



Adaptive MNN control for a class of non-affine NARMAX systems with disturbances

S.S. Ge*, J. Zhang, T.H. Lee

Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576, Singapore

Received 17 September 2002; received in revised form 28 November 2003; accepted 25 February 2004

Abstract

In this paper, adaptive multi-layer neural network (MNN) control is developed for a class of discrete-time non-affine nonlinear systems in nonlinear auto regressive moving average with eXogenous inputs (NARMAX) model. By using implicit function theorem, the existence of the implicit desired feedback control (IDFC) is proved. MNNs are used as the emulator of the desired feedback control. Projection algorithms are used to guarantee the boundedness of the neural network (NN) weights, which removes the need of persistent exciting (PE) condition for parameter convergence. Simulation results show the effectiveness of the proposed control scheme.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Non-affine nonlinear system; Implicit function theorem; Multi-layer neural networks; Projection algorithm

1. Introduction

Over the past few years, adaptive control for continuous-time nonlinear system has been studied extensively. In general, these methods cannot be directly applied to discrete-time systems due to some technical difficulties, such as the lack of applicability of Lyapunov techniques [9]. Furthermore, due to the fact that discrete-time Lyapunov differences are quadratic in the state first difference, while for continuous-time systems the Lyapunov derivative is linear in the state derivative, discrete-time adaptive control design is far more complex than continuous-time design. Therefore, it is academically

challenging and practically interesting to develop adaptive control scheme for discrete-time nonlinear systems, as plants are controlled at discrete-time instant. Most techniques for estimation and control of unknown nonlinear systems involve the use of a model based on past history of input–output data. These techniques depend largely on the accuracy of these models and the availability of sufficient historical input–output data. It is usually easier to identify discrete-time models and use these data as a basis for designing discrete-time control systems for computer implementation. This observation motivates us to study the control problems of discrete-time models. One of the most popular discrete-time model is nonlinear autoregressive moving average with eXogenous inputs (NARMAX) model, which has been studied in previous literature [1,2,5,6]. In this paper, we investigate a class of non-affine NARMAX systems by using multi-layer neural network (MNN)

* Corresponding author. Tel.: +65-68746821;
fax: +65-67791103.

E-mail address: elegez@nus.edu.sg (S.S. Ge).

approximation, for which, because input is in non-affine form, feedback linearization method cannot be implemented.

In recent years, there has been increasing interest in the application of neural networks (NNs) to process modelling and control. In [3,14], the authors showed how artificial NNs, including radial basis function networks, may be used as universal function approximators, which inspired the use of NNs as emulators to approximate unknown nonlinear functions to construct stable adaptive controller. Frequently used single-layer NNs include high-order neural networks (HONN) [10], radial basis function neural networks (RBF) [13], etc. For MNNs, its universal approximation abilities, parallel distributed processing abilities, learning and adaptation abilities make it one of the most popular tools in function approximation. In [11,12], MNNs were effectively used in nonlinear discrete-time system identification and control. But the MNN-based tracking control in those works was based on the realization that the linearization of a system around an equilibrium point is “well behaved” (controllable, observable, etc.). In this paper, we proposed a MNN control scheme which removes the restriction and is applicable to a wide class of nonlinear non-affine discrete-time systems. Firstly, the existence of the implicit desired feedback control (IDFC) control, which can drive the system output to track the desired trajectory, is proved. Then MNNs are used to emulate the IDFC control. Though single-layer NNs are also applicable to construct the stable adaptive control, MNNs are used in this paper for the following main reasons: (i) tuning of single-layer NNs has been studied extensively; (ii) MNNs make single-layer NN as a special case, if the hidden layers are fixed; and (iii) there is the “curse of dimensionality” problem for the frequently used RBF NNs, and MNNs are very good alternative in reducing the problem.

The main contributions of this paper are that (i) a discrete-time projection algorithm is proposed by extending the continuous-time projection algorithm used in [7,8,16]; (ii) MNNs are used to emulate the desired feedback control of non-affine discrete-time systems, which is not only a challenging topic but also of academic interest, and (iii) semi-global uniformly ultimate boundedness (SGUUB) stability is proposed for a class of non-affine NARMAX systems in the presence of bounded disturbances, for

which, feedback linearization method cannot be implemented.

This paper is organized as follows. In Section 2, the NARMAX model is proposed. Preliminaries about MNNs and projection algorithms are presented in Section 3, respectively. Controller design procedure and stability analysis are presented in Section 4. Finally, numerical simulation is carried out in Section 5.

