Article ID: 1000 - 8152(2001)06 - 0944 - 05

Identification of Nonlinear Systems Using Recurrent Neural Networks *

REN Xuemei

(Department of Automatic Control, Beijing Institute of Technology · Beijing, 100081, P. R. China)

Abstract: This paper proposes a new type of recurrent neural network for the identification of a class of unknown nonlinear system. It is proved that the proposed network with appropriate conditions can represent unknown input-output relationship of nonlinear systems. The dynamic backpropagation algorithm is employed to estimate the weights of both the feedforward and feedback connections in the networks. The proposed schemes have been successfully applied to modeling nonlinear plants.

Key words: recurrent neural network; dynamic backpropagation algorithm; system identification

Document code: A

非线性系统的回归网络辨识

任雪梅

(北京理工大学自动控制系·北京,100081)

摘要: 针对未知非线性系统的辨识问题,本文提出了一种新型的回归网络模型.证明了该网络模型在一定条件下能够逼近非线性系统的输入输出关系,提出了用于训练网络前向连接和反向连接权值的动态反向传播算法.仿真结果验证该方法的有效性.

关键词: 回归网络; 动态反向传播算法; 系统辨识

1 Introduction

In recent years, more and more attention has been paid to recurrent neural networks in the identification of dynamic systems since they have a dynamic memory not found in feedforward networks. The Elman network[1] is an important recurrent neural netwrok, which can model dynamic systems without requiring external feedback lines. Scott and Ray[2], Pham and Oh[3] demonstrated the performance of the Elman network for nonlinear process modeling. However, in the Elman network, only the weights of feedforward connections are trained by means of static backpropagation, while those of the feedback connections are set to be constant. With the objective of adding dynamic memory capability of the Elman network, this paper proposes a new type of recurrent neural network for system identification. The proposed network is an extended version of the Elman network. It is shown that the modified networks with proper neurons in the context layer have the capability to approximate unknown nonlinear systems. In the training of the network, a new dynamic backpropagation algorithm is proposed to achieve a faster learning. Simulation results demonstrate the proposed network has a better capability than the Elman network in modeling a complex system.

2 Proposed recurrent neural networks

The structure diagram of a modified Elman network is shown in Fig. 1, which has three layers including the input layer, the hidden layer and the output layer. The input layer is composed of two different groups of neurons, that is, the external input neurons and the internal neurons (also named context units). The modified Elman network has feedback connections from the outputs of the hidden units to all units in the context layer. Moreover, the activation function in the nonlinear hidden layer can be chosen as the hyperbolic tangent function

$$\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}},$$

while all neurons in the other layers have linear activations.

* Foundation item: supported by the National Natural Science Foundation of China (69804001).
Received date: 2000 - 02 - 24; Revised date: 2000 - 11 - 08.

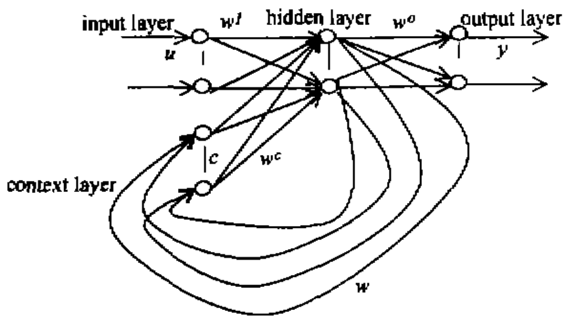

Fig. 1 The proposed recurrent neural network

Let $u(k) \in \mathbb{R}^p$ and $\hat{y}(k) \in \mathbb{R}^q$ denote the network input and output vectors at the discrete time k, respectively. Let $c(k) \in \mathbb{R}^s$ be the output of the context layer and $h(k) \in \mathbb{R}^l$, the output of the hidden layer. The proposed network can be described as

$$\begin{cases} \hat{y}(k+1) = w^o h(k), \\ h(k) = \sigma(w^c c(k) + w^I u(k)), \\ c(k) = w^h h(k-1), \end{cases} \tag{1}$$

where $w^o \in \mathbb{R}^{q\times l}, w^c \in \mathbb{R}^{l\times s}, w^I \in \mathbb{R}^{l\times p}, w^h \in \mathbb{R}^{s\times l}$ are weight matrices. The function $\sigma: \mathbb{R}^l \to \mathbb{R}^l$ is denoted by

$$\sigma((x_1,\cdots,x_l)^{\mathrm{T}}) = (\sigma(x_1),\cdots,\sigma(x_l))^{\mathrm{T}}.$$

The proposed network becomes the Elman network if $s = l$ and the feedback matrix w^h is a unit matrix. Since arbitrary connections can be allowed from the hidden layer to the context layer, the proposed network has more degrees of freedom to represent dynamic systems. In the sense of memory capability, the proposed recurrent network has more space than the Elman netwrok.

