

Repetitive Tracking Control of Nonlinear Systems Using Reinforcement Fuzzy-Neural Adaptive Iterative Learning Controller

Ying-Chung Wang and Chiang-Ju Chien

Department of Electronic Engineering, Huafan University, Shihding, New Taipei City, Taiwan

Received: Jan. 1, 2012; Revised April 4, 2012; Accepted May 6, 2012

Published online: 1 Sep. 2012

Abstract: This paper proposes a new fuzzy neural network based reinforcement adaptive iterative learning controller for a class of nonlinear systems. Different from some existing reinforcement learning schemes, the reinforcement adaptive iterative learning controller has the advantages of rigorous proofs without using an approximation of the plant Jacobian. The critic is appended into the reinforcement adaptive iterative learning controller to generate the simple discrete reinforcement signal, which provides a satisfaction about the tracking performance. In addition, the reinforcement signal can be further applied in the weight adaptation rules. Iterative learning components of the reinforcement adaptive iterative learning controller are designed to compensate for the uncertainties of plant nonlinearities. The overall adaptive scheme guarantees all adjustable parameters and the internal signals remain bounded for all iterations. Moreover, the norm of tracking error vector at each time instant will asymptotically converge to a tunable residual set as iteration goes to infinity even the initial state error exists. Finally, a simulation result is given to demonstrate the learning performance of the fuzzy neural network based reinforcement adaptive iterative learning controller.

Keywords: Iterative learning control, reinforcement learning control, adaptive control, nonlinear systems, fuzzy neural network.

1. Introduction

The neural or fuzzy adaptive control is a kind of the control approach which is capable of handling the system whose nonlinearity is not linearly parameterizable [1]–[3]. In addition to the aforementioned works, it is worth noting that there is an interesting design configuration, i.e., the neural network or fuzzy system based reinforcement learning control with the critic (evaluation network) and the controller (action network). The objective of the neural network or fuzzy system based reinforcement learning control [4]–[9] is to discover a control policy, a mapping from states to control actions, through stochastic exploration of the space of possible control actions. Similar to human and animal, if the selected control actions result in a good learning performance, then the reinforcement learning control is reward; otherwise it should be penalized. In contrast to other forms of learning approaches, there is no supervisor who teaches the reinforcement learning control what the target actions should be taken. Therefore, only perfor-

mance measurement in terms of failure or success can be used to find which control actions produce the maximum reward by using trial and error. The reinforcement signal in [4], a discrete number, which is usually -1 means a failure and 1 or 0 to express a success. Also, a real-valued reinforcement signal in [5]–[9] that indicates a more detailed and continuous degree of failure or success. The main problem of the fuzzy neural network (FNN) based reinforcement learning control [4]–[9] is that less Lyapunov stability analysis are applied to guarantee its convergence and it is a time-consuming control scheme.

For controlling the nonlinear systems with repeated tracking control or periodic disturbance rejection, iterative learning control (ILC) has become one of the most effective control strategies. The ILC system improves the control performances by some self-tuning processes without using accurate system models and can be applied to a lot of practical applications [10]–[12] such as robotics, servo motors, etc.. One of the main design concepts for ILC algorithms, called PID-type ILC, updates the control

* Corresponding author: e-mail: ycwang@cc.hfu.edu.tw

input directly by a learning mechanism using the information of error and input in the previous iteration. Another interesting ILC algorithm, called adaptive ILC, tunes the control parameters but not the control input itself between successive iterations [13]-[15]. Recently, adaptive nonlinear compensation ILCs by using fuzzy system or neural network approximation technique were applied to iterative learning control problems [16]-[18]. However the reinforcement adaptive ILC (reinforcement AILC) which is a combination of AILC with the reinforcement learning control, has never been presented in the literature. In this paper, we aim to present the FNN based reinforcement AILC for tracking control of unknown nonlinear systems. The critic is appended into the reinforcement AILC to generate the discrete reinforcement signal, which provides a satisfaction about the tracking performance. In addition, the reinforcement signal can be further applied in the weight adaptation rules. The FNN is used as an universal approximator to design the iterative learning component. Unfortunately, the optimal FNN parameters for a best approximation is generally unavailable. They will be tuned during the iteration processes in order to guarantee stability and learning performance. Based on the Lyapunov-like analysis, we show that all adjustable parameters as well as the internal signals will remain bounded. Moreover, the norm of tracking error vector will asymptotically converge to a tunable residual set as iteration goes to infinity even there exists variable initial state errors.

This paper is organized as follows. The problem formulation will be given in section 2. Section 3 describes the construction of reinforcement learning signal. The design concept of the proposed FNN based reinforcement AILC is presented in section 4. Analysis of stability and learning performance will be studied extensively in section 5. To clarify the performance of the proposed scheme, a simulation example is presented in section 6. Finally, a conclusion is made in section 7.

2. Problem Formulation

Consider a class of nonlinear systems which can repetitively perform a control task over a finite time interval $[0, T]$ as follows

$$\begin{aligned} \dot{x}_1^j(t) &= x_2^j(t) \\ \dot{x}_2^j(t) &= x_3^j(t) \\ &\vdots \\ \dot{x}_n^j(t) &= -f(X^j(t)) + b(X^j(t))u^j(t) \end{aligned} \quad (1)$$

where $X^j(t) = [x_1^j(t), \dots, x_n^j(t)]^\top \in \mathbb{R}^{n \times 1} \times [0, T]$ is the state vector of the system, $u^j(t)$ is the control input, $f(X^j(t))$ and $b(X^j(t))$ are unknown real continuous nonlinear functions of states. Here, j and t denote the index of iteration number and time variable. Given a specified desired trajectory $X_d(t) = [x_d(t), \dot{x}_d(t), \dots, x_d^{(n-1)}(t)]^\top$, $t \in [0, T]$ and a possible initial resetting error $X_d(0) \neq$

$X^j(0)$ for all $j \geq 1$, the control objective is to force the state vector $X^j(t)$ to follow some specified desired trajectory $X_d(t)$ for all $t \in [0, T]$ as close as possible. In order to achieve the above control objective, some assumptions on the nonlinear system and desired trajectory are given as follows :

- (A1) The nonlinear functions $f(X^j(t))$ and $b(X^j(t))$ are bounded if $X^j(t)$ is bounded.
- (A2) There exists a positive but unknown lower bound b_L , such that $0 < b_L \leq b(X^j(t))$.
- (A3) The desired state trajectory $X_d(t) = [x_d(t), \dot{x}_d(t), \dots, x_d^{(n-1)}(t)]^\top$ is bounded.
- (A4) Let the state errors $e_1^j(t), \dots, e_n^j(t)$ be defined as $e_1^j(t) = x_1^j(t) - x_d(t)$, $e_2^j(t) = \dot{x}_1^j(t) - \dot{x}_d(t)$, \dots , $e_n^j(t) = x_1^{(n-1),j}(t) - x_d^{(n-1)}(t)$. The initial state errors at each iteration are not necessarily zero, small and fixed, but assumed to be bounded.