2. System dynamics

Consider the following τ -step ahead non-affine NARMAX model [1]

$$y(k + \tau) = f(\bar{y}_k, \bar{u}_{k-1}, \bar{d}_{k+\tau-1}, u(k)), \quad (1)$$

where $\bar{y}_k = [y(k), \dots, y(k - n + 1)]^T$, $\bar{u}_{k-1} = [u(k - 1), \dots, u(k - n + 1)]^T$ and $\bar{d}_{k+\tau-1} = [d(k + \tau - 1), \dots, d(k)]^T$. Sequences $\{y(k)\}$, $\{u(k)\}$ and $\{d(k)\}$ represent system outputs, inputs and disturbances, respectively. τ denotes the system delay, or the relative degree of the system.

Assumption 1. The unknown nonlinear function $f(\cdot)$ is continuous and differentiable.

Assumption 2. System output $y(k)$ can be measured and its initial values are assumed to remain in a compact set Ω_{y_0} .

Assumption 3. Assume that there exist two constants $g_1, \epsilon > 0$ such that $0 < \epsilon < |\partial f / \partial u(k)| \leq g_1$, where both ϵ and g_1 are positive constants.

Assumption 3 states that $\partial f / \partial u(k)$ is either positive or negative. From now onwards, without loss of generality, assume that $\partial f / \partial u(k) > 0$.

Assumption 4. The disturbance $d(k)$ is bounded, $|d(k)| \leq d$, where d is a little unknown constant and the partial derivative $|\partial f / \partial d(k)| \leq g_2$, where g_2 is a positive constant.

The control objective is to design control $u(k)$ to drive the system output $y(k)$ follow desired trajectory $y_m(k)$. Define the tracking error as $e(k) = y(k) - y_m(k)$,

by using Mean Value Theorem, we have

$$\begin{aligned}
e(k + \tau) &= f(\bar{y}_k, u(k), \bar{u}_{k-1}, \bar{d}_{k+\tau-1}) - y_m(k + \tau) \\
&= f(\bar{y}_k, u(k), \bar{u}_{k-1}, 0) \\
&\quad + \delta_f^T \bar{d}_{k+\tau-1} - y_m(k + \tau) \\
&= f(\bar{y}_k, u(k), \bar{u}_{k-1}, 0) - y_m(k + \tau) + \delta_{d_k}, \quad (2)
\end{aligned}$$

where

$$\delta_f = \left[\frac{\partial f}{\partial d(k + \tau - 1)} \Big|_{d(k+\tau-1)=d_{\xi_{k+\tau-1}}}, \dots, \frac{\partial f}{\partial d(k)} \Big|_{d(k)=d_{\xi_k}} \right]^T,$$

$$d_{\xi} = [d_{\xi_{k+\tau-1}}, \dots, d_{\xi_k}]^T,$$

$$\delta_{d_k} = \delta_f^T \bar{d}_{k+\tau-1}$$

and $d_{\xi} \in L(0, \bar{d}_{k+\tau-1})$ with L indicate a line starting from 0 and end at $\bar{d}_{k+\tau-1}$.

Remark 1. Noticing the disturbance items in Eq. (1), at time instant k , the sequence $d(k + \tau - 1), \dots, d(k + 1)$ are the future unknown disturbances which cannot be controlled even if they are known. In the following sections, we can see by using the proposed control scheme, the system tracking error can be kept in a bounded compact set even with the presence of these unknown future and current disturbances.

Considering Assumption 4, we know that δ_{d_k} in (2) is bounded by

$$\begin{aligned}
\delta_{d_k} &= \delta_f^T \bar{d}_{k+\tau-1} \\
&= \frac{\partial f}{\partial d(k + \tau - 1)} \Big|_{d(k+\tau-1)=d_{\xi_{k+\tau-1}}} d(k + \tau - 1) \\
&\quad + \dots + \frac{\partial f}{\partial d(k)} \Big|_{d(k)=d_{\xi_k}} d(k) \\
&\leq g_2 d + g_2 d + \dots + g_2 d \\
&= \tau g_2 d. \quad (3)
\end{aligned}$$

Considering Eq. (2), if we can find an IDFC $u^*(k)$, such that

$$f(\bar{y}_k, u^*(k), \bar{u}_{k-1}, 0) - y_m(k + \tau) = 0 \quad (4)$$

then we have $e(k + \tau) = \delta_{d_k}$. Owing to the boundedness of δ_{d_k} , we know that the tracking error is bounded. In ideal condition, if there is no disturbance ($\delta_{d_k} = 0$), we know $e(k + \tau) = 0$, i.e, the exact tracking is obtained in τ steps.

Definition 1. Control input $u^*(k)$ in Eq. (4), which can drive the system output to track the desired trajectory exactly in ideal condition ($\delta_{d_k} = 0$), is called IDFC.