From (1), the relationship between the network output $\hat{y}(k)$ and state $c(k)$ is derived by

$$\begin{cases} \hat{y}(k+1) = w^o\sigma[w^c c(k) + w^I u(k)], \\ c(k) = w^h\sigma[w^c c(k-1) + w^I u(k-1)]. \end{cases} \tag{2}$$

The above equation can be rewritten as

$$\begin{cases} \hat{y}(k+1) = \psi[c(k), u(k)], \\ c(k) = \phi[c(k-1), u(k-1)], \end{cases} \tag{3}$$

where the mappings $\psi: \mathbb{R}^{s+p} \to \mathbb{R}^q, \phi: \mathbb{R}^{s+p} \to \mathbb{R}^s$ are nonlinear smooth functions. Since $\phi(0,0) = 0$, the origin is an equilibrium state of (2). From (3), we have

$$\begin{cases} \hat{y}(k+1) = \psi[c(k),u(k)] = \psi_1[c(k),u(k)], \\ \hat{y}(k+2) = \psi[\phi(c(k),u(k)),u(k+1)] = \\ \qquad \psi_2[c(k),u(k),u(k+1)], \\ \qquad \vdots \\ \hat{y}(k+s) = \psi\cdot\phi^{s-1}[\cdots] = \\ \qquad \psi_s[c(k),u(k),\cdots,u(k+s-1)], \end{cases} \tag{4}$$

where $\phi^{s-1}[\cdots]$ is an $(s-1)$-times iterated composition of ϕ. Denoting

$$\hat{Y}_s(k) = [\hat{y}(k+1),\cdots,\hat{y}(k+s)]^{\mathrm{T}}$$

and

$$U_s(k) = [u(k),u(k+1),\cdots,u(k+s-1)]^{\mathrm{T}},$$

(4) is of the form

$$\hat{Y}_s(k) = \Psi[c(k), U_s(k)], \tag{5}$$

where $\Psi = [\psi_1,\cdots,\psi_s]^{\mathrm{T}}$. If the Jacobian of Ψ with respect to c is nonsingular at $c = 0, U_s = 0$, by using the implicit function theorem, then there exists a unique local solution of (5) as follows

$$c(k) = \xi[\hat{Y}_s(k),\ U_s(k)] \tag{6}$$

in a neighborhood of the equilibrium state $c = 0, U_s = 0$, where $\xi: \mathbb{R}^{s(p+q)} \to \mathbb{R}^s$ is a smooth function. By (3), the state $c(k+s)$ depends on the state $c(k)$ and the input sequence $U_s(k) = [u(k),u(k+1),\cdots,u(k+s-1)]^{\mathrm{T}}$, this leads to the input-output representation of the network given by

$$\hat{y}(k+1) = G[\hat{y}(k),\cdots,\hat{y}(k-s+1),u(k),\cdots,u(k-s+1)], \tag{7}$$

where $G: \mathbb{R}^{s(p+q)} \to \mathbb{R}^q$ is a smooth function.

3 System identification

Consider the nonlinear dynamic systems described by

$$y(k+1) = F[y(k),y(k-1),\cdots,y(k-n+1), u(k),u(k-1),\cdots,u(k-m+1)], \tag{8}$$

where $u(k) \in \mathbb{R}^q$ and $y(k) \in \mathbb{R}^q$ are the control and output, respectively. The mapping $F: \mathbb{R}^{nq+mp} \to \mathbb{R}^q$ is a continuous function. For notational convenience, define

$$Y(k) = [y(k),\cdots,y(k-n+1)]^{\mathrm{T}},$$
$$U(k) = [u(k),\cdots,u(k-m+1)]^{\mathrm{T}}.$$

Here we consider the case that $n, m (n \geqslant m)$ are known, but the nonlinear function $F(\cdot)$ is unknown. As we know, multi-layer neural networks have been widely used in the identification of dynamic systems because they can approximate "well-behaved" any nonlinear function to any desired accuracy. The following fundamental approximation theorem is proved by Funahashi[4]:

Theorem 1 Let Ω be a compact subset of $\mathbb{R}^n$ and $f: \Omega \to \mathbb{R}^m$ be a continuous mapping. Then, for an arbitrary $\epsilon > 0$, there exist an integer N, an $m \times N$ matrix

A, an $N \times n$ matrix B and an N dimensional vector Θ such that

$$\max_{x \in \Omega} \| f(x) - A\sigma(BX + \Theta) \| < \varepsilon$$

holds, where $\sigma: \mathbb{R}^N \to \mathbb{R}^N$ is a sigmoid function.

