial Neural Networks*, T. Kohonen, K. Makisara, O.Simula, and J. Kangas (Eds), Elsevier Science Publishers B.V. (North-Holland), Vol. 1, pp. 961-966, 1991.

[3] A.D. Back, and A.C. Tsoi, "Internal Representation of Data in Multilayer Perceptrons with IIR Synapses", *Proc. IEEE Int. Symp. Circuits and Systems*, San Diego CA, 1991.

[4] S. Billings, "Identification of Nonlinear Systems - A Survey", *IEE Proc.*, Vol. 127, Pt. D, No. 6, pp. 272-285, 1981.

[5] Gallman P.G., "An Iterative Method for the Identification of Nonlinear Systems Using a Uryson Model", *IEEE Trans. Automat. Contr.,* pp. 771-775, 1975.

[6] R. Hecht-Nielsen, "Theory of the Backpropagation Neural Network", *International Joint Conference on Neural Networks*, Washington, Vol. I, pp. 593-605, 1989.

[7] K. Hornik, M. Stinchcombe, and H. White, "Multilayer Feedforward Networks are Universal Approximators", *Neural Networks*, 2, pp. 359-366, 1989.

[8] C.J. Johnson, "Adaptive IIR Filtering: Current Results and Open Issues", *IEEE Trans. Inform. Theory*, vol. 30, no. 2, pp. 237-250, 1984.

[9] A. Lapedes, R. Farber, "Nonlinear Signal Processing using Neural Networks: Prediction and System modelling", Tech Report LA-UR87-2662, Los Alamos National Laboratory, 1987.

[10] W. Leithead, "Systematic Approach to Linear Approximation of Non-Linear Stochastic Systems", *Int. J. Contr*, Vol. 51, No. 1, pp. 71-91, 1990.

[11] Ljung L., and T. Soderstrom, "Theory and Practice of Recursive Identification", MIT Press: Cambridge, MA, 1983.

[12] K.S. Narendra, and P.G. Gallman, "An Iterative Method for the Identification of Systems using a Hammerstein Model, *IEEE Trans. on Automatic Control*, pp. 546-550, 1966.

[13] J.P. Norton, "Introduction to Identification", Academic Press, 1986.

[14] E. Wan, "Temporal backpropagation for FIR neural networks", *Proc. Int. Joint Conf. Neural Networks*, San Diego, Vol. I, pp. 575-580, June 1990.

[15] S.A. White, "An Adaptive Recursive Digital Filter", *Proc. Ninth Asilomar Conf. Circuits, Systems and Computers*, Pacific Grove CA, pp. 21-25, 1975.

[16] H. Fan, "New Adaptive IIR Filtering Algorithms", University of Illinois at Urbana-Champaign, PhD Thesis, 1986.