It is obviously that if the practical control input $u(k)$ equals the IDFC control, then the error $e(k + \tau)$ will converge to a small neighborhood of the origin in the presence of boundedness disturbances. Furthermore, if there is no disturbance exist, the tracking error will be zero. Based on implicit function theorem, we have the following lemma to establish the existence of an IDFC $u^*(k)$.

Lemma 1. *If partial derivative $|\partial f / \partial u(k)| > \epsilon > 0$, then there exists a unique and continuous (smooth) function $u^*(k) = \alpha^c(\bar{y}_k, \bar{u}_{k-1}, y_m(k + \tau))$, such that Eq. (4) holds [4].*

3. Preliminaries

3.1. MNNs and function approximation

Because the IDFC input $u^*(k)$ is a continuous function on the compact set Ω_z , we know that there exists an integer l (the number of hidden neurons) and ideal constant weight matrices $W^* \in \mathbb{R}^{l+1}$ and $V^* = [v_1^*, \dots, v_l^*] \in \mathbb{R}^{(n+1) \times l}$, such that

$$u^*(k) = u^*(z) = W^{*T} S(V^{*T} \bar{z}) + \varepsilon_u(z), \forall z \in \Omega_z, \quad (5)$$

where $\bar{z} = [z^T, 1]^T$ and $z \in \mathbb{R}^n$ denotes the input vector, $S(V^{*T} \bar{z}) = [s(v_1^{*T} \bar{z}), \dots, s(v_l^{*T} \bar{z}), 1]^T$ with $s(*)$ denotes the sigmoid function [4]. In this paper, we choose $s(x) = 1/(1 + e^{-x})$ as the activation function. The derivative of $s(x)$ with respect to x is

$$s'(x) = \frac{d[s(x)]}{dx} = \frac{e^{-x}}{(1 + e^{-x})^2}.$$

It is easy to check that

$$0 \leq s'(x) \leq 0.25 \quad \text{and} \quad |xs'(x)| \leq 0.2239$$

for all $x \in \mathbb{R}$. (6)

Define \hat{W} and \hat{V} as the estimates of W^* and V^* , $\tilde{W} = \hat{W} - W^*$ and $\tilde{V} = \hat{V} - V^*$ as the estimation error. By noting $\hat{S} = S(\hat{V}^T \bar{z})$, we have

$$\begin{aligned} \|\hat{S}'\|_F &\leq \sum_{i=1}^l s'(\hat{v}_i^T \bar{z}) \leq 0.25l \|\hat{S}' \hat{V}^T \bar{z}\| \\ &\leq \sum_{i=1}^l |\hat{v}_i^T \bar{z} s'(\hat{v}_i^T \bar{z})| \leq 0.2239l, \end{aligned} \quad (7)$$

where $\hat{S}' = \text{diag}\{s'(\hat{v}_1^T \bar{z}), \dots, s'(\hat{v}_l^T \bar{z})\}$ is a diagonal matrix. $\|\cdot\|$ and $\|\cdot\|_F$ denote matrix 2-norm and Frobenius norm, respectively.

Assumption 5. On the compact set Ω_z , the ideal NN weights W^* , V^* and the NN approximation error are bounded by

$$\|W^*\| \leq w_m, \quad \|V^*\|_F \leq v_m, \quad |\varepsilon_u(z)| \leq \varepsilon_l \quad (8)$$

with w_m , v_m and ε_l being positive constants.

3.2. Projection algorithm

In this paper, we use the following projection mapping. Let $\hat{\rho}_{\theta_{ij}, \min}$ and $\hat{\rho}_{\theta_{ij}, \max}$ be the fictitious lower and upper bound for θ_{ij} , where θ could be any of the unknown weight vector or matrix. Based on these fictitious lower and upper bounds, the same as in [7,8,16], a discontinuous projection mapping $\text{Proj}(\cdot)$ can be defined as $\text{Proj}_{\hat{\theta}}(\cdot) = \{\text{Proj}_{\hat{\theta}}(*_{ij})\}$ with its ij th element being

$$\begin{aligned} &\text{Proj}_{\hat{\theta}}(*_{ij}) \\ &= \begin{cases} -*_{ij} & \text{if } \begin{cases} \hat{\theta}_{ij} = \hat{\rho}_{\theta_{ij}, \max} \text{ and } *_{ij} < 0, \\ \hat{\theta}_{ij} = \hat{\rho}_{\theta_{ij}, \min} \text{ and } *_{ij} > 0, \end{cases} \\ *_{ij} & \text{otherwise,} \end{cases} \end{aligned} \quad (9)$$

where $*$ denotes a vector or a matrix, then $*_{ij}$ denotes its element.