3. Reinforcement Fuzzy-Neural Adaptive Iterative Learning Controller

3.1. Construction of Reinforcement Learning Signal

In order to meet the control objective, we first define an extended error function as follows

$$s^j(t) = c_1 e_1^j(t) + c_2 e_2^j(t) + \dots + c_{n-1} e_{n-1}^j(t) + e_n^j(t) \quad (2)$$

where c_1, \dots, c_{n-1} are the coefficients of a Hurwitz polynomial $\Delta(D) = D^{n-1} + c_{n-1}D^{n-2} + \dots + c_1$. It is noted that there exists a known constant ε^j such that the initial value of $s^j(t)$ will satisfy $|s^j(0)| = |c_1 e_1^j(0) + c_2 e_2^j(0) + \dots + e_n^j(0)| \equiv \varepsilon^j$ by assumption (A4). To overcome the uncertainty from initial state errors, the performance measurement $s_\phi^j(t)$ is introduced as follows :

$$\begin{aligned} s_\phi^j(t) &= s^j(t) - \phi^j(t) \text{sat} \left(\frac{s^j(t)}{\phi^j(t)} \right) \\ \phi^j(t) &= \varepsilon^j e^{-kt}, \quad k > 0 \end{aligned} \quad (3)$$

where **sat** is the saturation function defined as

$$\text{sat} \left(\frac{s^j(t)}{\phi^j(t)} \right) = \begin{cases} 1 & \text{if } s^j(t) > \phi^j(t) \\ \frac{s^j(t)}{\phi^j(t)} & \text{if } |s^j(t)| \leq \phi^j(t) \\ -1 & \text{if } s^j(t) < -\phi^j(t) \end{cases} \quad (4)$$

and $\phi^j(t)$ is the width of boundary layer. Note that $\phi^j(t)$ is designed to decrease along time axis with the initial condition chosen as $\phi^j(0) = \varepsilon^j$ for j th iteration and $0 < \varepsilon^j e^{-kT} \leq \phi(t) \leq \varepsilon^j, \forall t \in [0, T], j \geq 1$. According to the definition of (3), $s_\phi^j(t)$ will satisfy $s_\phi^j(0) = 0$ for all $j \geq 1$. The critic for reinforcement signal $R^j(t)$ in term of the performance measurement can now be defined as follows :

$$R^j(t) = \text{sgn} \left(s_\phi^j(t) \right) = \begin{cases} 1 & \text{if } s_\phi^j(t) > 0 \\ 0 & \text{if } s_\phi^j(t) = 0 \\ -1 & \text{if } s_\phi^j(t) < 0 \end{cases} \quad (5)$$

3.2. Design of Iterative Learning Controller

We now propose a reinforcement AILC for the nonlinear systems (1) satisfying assumptions (A1)–(A4). The control input at j th iteration is designed as follows :

$$u^j(t) = u_m^j(t) - (\psi^j(t) + 1) R^j(t) |u_m^j(t)| \quad (6)$$

$$u_m^j(t) = -k s^j(t) - \sum_{i=1}^{n-1} c_i e_{i+1}^j(t) + x_d^{(n)}(t) + u_L^j(t) \quad (7)$$

where $\psi^j(t) > 0$ is a control parameter to be updated and $u_L^j(t)$ is a control force designed to compensate for the unknown nonlinearity $f(X^j(t))$ by using a fuzzy neural network (FNN) as shown in Figure 1 (for detailed, please see [16]).

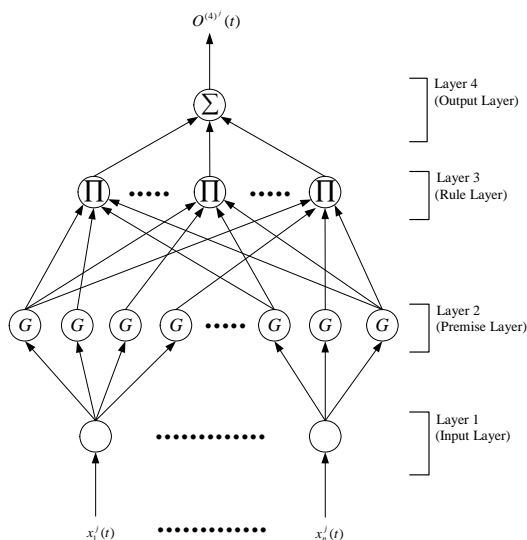


Figure 1 : Structure of the FNN

For this FNN, let $O^{(4)}(X^j(t), W^j(t))$, $O^{(3)}(X^j(t))$ and $W^j(t)$ denote the network output (output of layer 4), firing strength of layer 3 (output of layer 3) and network weight between layer 3 and layer 4, respectively at time t of the j th iteration. In this paper, the FNN will take the form of

$$O^{(4)}(X^j(t), W^j(t)) = W^j(t)^\top O^{(3)}(X^j(t)) \quad (8)$$

where $W^j(t) \in R^{M \times 1}$ with M being the numbers of rule nodes, $O^{(3)}(X^j(t)) = [O_1^{(3)}(X^j(t)), \dots, O_M^{(3)}(X^j(t))]^\top$ with elements $O_\ell^{(3)}(X^j(t))$, $\ell = 1, \dots, M$ being determined by the selected membership functions. Note that $0 < O_\ell^{(3)}(X^j(t)) \leq 1$. It is well known that the FNN (8) can uniformly approximate a real continuous nonlinear function vector X^j on a compact set $\mathcal{A}_c \subset \mathcal{R}^{n \times 1}$. An important aspect of the above approximation property is that there exist optimal weights W^* such that the function approximation error between the optimal $O^{(4)}(X^j(t), W^*)$ and $f(X^j(t))$ can be bounded by a prescribed constant θ^* on the compact set \mathcal{A}_c . More precisely, if we let $f(X^j(t)) = O^{(4)}(X^j(t), W^*) + \epsilon(X^j(t))$, then the approximation errors will satisfy $\|\epsilon(X^j(t))\| \leq \theta^*$, $\forall X^j \in \mathcal{A}_c$.

$u_L^j(t)$ is now designed by using the FNN in (8) as follows:

$$u_L^j(t) = W^j(t)^\top O^{(3)}(X^j(t)) - R^j(t) \theta^j(t) \quad (9)$$

where $W^j(t) \in R^{M \times 1}$ and $\theta^j(t) \in R$ are the fuzzy-neural network parameters and additional robust control parameter. The parameters $W^j(t)$ and $\theta^j(t)$, together with $\psi^j(t)$ in (6), will be tuned by suitable adaptive laws in order to guarantee the signal boundedness in time domain and error convergence in iteration domain. The adaptive laws combining time domain and iteration domain adaptation without deadzone or bounds of unknown parameters are proposed as follows:

$$(1 - \gamma_1) \dot{W}^j(t) = -\gamma_1 W^j(t) + \gamma_1 W^{j-1}(t) - \beta_1 R^j(t) O^{(3)}(X^j(t)) \quad (10)$$

$$(1 - \gamma_2) \dot{\theta}^j(t) = -\gamma_2 \theta^j(t) + \gamma_2 \theta^{j-1}(t) + \beta_2 |R^j(t)| \quad (11)$$

$$(1 - \gamma_3) \dot{\psi}^j(t) = -\gamma_3 \psi^j(t) + \gamma_3 \psi^{j-1}(t) + \beta_3 |R^j(t)| |u_m^j(t)| \quad (12)$$

with $W^j(0) = W^{j-1}(T)$, $\theta^j(0) = \theta^{j-1}(T)$, $\psi^j(0) = \psi^{j-1}(T)$ for $j \geq 1$, and $0 < \gamma_1, \gamma_2, \gamma_3 < 1$, $\beta_1, \beta_2, \beta_3 > 0$. In these adaptive laws, $\gamma_1, \gamma_2, \gamma_3$ and $\beta_1, \beta_2, \beta_3$ are defined as the weighting gains and adaptation gains, respectively. For the first iteration, we set $W^0(t) = W^0$ and $\theta^0(t) = \theta^0$ to be any constant (vector) and $\psi^0(t) = \psi^0 > 0$ to be a small positive number so that $\psi^j(t) > 0, \forall t \in [0, T]$ and $\forall j \geq 1$. (10), (11) and (12) will become pure time-domain adaptive laws if $\gamma_1 = \gamma_2 = 0 = \gamma_3 = 0$, or pure iteration-domain adaptive laws if $\gamma_1 = \gamma_2 = \gamma_3 = 1$.