In the following we shall show the proposed network (1) can represent the nonlinear system (8). For this purpose, we restate the basic points of the theory of dynamic systems, which are used to estabilish our main theorem.

Let D be an open subset of $\mathbb{R}^{nq+mp}$. A mapping $F: D \to \mathbb{R}^q$ is said to be Lipschitz on D if there exists a constant L such that

$$\| F(x) - F(y) \| \leqslant L \| x - y \|,$$

for all $x, y \in D$. We call L a Lipschitz constant for F. We also call F locally Lipschitz if each point of D has a neighborhood D_0 in D such that the restriction $F \mid D_0$ is Lipschitz.

Lemma 1[5] Let a mapping $F: D \to \mathbb{R}^q$ be continuous. Then F is locally Lipschitz. Moreover, if $A \subset D$ is compact, then the restriction $F \mid A$ is Lipschitz.

Lemma 2 Let $F, \hat{F}: D \to \mathbb{R}^q$ be Lipschitz continuous mappings. For any $\varepsilon < 0$, suppose that $\| F(x) - \hat{F}(x) \| < \varepsilon$ for all $x \in D$. Consider the following systems

$$y(k+1) = F[Y(k), U(k)], \tag{9}$$

and

$$\begin{cases} \hat{y}(k+1) = \hat{F}[\hat{Y}(k), U(k)], \\ \hat{Y}(k) = [\hat{y}(k), \cdots, \hat{y}(k-n+1)]^{\mathrm{T}}, \end{cases} \tag{10}$$

with the same initial condition $[Y(0), U(0)] = [\hat{Y}(0), U(0)] \in D$. Denote L being a Lipschitz constant of F, then there exists a positive constant $\Gamma = \Gamma(L)$ such that

$$\| y(k+1) - \hat{y}(k+1) \| \leqslant \Gamma(L)\varepsilon, \ k = 0, 1, \cdots, M.$$

Proof From (9) and (10), it follows that

$$\begin{aligned} & y(k+1) - \hat{y}(k+1) = \\ & F(Y(k), U(k)) - F(\hat{Y}(k), U(k)) + \\ & F(\hat{Y}(k), U(k)) - \hat{F}(\hat{Y}(k), U(k)). \end{aligned} \tag{11}$$

Taking norms in both sides of equality (11), we have

$$\begin{aligned} & \| y(k+1) - \hat{y}(k+1) \| \leqslant \\ & L \| Y(k) - \hat{Y}(k) \| + \\ & \| F(\hat{Y}(k), U(k)) - \hat{F}(\hat{Y}(k), U(k)) \|. \end{aligned} \tag{12}$$

Since $\| F(x) - \hat{F}(x) \| < \varepsilon$ for all $x \in D$, it follows that

$$\begin{aligned} & \| y(k+1) - \hat{y}(k+1) \| \leqslant \\ & L \| Y(k) - \hat{Y}(k) \| + \varepsilon \leqslant \\ & L \| y(k) - \hat{y}(k) \| + \cdots + L \| y(k - \\ & n+1) - \hat{y}(k-n+1) \| + \varepsilon. \end{aligned} \tag{13}$$

After straight manipulations, by the assumption $Y(0) = \hat{Y}(0)$, there exists a positive constant $\Gamma(L)$ such that

$$\| y(k+1) - \hat{y}(k+1) \| \leqslant \Gamma(L)\varepsilon,$$

for $k \in [0, 1, \cdots, M]$. The proof is completed.

Theorem 2 Let D be an open subset of $\mathbb{R}^{nq+mp}$, $F: D \to \mathbb{R}^q$ be a continuous mapping, and A be a compact subset of D. Given any $\varepsilon > 0$, there exists a proposed network (1) with the number of context units equal to the order of the system (8) such that

$$\max_{k \in [0, 1, \cdots, M]} \| y(k+1) - \hat{y}(k+1) \| < \varepsilon$$

holds, where $y(k+1), \hat{y}(k+1)$ are outputs of the plant and network, respectively.