In this paper, all parameter estimates will be updated by the projection type of adaptation laws given by

$$\hat{\Theta}(k + \tau) = \hat{\Theta}(k) - \text{Proj}_{\hat{\theta}}(\Gamma \eta), \quad (10)$$

where $\Gamma = \Gamma^T > 0$ is any diagonal positive-definite adaptation matrix with proper dimension, and η is any adaptation function. For simplicity, assume $\Gamma = \lambda I$ with λ being a positive constant. Similar to [7], we have the following lemma which indicates the nice properties of the above projection type of adaptation law.

Lemma 2. Considering the projection algorithm (9) and parameter adaptation laws (10) used in this paper, the following properties hold:

1. The parameter estimates are always within the known prescribed range, i.e., $\hat{\rho}_{\theta_{ij}, \min} \leq \hat{\theta}_{ij} \leq \hat{\rho}_{\theta_{ij}, \max}$.
2. In addition, if the true parameter Θ is actually within the prescribed range, noting $\tilde{\Theta} = \hat{\Theta} - \Theta$, then

$$\tilde{\Theta}^T (\Gamma^{-1} \text{Proj}_{\hat{\theta}}(\Gamma \eta) - \eta) \geq 0 \quad \text{if } \Theta \text{ is a vector,}$$

$$\text{trace}\{\tilde{\Theta}^T (\Gamma^{-1} \text{Proj}_{\hat{\theta}}(\Gamma \eta) - \eta)\} \geq 0$$

if Θ is a matrix.

Proof. According to the projection algorithm (9) and adaptation law (10), it is obvious that the first property always holds. Now we prove the second property.

If Θ is a vector, consider the diagonal positive-definite adaptation matrix Γ , noticing that the possible effect of projection operator $\text{Proj}_{\hat{\theta}}(*_{ij})$ is to change the sign of $*_{ij}$, we have

$$\begin{aligned} \tilde{\Theta}^T (\Gamma^{-1} \text{Proj}_{\hat{\theta}}(\Gamma \eta) - \eta) &= \tilde{\Theta}^T (\Gamma^{-1} \Gamma \text{Proj}_{\hat{\theta}}(\eta) - \eta) \\ &= \tilde{\Theta}^T (\text{Proj}_{\hat{\theta}}(\eta) - \eta) \\ &= \sum_{i=1}^l \tilde{\theta}_i (\text{Proj}_{\hat{\theta}}(\eta_i) - \eta_i). \end{aligned}$$

Then considering

$$\tilde{\theta}_i (\text{Proj}_{\hat{\theta}}(\eta_i) - \eta_i)$$

$$\begin{aligned}
&= (\hat{\Theta}_i - \Theta_i)(\text{Proj}_{\hat{\Theta}}(\eta_i) - \eta_i), \\
&= \begin{cases} (\hat{\Theta}_i - \Theta_i)(-\eta_i - \eta_i) > 0 \\ \text{if } \begin{cases} \hat{\Theta}_i = \hat{\rho}_{\Theta_i, \max} \text{ means } (\hat{\Theta}_i - \Theta_i) > 0 \\ \text{and } \eta_i < 0, \\ \hat{\Theta}_i = \hat{\rho}_{\Theta_i, \min} \text{ means } (\hat{\Theta}_i - \Theta_i) < 0 \\ \text{and } \eta_i > 0, \end{cases} \\ (\hat{\Theta}_i - \Theta_i)(\eta_i - \eta_i) = 0 \\ \text{otherwise,} \end{cases}
\end{aligned}$$

we have $\tilde{\Theta}^T(\Gamma^{-1}\text{Proj}_{\hat{\Theta}}(\Gamma\eta) - \eta) \geq 0$ holds.

If Θ is a matrix, following the same procedure, we have

$$\text{trace}\{\tilde{\Theta}^T(\Gamma^{-1}\text{Proj}_{\hat{\Theta}}(\Gamma\eta) - \eta)\} \geq 0.$$

Its proof is omitted here for clarity. \square

4. Controller design

Considering the universal approximate ability of MNN, in this section, we use MNN to approximate the IDFC control. The use of MNNs in discrete-time nonlinear system control is not only challenging but also of academic interest.

At first, considering the MNNs, neural weights adaptation laws and projection algorithms used in this paper, we have the following lemma.