4. Derivation of Error Equation

An error equation is required to analyze the stability and convergence of the proposed iterative learning control system. To begin with, the derivative of $|s_\phi^j(t)|$ with respect to time t is firstly computed as

$$\begin{aligned} \frac{d}{dt} |s_\phi^j(t)| &= \text{sgn} \left(s_\phi^j(t) \right) \dot{s}_\phi^j(t) \\ &= \begin{cases} \text{sgn} \left(s_\phi^j(t) \right) \left(\dot{s}^j(t) - \dot{\phi}^j(t) \right) & \text{if } s^j(t) > \phi^j(t) \\ 0 & \text{if } |s^j(t)| \leq \phi^j(t) \\ \text{sgn} \left(s_\phi^j(t) \right) \left(\dot{s}^j(t) + \dot{\phi}^j(t) \right) & \text{if } s^j(t) < -\phi^j(t) \end{cases} \\ &= \text{sgn} \left(s_\phi^j(t) \right) \left(\dot{s}^j(t) - \text{sgn} \left(s_\phi^j(t) \right) \dot{\phi}^j(t) \right) \\ &= \text{sgn} \left(s_\phi^j(t) \right) \left\{ \sum_{i=1}^{n-1} c_i e_{i+1}^j(t) - x_d^{(n)}(t) - f(X^j(t)) \right. \\ &\quad \left. + b(X^j(t)) u^j(t) - \text{sgn} \left(s_\phi^j(t) \right) \dot{\phi}^j(t) \right\} \end{aligned}$$

$$\begin{aligned}
&= \operatorname{sgn} \left(s_{\phi}^j(t) \right) \left\{ -k s^j(t) - \operatorname{sgn} \left(s_{\phi}^j(t) \right) \dot{\phi}^j(t) \right\} \\
&\quad + \operatorname{sgn} \left(s_{\phi}^j(t) \right) \left\{ \sum_{i=1}^{n-1} c_i e_{i+1}^j(t) - x_d^{(n)}(t) + k s^j(t) \right. \\
&\quad \left. - f(X^j(t)) + b(X^j(t)) u^j(t) \right\} \quad (13)
\end{aligned}$$

Note that $\operatorname{sgn} \left(s_{\phi}^j(t) \right)$ in (5) can be rewritten as

$$\operatorname{sgn} \left(s_{\phi}^j(t) \right) = \begin{cases} 1 & \text{if } s^j(t) > \phi^j(t) \\ 0 & \text{if } |s^j(t)| \leq \phi^j(t) \\ -1 & \text{if } s^j(t) < -\phi^j(t) \end{cases} \quad (14)$$

and hence it can be easily shown that

$$\operatorname{sgn} \left(s_{\phi}^j(t) \right) \operatorname{sat} \left(\frac{s^j(t)}{\phi^j(t)} \right) = \left(\operatorname{sgn} \left(s_{\phi}^j(t) \right) \right)^2 \quad (15)$$

using the definition of $\operatorname{sat} \left(\frac{s^j(t)}{\phi^j(t)} \right)$ in (4) and $\operatorname{sgn} \left(s_{\phi}^j(t) \right)$ in (14). The first term in the right hand side of equation (13) can now be rewritten by using the fact of (15) as

$$\begin{aligned}
&\operatorname{sgn} \left(s_{\phi}^j(t) \right) \left\{ -k s^j(t) - \operatorname{sgn} \left(s_{\phi}^j(t) \right) \dot{\phi}^j(t) \right\} \\
&= \operatorname{sgn} \left(s_{\phi}^j(t) \right) \left\{ -k s_{\phi}^j(t) - k \phi^j(t) \operatorname{sat} \left(\frac{s^j(t)}{\phi^j(t)} \right) \right. \\
&\quad \left. - \operatorname{sgn} \left(s_{\phi}^j(t) \right) \dot{\phi}^j(t) \right\} \\
&= -k |s_{\phi}^j(t)| - \left(\operatorname{sgn} \left(s_{\phi}^j(t) \right) \right)^2 \left(\dot{\phi}^j(t) + k \phi^j(t) \right) \\
&= -k |s_{\phi}^j(t)| \quad (16)
\end{aligned}$$

Substituting (16) into (13), it yields

$$\begin{aligned}
&\frac{d}{dt} |s_{\phi}^j(t)| \\
&= -k |s_{\phi}^j(t)| + \operatorname{sgn} \left(s_{\phi}^j(t) \right) \left\{ \sum_{i=1}^{n-1} c_i e_{i+1}^j(t) - x_d^{(n)}(t) \right. \\
&\quad \left. + k s^j(t) - f(X^j(t)) + b(X^j(t)) u^j(t) \right\} \quad (17)
\end{aligned}$$

If we apply the reinforcement signal $R^j(t)$ in (5) and the proposed controller $u^j(t)$ in (6) and (7) to (17), we have

$$\begin{aligned}
&\frac{d}{dt} |s_{\phi}^j(t)| \\
&= -k |s_{\phi}^j(t)| + R^j(t) \left\{ \sum_{i=1}^{n-1} c_i e_{i+1}^j(t) - x_d^{(n)}(t) + k s^j(t) \right. \\
&\quad \left. - f(X^j(t)) + u^j(t) + (b(X^j(t)) - 1) u^j(t) \right\} \\
&= -k |s_{\phi}^j(t)| + R^j(t) \left\{ -(\psi^j(t) + 1) R^j(t) |u_m^j(t)| \right. \\
&\quad \left. + (b(X^j(t)) - 1) u^j(t) + u_L^j(t) - f(X^j(t)) \right\} \quad (18)
\end{aligned}$$

Investigating the second and third terms in the right hand side of (18) by using assumption **(A2)**, we have

$$\begin{aligned}
&R^j(t) \left\{ -(\psi^j(t) + 1) R^j(t) |u_m^j(t)| + (b(X^j(t)) \right. \\
&\quad \left. - 1) \times u^j(t) \right\} \\
&= R^j(t) \left\{ -b(X^j(t)) (\psi^j(t) + 1) R^j(t) |u_m^j(t)| \right. \\
&\quad \left. + b(X^j(t)) u_m^j(t) - u_m^j(t) \right\} \\
&\leq -b(X^j(t)) (\psi^j(t) + 1) (R^j(t))^2 |u_m^j(t)| \\
&\quad + b(X^j(t)) |R^j(t)| |u_m^j(t)| + |R^j(t)| |u_m^j(t)| \\
&= -b(X^j(t)) \psi^j(t) |R^j(t)| |u_m^j(t)| + |R^j(t)| |u_m^j(t)| \\
&\leq -b_L \psi^j(t) |R^j(t)| |u_m^j(t)| + |R^j(t)| |u_m^j(t)| \\
&= -b_L \left(\psi^j(t) - \frac{1}{b_L} \right) |R^j(t)| |u_m^j(t)| \quad (19)
\end{aligned}$$

where we use the fact that $(R^j(t))^2 = |R^j(t)|$. The result of (19) implies that (18) can be simplified as

$$\begin{aligned}
&\frac{d}{dt} |s_{\phi}^j(t)| \leq -k |s_{\phi}^j(t)| + R^j(t) \left(u_L^j(t) - f(X^j(t)) \right) \\
&\quad - b_L |u_m^j(t)| |R^j(t)| \left(\psi^j(t) - \frac{1}{b_L} \right) \quad (20)
\end{aligned}$$