Proof For given $\varepsilon > 0$, choose η_1 such that $0 < \eta_1 < \min(\varepsilon, \lambda)$, where λ is the distance between A and the boundary ∂D of D. Define

$A_\eta = \{X: X \in \mathbb{R}^{nq+mp}, \ \exists z \in A, \| X - z \| \leqslant \eta_1\}$. Since A is compact, A_η is a compact subset of D. Hence, F is Lipschitz on A_η by Lemma 1. Denote L_F being a Lipschitz constant of $F \mid A_\eta$. We may choose $\varepsilon_1 > 0$ such that

$$\varepsilon_1 < \frac{\eta_1}{\Gamma(L)}. \tag{14}$$

Since the function F is continuous, by Theorem 1, for any $\varepsilon_1 > 0$, there exist an integer N and matrices $A_1 \in \mathbb{R}^{q \times N}, B_1 \in \mathbb{R}^{N \times nq}, B_2 \in \mathbb{R}^{N \times mp}$ and an N dimensional vector Θ such that

$$\begin{aligned} & \max_{[Y(k), U(k)] \in A_\eta} \| F(Y(k), U(k)) - \\ & A_1\sigma(B_1 Y(k) + B_2 U(k) + \Theta) \| < \varepsilon_1. \end{aligned}$$

Now, define a mapping $G: \mathbb{R}^{nq+mp} \to \mathbb{R}^q$:

$$G(\hat{Y}(k), U(k)) = A_1\sigma(B_1 \hat{Y}(k) + B_2 U(k) + \Theta), \tag{15}$$

such that the following inequality is satisfied

$$\max_{X \in A_\eta} \| F(X) - G(X) \| < \varepsilon_1.$$

According to the above discussion in Section 2, under the proper conditions, the input-output representation of the network (1) can be written as (7). In the following we consider the outputs of the following systems

$$y(k+1) = F(Y(k), U(k)), \tag{16}$$

and

$$\hat{y}(k+1) = G(\hat{Y}(k), U(k)) =$$

$$A_{1\sigma}(B_1\hat{Y}(k) + B_2U(k) + \Theta) \qquad (17)$$

with the same initial condition $[Y(0), U(0)] = [\hat{Y}(0), U(0)] \in A_\eta$, it can follow from Lemma 2 that $\max_{k\in[0,1,\cdots,M]} \| y(k+1) - \hat{y}(k+1) \| < \eta_1 < \varepsilon$ is satisfied.

The proof is completed.

Theorem 2 tells us that the proposed network (1) can be used to learn the input-output behaviour of dynamic system (8). In many cases, we do not know the order of the plant, the number of the context units can be set sufficiently large to meet this assumption.

To update the weights of the networks, we define the training objective function as

$$J = \frac{1}{2}\sum_{k=0}^{M} e^{\mathrm{T}}(k+1)e(k+1), \qquad (18)$$

where M represents the total number of data patterns used in the training process, $e(k+1) = y(k+1) - \hat{y}(k+1)$ is the error between the network and plant outputs at time $k+1$. Using the backpropagation algorithm, the weights of the neural network are adjusted as follows

$$w_{i+1} = w_i - \eta_i \frac{\partial J}{\partial w_i} + \alpha \Delta w_i, \qquad (19)$$

where i is the iteration number, η_i is a learning rate, α is a momentum factor, Δw_i represents the change in weight in ith iteration. Here, we use an adapted value of η_i as follows:

$$\eta_{i+1} = \begin{cases} 1.05\eta_i, & \text{if } J_i \leqslant 0.99J_{i-1}, \\ 0.85\eta_i, & \text{if } J_i \geqslant 1.02J_{i-1}, \\ \eta_i, & \text{otherwise} \end{cases} \qquad (20)$$

η_0 is chosen to lie between 0 and 1. From (18), we can obtain

$$\frac{\partial J}{\partial w_i} = -\sum_{k=0}^{M} e^{\mathrm{T}}(k+1)\frac{\partial \hat{y}(k+1)}{\partial w_i}.$$

In order to compute the gradient $\frac{\partial \hat{y}(k+1)}{\partial w}$, the mathematical model (1) for the proposed network can be rewritten as

$$\hat{y}_i(k+1) = \sum_j w_{ij}^0 h_j(k),$$
$$h_j(k) = \sigma(s_j(k)),$$
$$s_j(k) = \sum_r w_{jr}^c c_r(k) + \sum_d w_{jd}^I u_d(k),$$
$$c_r(k) = \sum_j w_{rj}^h h_j(k-1).$$