Lemma 3. *Considering the projection algorithms we used, on the compact set Ω_z , the estimate NN weights \hat{W} , \hat{V} , and the weight approximation errors \tilde{W} , \tilde{V} are bounded by*

$$\|\hat{W}\| \leq \hat{w}_m, \|\hat{V}\|_F \leq \hat{v}_m, \|\tilde{W}\| \leq \tilde{w}_m, \|\tilde{V}\|_F \leq \tilde{v}_m \quad (11)$$

where $\tilde{W} = \hat{W} - W^*$, $\tilde{V} = \hat{V} - V^*$ and \tilde{w}_m , \tilde{v}_m , \hat{w}_m , \hat{v}_m are positive constants.

In this paper, we use the following adaptive functions:

$$\eta_w(k) = \hat{S}(k - \tau)e(k), \quad (12)$$

$$\eta_v(k) = z_l \hat{W}^T(k - \tau) \hat{S}'(k - \tau)e(k), \quad (13)$$

where z_l is dimension compatible with $\hat{V}(k)$ and is defined as $z_l = [1/\sqrt{l}, \dots, 1/\sqrt{l}]^T$ with $\|z_l\| = 1$, $\hat{S}'(k) =$

$\text{diag}[\hat{S}'_1(k - \tau), \dots, \hat{S}'_l(k - \tau)]$ is a diagonal matrix and $\hat{S}'_i(k - \tau) = s'(\hat{v}_i^T \bar{z}(k - \tau))$.

Define the MNNs update laws as follows:

$$\hat{W}(k) = \hat{W}(k - \tau) - \text{Proj}_{\hat{W}}[\Gamma_w \eta_w(k)], \quad (14)$$

$$\hat{V}(k) = \hat{V}(k - \tau) - \text{Proj}_{\hat{V}}[\Gamma_v \eta_v(k)]. \quad (15)$$

Subtracting W^* and V^* to the both sides of the Eqs. (14) and (15), we obtain

$$\begin{aligned}
\tilde{W}(k) &= \tilde{W}(k - \tau) - \text{Proj}_{\hat{W}}(\Gamma_w \eta_w(k)) \\
&= \tilde{W}(k - \tau) - \text{Proj}_{\hat{W}}[\Gamma_w \hat{S}(k - \tau)e(k)], \quad (16)
\end{aligned}$$

$$\begin{aligned}
\tilde{V}(k) &= \tilde{V}(k - \tau) - \text{Proj}_{\hat{V}}(\Gamma_v \eta_v(k)) \\
&= \tilde{V}(k - \tau) \\
&\quad - \text{Proj}_{\hat{V}}[\Gamma_v z_l \hat{W}^T(k - \tau) \hat{S}'(k - \tau)e(k)], \quad (17)
\end{aligned}$$

where $\Gamma_w = \Gamma_w^T = \lambda_w I$ and $\Gamma_v = \Gamma_v^T = \lambda_v I$.

Lemma 4. *Considering Lemma 2, we have the following inequalities:*

$$\tilde{W}^T(\Gamma_w^{-1} \text{Proj}_{\hat{W}}(\Gamma_w \eta_w) - \eta_w) \geq 0, \quad (18)$$

$$\text{trace}\{\tilde{V}^T(\Gamma_v^{-1} \text{Proj}_{\hat{V}}(\Gamma_v \eta_v) - \eta_v)\} \geq 0. \quad (19)$$

Furthermore, we have

$$\text{Proj}_{\hat{W}}^T \Gamma_w^{-1} \text{Proj}_{\hat{W}} - \eta_w^T \Gamma_w^T \eta_w = 0, \quad (20)$$

$$\text{trace}\{\text{Proj}_{\hat{V}}^T \Gamma_v^{-1} \text{Proj}_{\hat{V}}\} - \text{trace}\{\eta_v^T \Gamma_v^T \eta_v\} = 0, \quad (21)$$

where $\text{Proj}_{\hat{W}} = \text{Proj}_{\hat{W}}(\Gamma_w \eta_w)$ and $\text{Proj}_{\hat{V}} = \text{Proj}_{\hat{V}}(\Gamma_v \eta_v)$.

Proof. It is obvious that following Lemma 2, inequalities (18) and (19) hold.

Considering Eq. (20), because $\Gamma_w = \lambda_w I$, we have

$$\begin{aligned}
&\text{Proj}_{\hat{W}}^T \Gamma_w^{-1} \text{Proj}_{\hat{W}} - \eta_w^T \Gamma_w^T \eta_w \\
&= \text{Proj}_{\hat{W}}^T(\eta_w) \Gamma_w^T \Gamma_w^{-1} \Gamma_w \text{Proj}_{\hat{W}}(\eta_w) - \eta_w^T \Gamma_w^T \eta_w \\
&= \lambda_w \text{Proj}_{\hat{W}}^T(\eta_w) \text{Proj}_{\hat{W}}(\eta_w) - \lambda_w \eta_w^T \eta_w \\
&= \lambda_w [\text{Proj}_{\hat{W}}^T(\eta_w) \text{Proj}_{\hat{W}}(\eta_w) - \eta_w^T \eta_w] \\
&= 0;
\end{aligned}$$

then Eq. (20) holds.