It is clear now that $u_L^j(t)$ and $\psi^j(t)$ are designed to compensate for the unknown nonlinear function $f(X^j(t))$ and unknown constant $\frac{1}{b_L}$ respectively. If we define the parameter errors as $\widetilde{W}^j(t) = W^j(t) - W^*$, $\widetilde{\theta}^j(t) = \theta^j(t) - \theta^*$, $\widetilde{\psi}^j(t) = \psi^j(t) - \frac{1}{b_L}$ and substitute (9) into (20), we can get the following error equation for analyzing the stability and convergence of the iterative learning control system,

$$\begin{aligned}
&\frac{d}{dt} |s_{\phi}^j(t)| \\
&\leq -k |s_{\phi}^j(t)| + R^j(t) \left\{ \widetilde{W}^j(t)^{\top} O^{(3)}(X^j(t)) - R^j(t) \theta^j(t) \right. \\
&\quad \left. - \epsilon^j(t) \right\} - b_L |u_m^j(t)| |R^j(t)| \widetilde{\psi}^j(t) \\
&\leq -k |s_{\phi}^j(t)| + R^j(t) \widetilde{W}^j(t)^{\top} O^{(3)}(X^j(t)) - |R^j(t)| \widetilde{\theta}^j(t) \\
&\quad - b_L |u_m^j(t)| |R^j(t)| \widetilde{\psi}^j(t) \quad (21)
\end{aligned}$$

5. Analysis of Stability and Convergence

In order to prove stability and convergence of the proposed reinforcement AILC system, we first give the following two facts.

Fact 1 : The proposed reinforcement AILC system guarantees that all the control parameters and internal signals are bounded for the first iteration.

Choose a Lyapunov-like positive function as

$$V_a^j(t) = |s_\phi^j(t)| + \frac{(1-\gamma_1)}{2\beta_1} \widetilde{W}^j(t)^\top \widetilde{W}^j(t) + \frac{(1-\gamma_2)}{2\beta_2} (\tilde{\theta}^j(t))^2 + \frac{(1-\gamma_3)b_L}{2\beta_3} (\tilde{\psi}^j(t))^2 \quad (22)$$

and compute its derivative with respect to time t along (21), (10), (11) and (12) as follows,

$$\begin{aligned} \dot{V}_a^j(t) &= \frac{d}{dt} |s_\phi^j(t)| + \frac{(1-\gamma_1)}{\beta_1} \widetilde{W}^j(t)^\top \dot{\widetilde{W}}^j(t) \\ &\quad + \frac{(1-\gamma_2)}{\beta_2} \tilde{\theta}^j(t) \dot{\tilde{\theta}}^j(t) + \frac{(1-\gamma_3)b_L}{\beta_3} \tilde{\psi}^j(t) \dot{\tilde{\psi}}^j(t) \\ &\leq -k |s_\phi^j(t)| + R^j(t) \widetilde{W}^j(t)^\top O^{(3)}(X^j(t)) \\ &\quad - |R^j(t)| \tilde{\theta}^j(t) - b_L |u_m^j(t)| |R^j(t)| \tilde{\psi}^j(t) \\ &\quad + \frac{1}{\beta_1} \widetilde{W}^j(t)^\top (1-\gamma_1) \dot{\widetilde{W}}^j(t) \\ &\quad + \frac{1}{\beta_2} \tilde{\theta}^j(t) (1-\gamma_2) \dot{\tilde{\theta}}^j(t) \\ &\quad + \frac{b_L}{\beta_3} \tilde{\psi}^j(t) [(1-\gamma_3) \dot{\tilde{\psi}}^j(t)] \end{aligned} \quad (23)$$

Since $-\gamma_1 W^j(t) + \gamma_1 W^{j-1}(t) = -\gamma_1 \widetilde{W}^j(t) + \gamma_1 \widetilde{W}^{j-1}(t)$, $-\gamma_2 \theta^j(t) + \gamma_2 \theta^{j-1}(t) = -\gamma_2 \tilde{\theta}^j(t) + \gamma_2 \tilde{\theta}^{j-1}(t)$ and $-\gamma_3 \psi^j(t) + \gamma_3 \psi^{j-1}(t) = -\gamma_3 \tilde{\psi}^j(t) + \gamma_3 \tilde{\psi}^{j-1}(t)$, $\dot{V}_a^j(t)$ in (23) can be simplified by using (10)–(12) as

$$\begin{aligned} \dot{V}_a^j(t) &\leq -k |s_\phi^j(t)| \\ &\quad - \frac{\gamma_1}{\beta_1} \widetilde{W}^j(t)^\top \widetilde{W}^j(t) + \frac{\gamma_1}{\beta_1} \widetilde{W}^j(t)^\top \widetilde{W}^{j-1}(t) \\ &\quad - \frac{\gamma_2}{\beta_2} (\tilde{\theta}^j(t))^2 + \frac{\gamma_2}{\beta_2} \tilde{\theta}^j(t) \tilde{\theta}^{j-1}(t) \\ &\quad - \frac{\gamma_2 b_L}{\beta_2} (\tilde{\psi}^j(t))^2 + \frac{\gamma_2 b_L}{\beta_2} \tilde{\psi}^j(t) \tilde{\psi}^{j-1}(t) \\ &= -k |s_\phi^j(t)| - \frac{\gamma_1}{2\beta_1} \widetilde{W}^j(t)^\top \widetilde{W}^j(t) - \frac{\gamma_2}{2\beta_2} (\tilde{\theta}^j(t))^2 \\ &\quad - \frac{\gamma_3 b_L}{2\beta_3} (\tilde{\psi}^j(t))^2 - \frac{\gamma_1}{2\beta_1} (\widetilde{W}^j(t) - \widetilde{W}^{j-1}(t))^\top \\ &\quad \times (\widetilde{W}^j(t) - \widetilde{W}^{j-1}(t)) - \frac{\gamma_2}{2\beta_2} (\tilde{\theta}^j(t) - \tilde{\theta}^{j-1}(t))^2 \\ &\quad - \frac{\gamma_3 b_L}{2\beta_3} (\tilde{\psi}^j(t) - \tilde{\psi}^{j-1}(t))^2 \\ &\quad + \frac{\gamma_1}{2\beta_1} \widetilde{W}^{j-1}(t)^\top \widetilde{W}^{j-1}(t) + \frac{\gamma_2}{2\beta_2} (\tilde{\theta}^{j-1}(t))^2 \\ &\quad + \frac{\gamma_3 b_L}{2\beta_3} (\tilde{\psi}^{j-1}(t))^2 \\ &\leq V_b^{j-1}(t) \end{aligned} \quad (24)$$

where $V_b^{j-1}(t) = \frac{\gamma_1}{2\beta_1} \widetilde{W}^{j-1}(t)^\top \widetilde{W}^{j-1}(t) + \frac{\gamma_2}{2\beta_2} (\tilde{\theta}^{j-1}(t))^2 + \frac{\gamma_3 b_L}{2\beta_3} (\tilde{\psi}^{j-1}(t))^2$. Since $\widetilde{W}^0(t) = W^0(t) - W^* =$