The output gradients with respect to w^o, w^c, w^I and w^h are given by

$$\frac{\partial \hat{y}_i(k+1)}{\partial w_{ij}^o} = h_j(k), \qquad (21)$$

$$\frac{\partial \hat{y}_i(k+1)}{\partial w_{jr}^c} = w_{ij}^o P_{W_{jr}^c}, \quad P_{W_{jr}^c} = \frac{\partial h_j(k)}{\partial w_{jr}^c}, \qquad (22)$$

$$\frac{\partial \hat{y}_i(k+1)}{\partial w_{jd}^I} = w_{ij}^o P_{W_{jd}^I}, \quad P_{W_{jd}^I} = \frac{\partial h_j(k)}{\partial w_{jd}^I}, \qquad (23)$$

$$\frac{\partial \hat{y}_i(k+1)}{\partial w_{rj}^h} = w_{ij}^o P_{W_{rj}^h}, \quad P_{W_{rj}^h} = \frac{\partial h_j(k)}{\partial w_{rj}^h}, \qquad (24)$$

where $P_{W_{jr}^c}$, $P_{W_{jd}^I}$, $P_{W_{rj}^h}$ satisfy

$$\begin{cases} P_{W_{jr}^c}(k) = \sigma'(s_j)[c_r(k) + w_{jr}^c w_{rj}^h P_{W_{jr}^c}(k-1)], \\ P_{W_{rj}^c}(0) = 0, \end{cases} \qquad (25)$$

$$\begin{cases} P_{W_{jd}^I}(k) = \sigma'(s_j)[u_d(k) + w_{jr}^c w_{rj}^h P_{W_{jd}^I}(k-1)], \\ P_{W_{jd}^I}(0) = 0, \end{cases} \qquad (26)$$

$$\begin{cases} P_{W_{rj}^h}(k) = \sigma'(s_j)[w_{jr}^c(k)h_j(k-1) + \\ \qquad w_{jr}^c w_{rj}^h P_{W_{rj}^h}(k-1)], \\ P_{W_{rj}^h}(0) = 0. \end{cases} \qquad (27)$$

4 Simulation

In this section, the ability of the proposed recurrent neural network to model nonlinear dynamic systems has been demonstrated in simulation. The initial weights are randomly set between $(-0.5, 0.5)$ and the parameters of the networks are adjusted by (18) ~ (27). The input sequence is random in the range $[-1, 1]$. To test the performance of the trained network, the network is used in recall mode to produce the response to a sinusoidal input or random input signal in the range $[-0.5, 0.5]$. The performance of the trained network is measured by

$$\mathrm{MSE} = \frac{1}{N}\sum_{k=1}^{N_2}[y(k) - \hat{y}(k)]^2,$$

where N_2 represents the number of data pairs $[u, y]$ in the recall set.

The modeling is carried out using the Elman network and the proposed network. For demonstration, the same number of units in the context layer for two types of networks is used.

Example 1 The plant is described by

$$y(k+1) = \frac{y(k)}{1.2 + y^2(k)} + 0.5u(k). \qquad (28)$$

During training, choose $\eta_0 = 0.001, \alpha = 0.95, c = 10, s = 10, i = 1000$, A random signal in the range $[-1,1]$ is used as the input to the plant and the proposed network. In the recall model, responses from the plant and the network are obtained for two kinds of signals: $u(k) = 0.62\sin(2\pi/25) + 0.5\sin(2\pi/40)$ and a random input $-0.5 < u(k) < 0.5, k = 1,2,\cdots,150$. The simulation results are shown in Figs. 2 and 3.

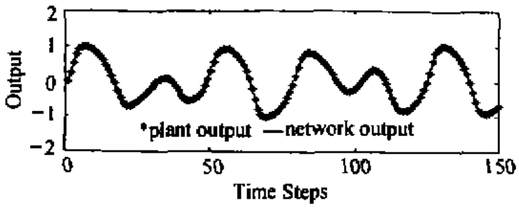

Fig. 2 Sinusoidal responses of plant (28) and the proposed network

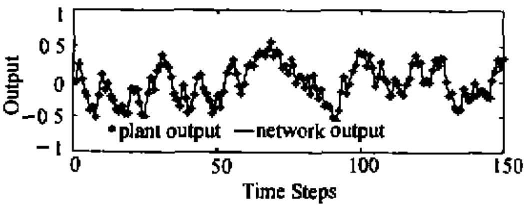

Fig. 3 Random responses of plant (28) and the proposed network

For comparison, the MSE values computed for the Elman network and the proposed network are presented in Table 1. In this example, both types of networks capture the dynamics of the given system. However, the proposed network gives a better performance.