Considering Eq. (21), because $\Gamma_v = \lambda_v I$, we have

$$\begin{aligned} & \text{trace}\{\text{Proj}_{\hat{V}}^T \Gamma_v^{-1} \text{Proj}_{\hat{V}}\} - \text{trace}\{\eta_v^T \Gamma_v^T \eta_v\} \\ &= \text{trace}\{\text{Proj}_{\hat{V}}^T (\eta_v) \Gamma_v^T \Gamma_v^{-1} \Gamma_v \text{Proj}_{\hat{V}} (\eta_v)\} \\ & \quad - \text{trace}\{\eta_v^T \Gamma_v^T \eta_v\} \\ &= \lambda_v \text{trace}\{\text{Proj}_{\hat{V}}^T (\eta_v) \text{Proj}_{\hat{V}} (\eta_v) - \eta_v^T \eta_v\} \\ &= 0; \end{aligned}$$

then Eq. (21) holds. \square

Choose the practical control input as

$$u(k) = \hat{W}^T(k) S(\hat{V}^T(k) \bar{z}), \quad (22)$$

where $\bar{z} = [z^T, 1]^T$ with $z = [\bar{y}_k, \bar{u}_{k-1}, y_m(k + \tau)]^T$.

Noticing Eq. (5), then we have

$$\begin{aligned} & u(k) - u^*(k) \\ &= \hat{W}^T(k) S(\hat{V}^T(k) \bar{z}) \\ & \quad - W^{*T}(k) S(V^{*T}(k) \bar{z}) - \varepsilon_u(z) \\ &= \hat{W}^T(k) S(\hat{V}^T(k) \bar{z}) - W^{*T}(k) S(\hat{V}^T(k) \bar{z}) \\ & \quad + W^{*T}(k) S(\hat{V}^T(k) \bar{z}) \\ & \quad - W^{*T}(k) S(V^{*T}(k) \bar{z}) - \varepsilon_u(z) \\ &= \tilde{W}^T(k) S(\hat{V}^T(k) \bar{z}) \\ & \quad + W^{*T}(k) [S(\hat{V}^T(k) \bar{z}) - S(V^{*T}(k) \bar{z})] - \varepsilon_u(z) \\ &= \tilde{W}^T \hat{S} + W^{*T}(\hat{S} - S^*) - \varepsilon_u(z). \end{aligned}$$

Substituting $u(k)$ into the error equation (2), then we have

$$\begin{aligned} e(k + \tau) &= -y_m(k + \tau) + f(\bar{y}_k, \hat{W}^T(k) \\ & \quad \times S(\hat{V}^T(k) \bar{z}), \bar{u}_{k-1}, 0) + \delta_{d_k} \\ &= -y_m(k + \tau) + f(\bar{y}_k, u^*(k) + \tilde{W}^T \hat{S} \\ & \quad + W^{*T}(\hat{S} - S^*) - \varepsilon_u(z), \bar{u}_{k-1}, 0) \\ & \quad + \delta_{d_k}. \end{aligned} \quad (23)$$

Using Mean Value Theorem, noticing Eq. (4), then the above equation becomes

$$\begin{aligned} & e(k + \tau) \\ &= -y_m(k + \tau) + f(\bar{y}_k, u^*(k), \bar{u}_{k-1}, 0) \\ & \quad + \left. \frac{\partial f}{\partial u} \right|_{u=\xi} (\tilde{W}^T \hat{S} + W^{*T}(\hat{S} - S^*) - \varepsilon_u(z)) \\ & \quad + \delta_{d_k} \\ &= [\tilde{W}^T \hat{S} + W^{*T}(\hat{S} - S^*) - \varepsilon_u(z)] f_u + \delta_{d_k}, \end{aligned} \quad (24)$$

where

$$f_u = \left. \frac{\partial f}{\partial u} \right|_{u=\xi}, \quad \text{where } \xi \in [u^*(k), u(k)].$$

Theorem 1. For the non-affine discrete-time system (1), NN controller (22) and NN weight update laws (14) and (15). There exist compact sets Ω_y , Ω_w , Ω_v and positive constants l^* , λ_w^* and λ_v^* such that if