$W^0 - W^* \equiv \widetilde{W}^0, \tilde{\theta}^0(t) = \theta^0(t) - \theta^* = \theta^0 - \theta^* \equiv \tilde{\theta}^0$ and $\tilde{\psi}^0(t) = \psi^0(t) - \frac{1}{b_L} = \psi^0 - \frac{1}{b_L} \equiv \tilde{\psi}^0$ are bounded for all $t \in [0, T]$ so that if $j = 1$, (24) can be rewritten as

$$\dot{V}_a^1(t) \leq V_b^0(t) = \frac{\gamma_1}{2\beta_1} \widetilde{W}^{0\top} \widetilde{W}^0 + \frac{\gamma_2}{2\beta_2} (\tilde{\theta}^0)^2 + \frac{\gamma_3 b_L}{2\beta_3} (\tilde{\psi}^0)^2 \quad (25)$$

Note that the initial value $V_a^1(0)$ is bounded since $s_\phi^1(0) = 0, \widetilde{W}^1(0) = W^1(0) - W^* = W^0(T) - W^* = \widetilde{W}^0, \tilde{\theta}^1(0) = \theta^1(0) - \theta^* = \theta^0(T) - \theta^* = \tilde{\theta}^0$ and $\tilde{\psi}^1(0) = \psi^1(0) - \frac{1}{b_L} = \psi^0(T) - \frac{1}{b_L} = \tilde{\psi}^0$. Together with the result of (25), it implies $V_a^1(t), s_\phi^1(t), \widetilde{W}^1(t), \tilde{\theta}^1(t), \tilde{\psi}^1(t) \in L_{\infty e}[0, T]$ and hence, $s^1(t)$ (by (3)), $\dot{e}_1^1(t) \dots, \dot{e}_n^1(t)$ (by (2)), $u^1(t)$ (by (6) and (7)), $\dot{W}^1(t)$ (by (10)), $\dot{\theta}^1(t)$ (by (11)), $\dot{\psi}^1(t)$ (by (12)) $\in L_{\infty e}[0, T]$.

Fact 2 : The proposed reinforcement AILC system guarantees that $\widetilde{W}^j(T), \tilde{\theta}^j(T), \tilde{\psi}^j(T)$ are bounded for all $j \geq 1$.

Define a positive function $V^j(T)$ as

$$\begin{aligned} V^j(T) &= \int_0^T \left[\frac{\gamma_1}{2\beta_1} \widetilde{W}^j(t)^\top \widetilde{W}^j(t) + \frac{\gamma_2}{2\beta_2} (\tilde{\theta}^j(t))^2 \right. \\ &\quad \left. + \frac{\gamma_3 b_L}{2\beta_3} (\tilde{\psi}^j(t))^2 \right] dt + \frac{1-\gamma_1}{2\beta_1} \widetilde{W}^j(T)^\top \widetilde{W}^j(T) \\ &\quad + \frac{1-\gamma_2}{2\beta_2} (\tilde{\theta}^j(T))^2 + \frac{(1-\gamma_3)b_L}{2\beta_3} (\tilde{\psi}^j(T))^2 \end{aligned} \quad (26)$$

Using the technique of integration by parts, we have

$$\begin{aligned} &\frac{(1-\gamma_1)}{2\beta_1} \widetilde{W}^j(T)^\top \widetilde{W}^j(T) \\ &= \frac{(1-\gamma_1)}{\beta_1} \int_0^T \widetilde{W}^j(t)^\top \dot{\widetilde{W}}^j(t) dt + \frac{(1-\gamma_1)}{2\beta_1} \\ &\quad \times \widetilde{W}^j(0)^\top \widetilde{W}^j(0) \\ &= \frac{(1-\gamma_2)}{2\beta_2} (\tilde{\theta}^j(T))^2 \\ &= \frac{(1-\gamma_2)}{\beta_2} \int_0^T \tilde{\theta}^j(t) \dot{\tilde{\theta}}^j(t) dt + \frac{(1-\gamma_2)}{2\beta_2} (\tilde{\theta}^j(0))^2 \\ &= \frac{(1-\gamma_3)b_L}{2\beta_3} (\tilde{\psi}^j(T))^2 \\ &= \frac{(1-\gamma_3)b_L}{\beta_3} \int_0^T \tilde{\psi}^j(t) \dot{\tilde{\psi}}^j(t) dt + \frac{(1-\gamma_3)b_L}{2\beta_3} (\tilde{\psi}^j(0))^2 \end{aligned}$$

The difference between $V^j(T)$ and $V^{j-1}(T)$ can be derived by the facts of $\widetilde{W}^j(0) = \widetilde{W}^{j-1}(T), \tilde{\theta}^j(0) = \tilde{\theta}^{j-1}(T)$ and $\tilde{\psi}^j(0) = \tilde{\psi}^{j-1}(T)$ as follows :

$$\begin{aligned} V^j(T) - V^{j-1}(T) &= \int_0^T \left[\frac{\gamma_1}{2\beta_1} (\widetilde{W}^j(t)^\top \widetilde{W}^j(t) - \widetilde{W}^{j-1}(t)^\top \widetilde{W}^{j-1}(t)) \right. \\ &\quad \left. + \frac{\gamma_2}{2\beta_2} \left((\tilde{\theta}^j(t))^2 - (\tilde{\theta}^{j-1}(t))^2 \right) \right] \end{aligned}$$

$$\begin{aligned}
 & + \frac{\gamma_3 b_L}{2\beta_3} \left[\left(\tilde{\psi}^j(t) \right)^2 - \left(\tilde{\psi}^{j-1}(t) \right)^2 \right] dt \\
 & + \frac{(1-\gamma_1)}{\beta_1} \int_0^T \tilde{W}^j(t)^\top \dot{\tilde{W}}^j(t) dt \\
 & + \frac{(1-\gamma_1)}{2\beta_1} \tilde{W}^j(0)^\top \tilde{W}^j(0) \\
 & - \frac{(1-\gamma_1)}{2\beta_1} \tilde{W}^{j-1}(T)^\top \tilde{W}^{j-1}(T) \\
 & + \frac{(1-\gamma_2)}{\beta_2} \int_0^T \tilde{\theta}^j(t) \dot{\tilde{\theta}}^j(t) dt \\
 & + \frac{(1-\gamma_2)}{2\beta_2} \left(\tilde{\theta}^j(0) \right)^2 - \frac{(1-\gamma_2)}{2\beta_2} \left(\tilde{\theta}^{j-1}(T) \right)^2 \\
 & + \frac{(1-\gamma_3) b_L}{\beta_3} \int_0^T \tilde{\psi}^j(t) \dot{\tilde{\psi}}^j(t) dt \\
 & + \frac{(1-\gamma_3) b_L}{2\beta_3} \left(\tilde{\psi}^j(0) \right)^2 - \frac{(1-\gamma_3) b_L}{2\beta_3} \left(\tilde{\psi}^{j-1}(T) \right)^2 \\
 = & \int_0^T \left[-\frac{\gamma_1}{2\beta_1} \left(\tilde{W}^j(t) - \tilde{W}^{j-1}(t) \right)^\top \right. \\
 & \times \left(\tilde{W}^j(t) - \tilde{W}^{j-1}(t) \right) \\
 & - \frac{\gamma_2}{2\beta_2} \left(\tilde{\theta}^j(t) - \tilde{\theta}^{j-1}(t) \right)^2 \\
 & \left. - \frac{\gamma_3 b_L}{2\beta_3} \left(\tilde{\psi}^j(t) - \tilde{\psi}^{j-1}(t) \right)^2 \right] dt \\
 & + \int_0^T \left[-R^j(t) \tilde{W}^j(t)^\top O^{(3)}(X^j(t)) \right. \\
 & \left. + |R^j(t)| \tilde{\theta}^j(t) + b_L |u_m^j(t)| |R^j(t)| \tilde{\psi}^j(t) \right] dt \\
 \leq & \int_0^T \left[-R^j(t) \tilde{W}^j(t)^\top O^{(3)}(X^j(t)) + |R^j(t)| \tilde{\theta}^j(t) \right. \\
 & \left. + b_L |u_m^j(t)| |R^j(t)| \tilde{\psi}^j(t) \right] dt \tag{27}
 \end{aligned}$$