Table 1 Performance indices of different examples

Example	Input	MSE (the Elman network)	MSE (the proposed network)
Example 1	Sinusoidal input	0.000611	0.000575
Example 1	Random input	0.000401	0.000273
Example 2	Sinusoidal input	0.1041	0.0423
Example 2	Random input	0.0233	0.0040

Example 2 The system can be found in the paper by Sales and Billings[6] and is described by

$$\begin{aligned} & y(k+1) = \\ & 0.9722y(k) + 0.3578u(k) - 0.1295u(k-1) - \\ & 0.3103y(k)u(k) - 0.04228y^2(k-1) + \\ & 0.1663y(k-1)u(k-1) + 0.2573y(k-1)e(k) - \\ & 0.03259y^2(k)y(k-1) - 0.3513y^2(k)u(k-1) + \\ & 0.3084y(k)y(k-1)u(k-1) + 0.2939y^2(k-1)e(k) + \\ & 0.1087y(k-1)u(k)u(k-1) + 0.4770y(k-1)u(k)e(k) + \\ & 0.6389u^2(k-1)e(k) + e(k+1). \end{aligned} \tag{29}$$

The model used in the experiment was identified from a laboratory scale liquid level system. The system contains a DC water pump feeding a conical flask which in turn feeds a square tank, giving the system secondorder dynamics. The input is the voltage to the pump motor and the output of the system is the height of water in the conical flask. The noise sequence $e(k)$ was gaussian and its variance is 0.05. The training input is the same as in Example 1. Let $\eta_0 = 0.003, \alpha = 0.65, c = 20, s = 20, i = 3500$. Figs. 4 and 5 show the responses of the proposed network during recall for the case where the input was, respectively, a sinusoidal function $u(k) = 0.3\sin(2\pi/15) + 0.5\sin(2\pi/25) + 0.23\sin(2\pi/k50)$ and a random input in the range $[-0.5, 0.5]$. Table 1 gives the MSE performance indices of the Elman network and the proposed network. The results show that both types of networks can model the system (29), but the proposed network has a better capability than the Elman network.

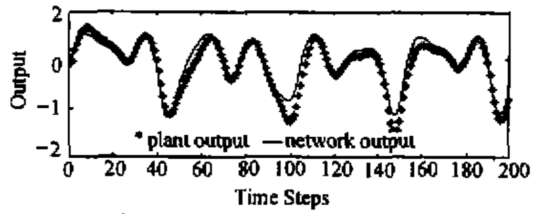

Fig. 4 Sinusoidal responses of plant (29) and the proposed network

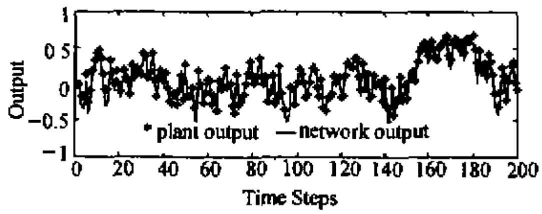

Fig. 5 Random responses of plant (29) and the proposed network

5 Conclusion

This paper investigates the use of the proposed recurrent neural networks for the identification of dynamic systems. The architecture of the proposed network is an extended version of the network described by Elman. The input-output dynamics of the proposed network is discussed in this paper. It is shown that the proposed network can represent nonlinear system under the assumption that the order of the system is equal to or less than the number of the context units. The dynamic

(Continued on page 957)

$$\dot V(t) \leqslant \sum_{i=1}^{N}\{x_i^{\mathrm T}(t)(P_iA_i + A_i^{\mathrm T}P_i + P_iB_iK_i + K_i^{\mathrm T}B_i^{\mathrm T}P_i + 2P_iD_iD_i^{\mathrm T}P_i + E_{ai}^{\mathrm T}E_{ai} + K_i^{\mathrm T}E_{bi}^{\mathrm T}E_{bi}K_i)x_i(t) + x_i^{\mathrm T}(t)P_iG_iw_i(t) + w_i^{\mathrm T}(t)G_i^{\mathrm T}P_ix_i(t) + x_i^{\mathrm T}(t)P_i\sum_{j\in J_i(A)}(A_{ij} + \Delta A_{ij}(t))P_j^{-1}H_j^{-1}P_j^{-1}(A_{ij} + \Delta A_{ij}(t))^{\mathrm T}P_ix_i(t) + \sum_{j\in J_i(A)}x_j^{\mathrm T}(t)P_jH_jP_jx_j(t)\}.$$