- (i) the initial parameter set $\Omega_{y_0} \in \Omega_y$, $\Omega_{w_0} \in \Omega_w$, $\Omega_{v_0} \in \Omega_v$;
- (ii) the neural number $l > l^*$, adaptive gain $\lambda_w < \lambda_w^*$, with λ_w^* being the eigenvalue of Γ_w , $\lambda_v < \lambda_v^*$, with λ_v^* being the eigenvalue of Γ_v , satisfying the following parameter condition:
$$\frac{1}{g_1} - \lambda_w l - 0.0625 \lambda_v \hat{w}_m^2 l^2 > 0; \quad (25)$$
- (iii) the initial future output sequence $y(k_0), \dots, y(k_0 + \tau - 1)$ are kept in the compact set Ω_y , initial input sequence $u(k_0)$ are kept in the compact set Ω_u ;

then the output of system (1) will track the desired trajectory and the tracking error is bounded. The closed-loop system is semi globally uniformly ultimately bounded (SGUUB).

Proof. See Appendix A. \square

Remark 2. In the literature, to guarantee the convergence of the parameter estimates, persistently exciting (PE) condition is usually required [15], though it is very difficult to satisfy in practice. Recognizing this difficulty, projection algorithm is used to guarantee the estimated MNN weights remain bounded as

detailed in Lemma 2, rather than converge to their optimal values.

Remark 3. In practical applications, if the bound of system sensitivity g_1 is known, we can pre-define \hat{w}_m which is sufficient large. Then, by choosing sufficient small learning rates λ_w and λ_v , we can guarantee the existence of parameter condition (25)

$$\frac{1}{g_1} - \lambda_w l - 0.0625 \lambda_v \hat{w}_m^2 l^2 > 0.$$

However, if the bound of system sensitivity g_1 is unknown, trial and error may have to be used starting with very small learning rates.

5. Simulation studies

Consider the following non-affine continuous stirred tank reactors (CSTR) system [4]:

$$\begin{aligned} \dot{x}_1 &= 1 - x_1 - a_0 x_1 e^{-10^4/x_2}, \\ \dot{x}_2 &= 350 - x_2 + a_1 x_1 e^{-10^4/x_2} \\ &\quad + a_3 u (1 - e^{-a_2/u}) (350 - x_2), \\ y &= x_1, \end{aligned} \quad (26)$$

where $a_1 = 1.44 \times 10^{13}$, $a_2 = 6.987 \times 10^2$ and $a_3 = 0.01$. (Detailed definition can be found in [4].) The major challenge of this control problem is that the plant does not assume the customary control affine system structure because the control input u appears nonlinearly.

The control objective is to make the output $y(t)$ track the set-point step change signal $y_d(t)$. In order to get a smooth reference signal, a linear reference model is used to shape the discontinuous reference signal for providing the desired signals y_d . The following reference model is to be implemented: $y_d(s)/r(s) = \omega_n^2/(s^2 + 2\zeta_n\omega_n s + \omega_n^2)$ where the natural frequency $\omega_n = 5.0$ rad/min and the damping ratio $\zeta_n = 1.0$.

The operating condition of the CSTR system are restricted to

$$\Omega_x = \{(x_1, x_2, u) | 0.02 < x_1 < 0.8, 350 \leq x_2 < T_{\max}, 0 \leq u \leq u_{\max}\} \quad (27)$$

where the constants T_{\max} and u_{\max} are the maximum values of the coolant flow rate and the tank temperature, respectively.

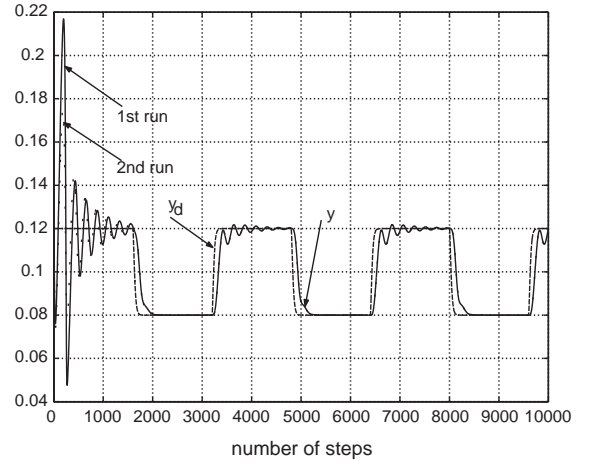


Fig. 1. Tracking performance.

By using first-order approximation, $x_1(k+1) = x_1(k) + \dot{x}_1(k)T$ and $x_2(k+1) = x_2(k) + \dot{x}_2(k)T$, we can see that the discretization of (26) can be written as follows: (Details are omitted here due to the space limitation.)

$$y(k+2) = f_0(y(k), y(k-1), u(k), u(k-1))$$

with $f_0(\cdot)$ being a nonlinear function. We can see that the discretized system is in the form of (1). Firstly, we can easily obtain that Assumptions 1, 2 and 5 hold. It is shown in [4] that Assumption 3 holds. By choosing sufficiently small learning rates λ_w and λ_v , the parameter condition (25) in Theorem 1 can be guaranteed. Therefore, all the numerical conditions hold.