Integrating both side of (21) from 0 to T gives

$$\begin{aligned}
 & \int_0^T \left[-R^j(t) \tilde{W}^j(t)^\top O^{(3)}(X^j(t)) + |R^j(t)| \tilde{\theta}^j(t) \right. \\
 & \left. + b_L |u_m^j(t)| |R^j(t)| \tilde{\psi}^j(t) \right] dt \\
 \leq & \int_0^T -k |s_\phi^j(t)| dt - |s_\phi^j(T)| \tag{28}
 \end{aligned}$$

where we use the property of $|s_\phi^j(0)| = 0$. Substituting (28) into (27), it yields

$$V^j(T) - V^{j-1}(T) \leq \int_0^T -k |s_\phi^j(t)| dt - |s_\phi^j(T)| \tag{29}$$

Since $V^1(T)$ is bounded by Fact 1, and $V^j(T)$ is positive and monotonically decreasing, $V^j(T)$ is bounded for all

$j \geq 1$ and will converge as j approaches infinity to some limit value $V(T)$ (independent of j). The boundedness of $V^j(T)$ also ensures the boundedness of $\tilde{W}^j(T)$, $\tilde{\theta}^j(T)$, $\tilde{\psi}^j(T)$ for all $j \geq 1$.

The boundedness of $\tilde{W}^j(T)$, $\tilde{\theta}^j(T)$ and $\tilde{\psi}^j(T)$ (or equivalently the boundedness of $\tilde{W}^j(0)$, $\tilde{\theta}^j(0)$ and $\tilde{\psi}^j(0)$) for all $j \geq 1$ is shown in Fact 2, the convergence of $s_\phi^j(t)$, $s^j(t)$, $e_1^j(t), \dots, e_n^j(t)$ and boundedness of all internal signals for all $j \geq 1$ are now established in the following Lemma.

Lemma 1 : The proposed reinforcement AILC system ensures that all adjustable control parameters and internal signals $s_\phi^j(t)$, $s^j(t)$, $e_1^j(t), \dots, e_n^j(t)$, $W^j(t)$, $\theta^j(t)$, $\psi^j(t)$, $u^j(t)$, $\dot{W}^j(t)$, $\dot{\theta}^j(t)$, $\dot{\psi}^j(t) \in L_\infty[0, T]$ for all $j \geq 1$.

Proof: Integrating (24) from 0 to t , we have

$$\begin{aligned}
 V_a^j(t) & \leq V_a^j(0) + \int_0^t V_b^{j-1}(t') dt' \\
 & \leq V_a^j(0) + \int_0^T V_b^{j-1}(t) dt \tag{30}
 \end{aligned}$$

Since $V^j(T)$, defined in (26), is bounded $\forall j \geq 1$ according to Fact 2, we conclude that $\int_0^T V_b^{j-1}(t) dt$ is bounded $\forall j \geq 1$. Furthermore, the initial value $V_a^j(0)$ of $V_a^j(t)$ is also bounded for all $j \geq 1$ due to Fact 2. This implies from (30) that $V_a^j(t)$ and hence, $s_\phi^j(t)$, $\tilde{W}^j(t)$, $\tilde{\theta}^j(t)$, $\tilde{\psi}^j(t)$, $\in L_\infty[0, T]$. Using the same argument as in Fact 1, it can be easily shown that $s_\phi^j(t)$, $s^j(t)$, $e_1^j(t), \dots, e_n^j(t)$, $W^j(t)$, $\theta^j(t)$, $\psi^j(t)$, $u^j(t)$, $\dot{W}^j(t)$, $\dot{\theta}^j(t)$, $\dot{\psi}^j(t) \in L_\infty[0, T]$ for all $j \geq 1$.

Theorem 1 : Define $E^j(t) = [e_1^j(t), e_2^j(t), \dots, e_{n-1}^j(t)]^\top$. The proposed reinforcement AILC system guarantees that the tracking performance and system stability will satisfy the following results,

- (t1) $\lim_{j \rightarrow \infty} |s_\phi^j(t)| = |s_\phi^\infty(t)| = 0$ for all $t \in [0, T]$.
- (t2) $\lim_{j \rightarrow \infty} |s^j(t)| = |s^\infty(t)| \leq \phi^\infty(t) = e^{-kt} \varepsilon^\infty$, for all $t \in [0, T]$.
- (t3) Let λ be the positive constant such that $\Delta(D - \lambda)$ is still a Hurwitz polynomial, then

$$\begin{aligned}
 \lim_{j \rightarrow \infty} \|E^j(t)\| & \leq m_1 e^{-\lambda t} \|E^\infty(0)\| \\
 & \quad + m_1 \varepsilon^\infty \frac{e^{-kt} - e^{-\lambda t}}{\lambda - k} \tag{31}
 \end{aligned}$$

$$\lim_{j \rightarrow \infty} |e_n^j(t)| \leq \sum_{i=1}^{n-1} c_i |e_i^\infty(t)| + e^{-kt} \varepsilon^\infty \tag{32}$$

for some positive constant m_1 and for all $t \in [0, T]$.

Proof :

- (t1) According to Lemma 1, we have $|s_\phi^j(t)| \in L_\infty[0, T]$ and $\frac{d}{dt} |s_\phi^j(t)| = \text{sgn} \left(s_\phi^j(t) \right) \dot{s}_\phi^j(t) \in L_\infty[0, T]$ for

all $j \geq 1$. This implies that $|s_\phi^j(t)|$ is uniformly continuous over $[0, T]$ for all $j \geq 1$. On the other hand, we have

$$\int_0^T k|s_\phi^j(t)|dt \leq V^{j-1}(T) - V^j(T) \leq V^1(T) \quad (33)$$

for all $j \geq 1$ by using (29). (33) implies that

$$\lim_{j \rightarrow \infty} \int_0^T k|s_\phi^j(t)|dt = 0 \quad (34)$$

Together with the result that $|s_\phi^j(t)| \in L_{\infty e}[0, T]$ for all $j \geq 1$ according to Lemma 1, we can conclude, by using Barbalat's lemma (e.g., Lemma 3.2.6 in [19]), that $\lim_{j \rightarrow \infty} |s_\phi^j(t)| = 0$ for all $t \in [0, T]$.

(t2) Since $\lim_{j \rightarrow \infty} s_\phi^j(t) = 0$, we have the bound of $s^\infty(t)$ by (3) as

$$\lim_{j \rightarrow \infty} |s^j(t)| \leq \phi^\infty(t) = e^{-kt} \varepsilon^\infty, \forall t \in [0, T]$$

This prove (t2) of Theorem 1.