Using Lemma 3, we get

$$\sum_{j\in J_i(A)}(A_{ij} + \Delta A_{ij}(t))P_j^{-1}H_j^{-1}P_j^{-1}(A_{ij} + \Delta A_{ij}(t))^{\mathrm T} = \sum_{j\in J_i(A)}(A_{ij}+M_{ij}F_{ij}(t)N_{ij})P_j^{-1}H_j^{-1}P_j^{-1}(A_{ij}+M_{ij}F_{ij}(t)N_{ij})^{\mathrm T}\leqslant \sum_{j\in J_i(A)}A_{ij}(P_jH_jP_j - \varepsilon_{ij}N_{ij}^{\mathrm T}N_{ij})^{-1}A_{ij}^{\mathrm T} + \sum_{j\in J_i(A)}\varepsilon_{ij}^{-1}M_{ij}M_{ij}^{\mathrm T}.$$

Furthermore, we have

$$\dot V(t) \leqslant \sum_{i=1}^{N}\{x_i^{\mathrm T}(t)[P_iA_i + A_i^{\mathrm T}P_i+P_iB_iK_i+K_i^{\mathrm T}B_i^{\mathrm T}P_i+ 2P_iD_iD_i^{\mathrm T}P_i + E_{ai}^{\mathrm T}E_{ai} + K_i^{\mathrm T}E_{bi}^{\mathrm T}E_{bi}K_i + P_i\sum_{j\in J_i(A)}A_{ij}(P_jH_jP_j - \varepsilon_{ij}N_{ij}^{\mathrm T}N_{ij})^{-1}A_{ij}^{\mathrm T}P_i + \sum_{j\in J_i(A)}\varepsilon_{ij}^{-1}P_iM_{ij}M_{ij}^{\mathrm T}P_i]x_i(t) + x_i^{\mathrm T}(t)P_iG_iw_i(t) + w_i^{\mathrm T}(t)G_i^{\mathrm T}P_ix_i(t) + \sum_{j\in J_i(A)}x_j^{\mathrm T}(t)P_jH_jP_jx_j(t)\} =$$

$$\sum_{i=1}^{N}\{x_i^{\mathrm T}(t)[P_iA_i + A_i^{\mathrm T}P_i + P_iB_iK_i + K_i^{\mathrm T}B_i^{\mathrm T}P_i + 2P_iD_iD_i^{\mathrm T}P_i + E_{ai}^{\mathrm T}E_{ai} + K_i^{\mathrm T}E_{bi}^{\mathrm T}E_{bi}K_i + P_i\sum_{j\in J_i(A)}A_{ij}(P_jH_jP_j - \varepsilon_{ij}N_{ij}^{\mathrm T}N_{ij})^{-1}A_{ij}^{\mathrm T}P_i + \sum_{j\in J_i(A)}\varepsilon_{ij}^{-1}P_iM_{ij}M_{ij}^{\mathrm T}P_i + \tilde N_A(i)P_iH_iP_i]x_i(t) + x_i^{\mathrm T}(t)P_iG_iw_i(t) + w_i^{\mathrm T}(t)G_i^{\mathrm T}P_ix_i(t)\}.$$

First, we prove asymptotic stability of closed-loop system (3), let $w(t) = 0$, that is $w_i(t) = 0, i = 1,2, \cdots, N$, obviously $\dot V(t) < 0$ if

$$P_iA_i + A_i^{\mathrm T}P_i+P_iB_iK_i+K_i^{\mathrm T}B_i^{\mathrm T}P_i+ 2P_iD_iD_i^{\mathrm T}P_i + E_{ai}^{\mathrm T}E_{ai} + K_i^{\mathrm T}E_{bi}^{\mathrm T}E_{bi}K_i + P_i\sum_{j\in J_i(A)}A_{ij}(P_jH_jP_j - \varepsilon_{ij}N_{ij}^{\mathrm T}N_{ij})^{-1}A_{ij}^{\mathrm T}P_i + \sum_{j\in J_i(A)}\varepsilon_{ij}^{-1}P_iM_{ij}M_{ij}^{\mathrm T}P_i +\tilde N_A(i)P_iH_iP_i < 0. \tag{9}$$

Next, to establish that $\| z(t) \|_2 < \gamma \| w(t) \|_2$ whenever $w(t) \neq 0$ which implies that the desired robust H$_\infty$ performance is achieved, we introduce

$$J = \int_0^\infty(z^{\mathrm T}(t)z(t) - \gamma^2 w^{\mathrm T}(t)w(t))\mathrm dt = \sum_{i=1}^{N}\int_0^\infty[z_i^{\mathrm T}(t)z_i(t) - \gamma^2 w_i^{\mathrm T}(t)w_i(t)]\mathrm dt.$$