(t3) In order to investigate the tracking performance in the final iteration when (t1) and (t2) of theorem 1 are achieved, we consider the following state space equation :

$$\dot{E}^\infty(t) = A_c E^\infty(t) + B_c s^\infty(t) \quad (35)$$

where

$$A_c = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -c_1 & -c_2 & -c_3 & \cdots & -c_{n-1} \end{bmatrix}$$

$$B_c = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

by using assumption (A4) and the definition of control function $s^j(t)$ in (2). Solution of (35) in time domain is given by

$$E^\infty(t) = \Phi(t)E^\infty(0) + \int_0^t \Phi(t-\tau)B_c s^\infty(\tau)d\tau \quad (36)$$

where the state transition matrix $\Phi(t)$ satisfies $\|\Phi(t)\| \leq m_1 e^{-\lambda t}$ for some suitable positive constant m_1 . Taking norms on (36), it yields

$$\|E^\infty(t)\| \leq m_1 e^{-\lambda t} \|E^\infty(0)\| + m_1 \int_0^t e^{-\lambda(t-\tau)} e^{-k\tau} \varepsilon^\infty d\tau$$

$$\leq m_1 e^{-\lambda t} \|E^\infty(0)\| + m_1 \varepsilon^\infty \frac{e^{-kt} - e^{-\lambda t}}{\lambda - k}$$

This concludes (31) of (t3). Finally, tracking performance of $e_n^\infty(t)$ which is shown in (32) can be easily derived by using the definition of (2). Q.E.D.

6. A Simulation Example

In this section, we apply the proposed reinforcement AILC system to a mass-spring-damper system [20] whose state equation is described by

$$\dot{x}_1^j(t) = x_2^j(t)$$

$$\dot{x}_2^j(t) = -0.1(x_2^j(t))^3 - 0.02x_1^j(t) - 0.67(x_1^j(t))^3 + u^j(t)$$

The control objective is to make the state vector $X^j(t) = [x_1^j(t), x_2^j(t)]^T$ to track a desired trajectory $X_d(t) = [x_d(t), \dot{x}_d(t)]^T = [\sin(t), \cos(t)]^T$ for $t \in [0, 15]$. The design steps are summarized as follows :

(D1) Define $s^j(t) = c_1 e_1^j(t) + e_2^j(t)$ where $e_1^j(t) = x_1^j(t) - \sin t$, $e_2^j(t) = x_2^j(t) - \cos t$ and $s_\phi^j(t) = s^j(t) - \phi^j(t) \text{sat} \left(\frac{s^j(t)}{\phi^j(t)} \right)$ with $\dot{\phi}^j(t) + k\phi^j(t) = 0$ and $k > 0$.

(D2) Design the controller as in (6) and (7), $u_L^j(t)$ is constructed in (9). Since the working domain of the desired trajectory $X_d(t) = [\sin(t), \cos(t)]^T$ is within the interval $[-1, 1]$, the centers and variances are chosen as $m^0(t) = m^0 = [m_{11}^0, m_{12}^0, m_{13}^0, m_{14}^0, m_{15}^0, m_{21}^0, m_{22}^0, m_{23}^0, m_{24}^0, m_{25}^0] = [-1.5, -0.75, 0, 0.75, 1.5, -1.5, -0.75, 0, 0.75, 1.5]$ and $\sigma^0(t) = \sigma^0 = \sigma_{i\ell} = 5, i = 1, 2, \ell = 1, \dots, 5$ at the first iteration to cover this interval, respectively. In addition, we set the control parameter $\theta^0(t) = \theta^0 = 0.1$ at the first iteration for all $t \in [0, 15]$. It is noted that the initial values of the consequent parameters can be roughly estimated if the nonlinear functions $f(X^j(t))$ and $b(X^j(t))$ are partially known. However, we often arbitrarily choose the initial parameters.

(D3) Finally, the adaptation algorithms as in (10)-(12) are adopted to update the FNN parameters and control parameters.

In order to show the robustness to the varying initial state errors, we assume the initial states take the arbitrary values for the first 5 iterations : $[x_1^j(0), x_2^j(0)] = [0.9003, 0], [-0.028, 0.1], [-0.0871, -0.1], [-0.0829, 0.12], [0.25, 0]$. The initial value of the boundary layer $\phi^j(t)$ is then chosen according to $\phi^j(0) = \varepsilon^j = |s^j(0)| = |2e_1^j(0) + e_2^j(0)| = |2(x_1^j(0) - \sin(0)) + (x_2^j(0) - \cos(0))|$ at the beginning of each iteration. The AILC (6), (7), (10)-(12) is applied with the design parameters $c_1 = 5, k = 10, \gamma_1 = \gamma_2 = \gamma_3 = 0.5, \beta_1 = \beta_2 = \beta_3 = 10$. To study the effects of learning performances, we first show the reinforcement signal $R^j(t)$ in term of the performance measurement for critic is shown in Figure 2 (a). It is shown that the reinforcement signal provides a satisfaction about the nice tracking performance. Compared with a similar work in [16], we apply the reinforcement AILC and the hybrid AILC to control the same nonlinear plant. The supremum value of $|s_\phi^j(t)|$ via the reinforcement AILC and the hybrid AILC with respective to iteration j are shown in Figure 2 (b). In this simulation, it is clear that the asymptotic convergence by

using the reinforcement AILC can be used to verify (t1) of the main theorem. Since the learning process is almost completed at the 5th iteration, we demonstrate the learning error $s^5(t)$ in Figure 2 (c). The trajectory of $s^5(t)$ satisfies $-s^5(0) \times e^{-kt} \leq s^5(t) \leq s^5(0) \times e^{-kt}$ which clearly proves (t2) of the main theorem. In addition, it is necessary to see the relation between system states $x_1^j(t), x_2^j(t)$ and desired states $x_d(t), \dot{x}_d(t)$. The nice tracking performances of both states at the 5th iteration are provided in Figure 2 (d) and Figure 2 (e), respectively. Finally, the bounded learned control force $u^5(t)$ is plotted in Figure 2 (f).

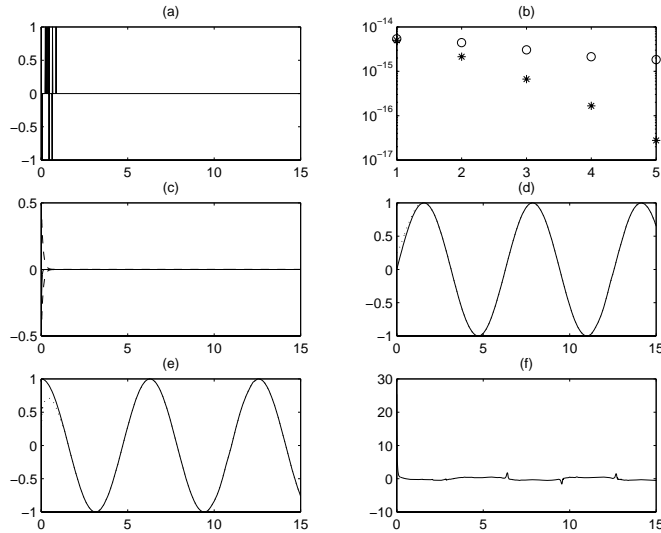


Figure 2 :

- (a) $R^5(t)$ versus time t ;
 (b) $\sup_{t \in [0,15]} |s^j(t)|$ (* * * for Reinforcement AILC and $\circ \circ \circ$ for Hybrid AILC) versus iteration j ;
 (c) $s^5(t)$ (solid line) and $\pm \phi^5(t)$ (dashed lines) versus time t ;
 (d) $x_1^5(t)$ (dotted line) and $x_d(t)$ (solid line) versus time t ;
 (e) $x_2^5(t)$ (dotted line) and $\dot{x}_d(t)$ (solid line) versus time t ;
 (f) $u^5(t)$ versus time t , $k = 10$, $\gamma_1 = \gamma_2 = \gamma_3 = 0.5$, $\beta_1 = \beta_2 = \beta_3 = 10$.

7. Conclusion

This paper presents a method which applies the FNN, reinforcement learning control and AILC to construct a new reinforcement adaptive learning controller. The FNN based reinforcement AILC is design for iterative learning tracking control of nonlinear systems. Compared with some existing reinforcement learning method which needs the gradient information and an approximation of plant Jacobian, the proposed reinforcement AILC doesn't require a prior offline training phase. The discrete reinforcement signal generated from the critic is an evaluation in terms of the performance measurement. Based on the reinforcement signal, an iterative learning component is designed to meet the control objective. A Lyapunov-like analysis is

given to analyze the stability and convergence of the learning system. It is shown that the tracking error asymptotically converge to a tunable residual set as iteration goes to infinity and all adjustable parameters as well as the internal signals will remain bounded.

Acknowledgement

This work is supported by National Science Council, R.O.C., under Grant NSC 99-2221-E-211-011-MY2

References

- [1] S. J. Yoo and J. B. Park, Neural-network-based decentralized adaptive control for a class of large-scale nonlinear systems with unknown time-varying, *IEEE Trans. on Systems, Man, and Cybernetics, Part-B: Cybernetics*, Vol. 39, No. 5, (2009), 1316-1323.
- [2] M. Chen, S. S. Ge and B. How, Robust adaptive neural network control for a class of uncertain MIMO nonlinear systems with input nonlinearities, *IEEE Trans. Neural Networks*, Vol. 21, No. 5, (2010), 796-812.
- [3] S. Tong and Y. Li, Observer-based fuzzy adaptive control for strict-feedback nonlinear systems, *Fuzzy Sets and Systems*, Vol. 160, No. 12, (2009), 1749-1764.
- [4] H. R. Berenji and P. Khedkar, Learning and tuning fuzzy logic controllers through reinforcements, *IEEE Trans. Neural Networks*, Vol. 3, No. 5, (1992), 724-740.
- [5] C. F. Juang and C. H. Hsu, Reinforcement Interval Type-2 Fuzzy Controller Design by Online Rule Generation and Q-Value-Aided Ant Colony Optimization, *IEEE Transactions on Systems Man and Cybernetics, Part B: Cybernetics*, Vol. 39, No. 6, (2009), 1528-1542.
- [6] C. F. Juang and C. H. Hsu, Reinforcement ant optimized fuzzy controller for mobile-robot wall-following control, *IEEE Trans. on Industrial Electronics*, Vol. 56, No. 10, (2009), 3931-3940.
- [7] C. K. Lin, H_∞ reinforcement learning control of robot manipulators using fuzzy wavelet networks, *Fuzzy Sets and Systems*, Vol. 160, Issue 12, (2009), 1765-1786.
- [8] C. Y. Hsu, Y. C. Hsu and S. F. Lin, Reinforcement evolutionary learning using data mining algorithm with TSK-type fuzzy controllers, *Applied Soft Computing*, Vol. 11, Issue 3, (2011), 3247-3259.
- [9] C. J. Lin and C. H. Chen, Nonlinear system control using self-evolving neural fuzzy inference networks with reinforcement evolutionary learning, *Applied Soft Computing*, Vol. 11, Issue 8, (2011), 5463-5476.
- [10] K.L. Moore, J. X. Xu, Special issue on iterative learning control, *Int. J. Contr.*, Vol. 73, (2000), 819-823.
- [11] D.A. Bristow, M. Tharayil, A.G. Allyn, A survey of iterative learning, *IEEE Control Systems Magazine*, Vol. 26, (2006), 96-114.
- [12] H.S. Ahn, Y. Chen, and K.L. Moore, Iterative learning control: Brief survey and categorization, *IEEE Transactions on Systems Man and Cybernetics, Part C - Applications and Reviews*, Vol. 37, No. 6, (2007), 1099-1121.

- [13] A. Tayebi and C.J. Chien, A Unified adaptive iterative learning control framework for uncertain nonlinear systems, *IEEE Transactions on Automatic Control*, Vol. 52, No. 10, (2007), 1907-1913.
- [14] I. Rotariu, M. Steinbuch and R. Ellenbroek, Adaptive iterative learning control for high precision motion systems, *IEEE Transactions on Control Systems Technology*, Vol. 16, No. 5, (2008), 1075 - 1082.
- [15] W. S. Chen, J. Li and J. Li, Practical adaptive iterative learning control framework based on robust adaptive approach, *Asian Journal of Control*, Vol. 13, No. 1, (2011), 85-93.
- [16] Y. C. Wang, C. J. Chien, and D. T. Lee, A hybrid adaptive scheme of fuzzy-neural iterative learning controller for nonlinear dynamic systems, *International Journal of Fuzzy Systems*, Vol. 7, No. 4, (2005), 147-157.
- [17] C.J. Chien, A combined adaptive law for fuzzy iterative learning control of nonlinear systems with varying control tasks, *IEEE Trans. Fuzzy Syst.*, Vol. 16, No. 1, (2008), 40-51.
- [18] Y. C. Wang and C. J. Chien, Decentralized adaptive fuzzy neural iterative learning control for nonaffine nonlinear interconnected systems," *Asian Journal of Control*, Vol. 13, No. 1, (2011), 94-106.
- [19] P.A. Ioannou and J. Sun, *Robust Adaptive Control*, (Englewood Cliffs, NJ, Prentice-Hall, 1996).
- [20] K. Tanaka, T. Ikeda and H. O. Wang, Robust stabilization of a class of uncertain nonlinear systems via fuzzy control: quadratic stabilizability, H^∞ control theory, and linear matrix inequalities, *IEEE Trans. Fuzzy Syst.*, Vol. 4, Issue: 1, (1996), 14-23.



Ying-Chung Wang was born in Taipei, Taiwan, in 1972. He received his B.S. degree in Electronic Engineering and M.S. degree in Mechatronic Engineering from the Huafan University, Taipei, Taiwan, in 1996 and 1998, respectively, and his Ph.D. degree in Electrical and Control Engineering from the National Chiao-Tung University, Hsinchu, Taiwan, in 2003. From 2004 to 2008, he was a Postdoctoral Fellow of the Institute of Information Science, Academia Sinica. In 2009, he joined Huafan University, Taipei, Taiwan, where he is currently an Assistant Professor of Department of Electronic Engineering. His main research interests are fuzzy systems, neural networks, iterative learning control, intelligent adaptive control, and energy management systems.



Chiang-Ju Chien was born in Taiwan, in 1963. He received his B.S. degree in nuclear engineering from National Tsing Hua University, Hsinchu, Taiwan, in 1985 and his Ph.D. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1992, respectively. Since 1993, he has been with the Department of Electronic Engineering, Huafan University, Taipei, Taiwan, where he is currently a Professor. From 2000 to 2005, he served as the Chair of the same Department, and he currently serves as the Secretary General of Secretariat in Huafan University. His current research interests are mainly in iterative learning control, adaptive control, fuzzy-neuro systems, and control circuit design. Dr. Chien is a member of IEEE and the Chinese Automatic Control Society.