Note that for zero initial conditions, $V(0) = 0$, we get

$$J = \sum_{i=1}^{N}\int_0^\infty[z_i^{\mathrm T}(t)z_i(t)-\gamma^2w_i^{\mathrm T}(t)w_i(t)+ \dot V_i(t)]\mathrm dt - \sum_{i=1}^{\infty}V_i(\infty)\leqslant \sum_{i=1}^{N}\int_0^\infty[z_i^{\mathrm T}(t)z_i(t) - \gamma^2w_i^{\mathrm T}(t)w_i(t) + \dot V_i(t)]\mathrm dt \leqslant$$

$$\sum_{i=1}^{N}\int_0^\infty\{x_i^{\mathrm T}(t)[P_iA_i + A_i^{\mathrm T}P_i + \tilde N_A(i)P_iH_iP_i + P_iB_iK_i + K_i^{\mathrm T}B_i^{\mathrm T}P_i + 2P_iD_iD_i^{\mathrm T}P_i + E_{ai}^{\mathrm T}E_{ai} + K_i^{\mathrm T}E_{bi}^{\mathrm T}E_{bi}K_i + \sum_{j\in J_i(A)}P_iA_{ij}(P_jH_jP_j - \varepsilon_{ij}N_{ij}^{\mathrm T}N_{ij})^{-1}A_{ij}^{\mathrm T}P_i + C_i^{\mathrm T}C_i + \sum_{j\in J_i(A)}\varepsilon_{ij}^{-1}P_iM_{ij}M_{ij}^{\mathrm T}P_i + C_i^{\mathrm T}T_iK_i + K_i^{\mathrm T}T_i^{\mathrm T}C_i + K_i^{\mathrm T}T_i^{\mathrm T}T_iK_i]x_i(t) + x_i^{\mathrm T}(t)P_iG_iw_i(t) + w_i^{\mathrm T}(t)G_i^{\mathrm T}P_ix_i(t) - \gamma^2w_i^{\mathrm T}(t)w_i(t)\}\mathrm dt.$$

Using Lemma2, we have

$$J \leqslant \sum_{i=1}^{N}\int_0^\infty\{x_i^{\mathrm T}(t)[P_iA_i + A_i^{\mathrm T}P_i + \tilde N_A(i)P_iH_iP_i + P_iB_iK_i + K_i^{\mathrm T}B_i^{\mathrm T}P_i + 2P_iD_iD_i^{\mathrm T}P_i + E_{ai}^{\mathrm T}E_{ai} + K_i^{\mathrm T}E_{bi}^{\mathrm T}E_{bi}K_i + \sum_{j\in J_i(A)}P_iA_{ij}(P_jH_jP_j - \varepsilon_{ij}N_{ij}^{\mathrm T}N_{ij})^{-1}A_{ij}^{\mathrm T}P_i + \sum_{j\in J_i(A)}\varepsilon_{ij}^{-1}P_iM_{ij}M_{ij}^{\mathrm T}P_i + 2C_i^{\mathrm T}C_i + 2K_i^{\mathrm T}T_i^{\mathrm T}T_iK_i]x_i(t) + x_i^{\mathrm T}(t)P_iG_iw_i(t)+w_i^{\mathrm T}(t)G_i^{\mathrm T}P_ix_i(t)-\gamma^2w_i^{\mathrm T}(t)w_i(t)\}\mathrm dt = \sum_{i=1}^{N}\int_0^\infty\xi_i^{\mathrm T}(t)\Psi_i\xi_i(t)\mathrm dt.$$

where

$$\xi_i = [x_i^{\mathrm T}(t) \quad w_i^{\mathrm T}(t)]^{\mathrm T},$$

$$\Psi_i = \begin{bmatrix}\Omega_i & P_iG_i\\ G_i^{\mathrm T}P_i & -\gamma^2 I\end{bmatrix},$$

$$\Omega_i = P_iA_i + A_i^{\mathrm T}P_i + \tilde N_A(i)P_iH_iP_i + P_iB_iK_i + K_i^{\mathrm T}B_i^{\mathrm T}P_i + 2C_i^{\mathrm T}C_i + 2K_i^{\mathrm T}T_i^{\mathrm T}T_iK_i + 2P_iD_iD_i^{\mathrm T}P_i + E_{ai}^{\mathrm T}E_{ai} +$$