It can be obtained that the delay of the discretized system is $\tau = 2$. The control action starts from $k = 1$. System initial conditions are $x(k) = [0.1, 440]^T$ ($k = 0, 1, 2$). Number of neurons used is $l = 40$. NN weights are initialized as $\hat{W}(k) = 0$ and $\hat{V}(k) = 0$ ($k = 0, 1, 2$). The adaptation gain matrix are chosen as $\Gamma_w = 0.1I$ and $\Gamma_v = 0.15I$. The fictitious lower and upper bounds for every element of weights $\hat{W}(k)$ and $\hat{V}(k)$ are assumed to be -5 and $+5$.

Simulation results are shown in Figs. 1–3. It can be seen that all the practical output of the system converge to a neighborhood of the desired trajectory and all the signals in the closed-loop system are bounded.

Remark 4. In practical applications, the high oscillation of the output is an undesirable behavior and

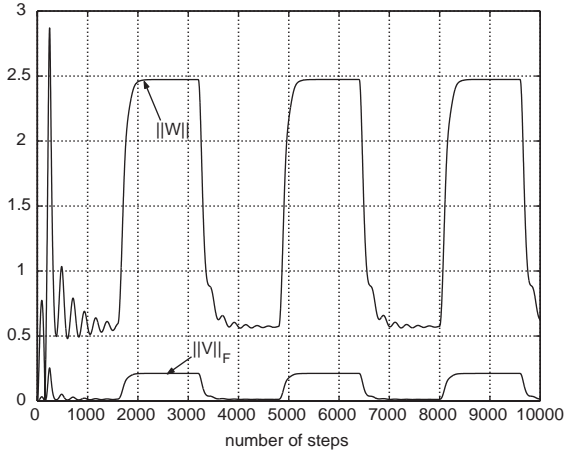


Fig. 2. Weight norm.

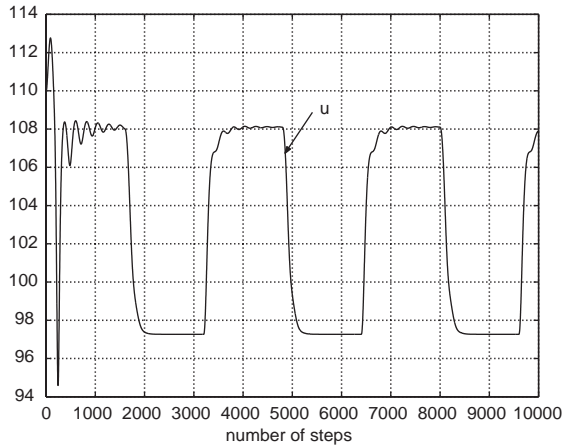


Fig. 3. Control trajectory.

should be reduced. By training the NN weights, the high oscillation can be reduced. As clearly indicated in Fig. 1, by using the NN weights at the end of the first time run as the initial value of the second time run, the oscillation peak was reduced, though was not completely eliminated.

6. Conclusion

In this paper, an adaptive MNN control scheme was proposed for a class of discrete-time non-affine non-linear systems in NARMAX form. The existence of

IDFC, which can drive the system output to track desired trajectory, was proved by using implicit function theorem. Based on the input–output model, MNNs were used to emulate the IDFC. Lyapunov stability techniques and projection algorithms were used to develop the control scheme and adaptive learning laws. The proposed controller guarantees the stability and of closed-loop system. A non-affine CSTR system was studied as a simulation example. The effectiveness of the proposed control method was illustrated through numerical simulation study.

Appendix A. Proof of Theorem 1

We have illustrated that there exists an ideal control $u^*(k)$ which guarantees that $e(k + \tau) = 0$ if there is no unknown disturbances. Since all the assumptions are only valid in compact sets Ω_y and Ω_u , we must prove that the system outputs and inputs will remain in these compact sets all the time indeed. At time instant k , suppose that all past inputs are in Ω_u , current output and all past outputs are in Ω_y , we will prove that all these conditions still hold after time instant k and the tracking error converges into a small neighborhood of zero.

Choose the Lyapunov function as follows:

$$J(k) = \frac{1}{g_1} \sum_{j=0}^{\tau-1} e^2(k+j) + \sum_{j=0}^{\tau-1} \tilde{W}^T(k+j) \Gamma_w^{-1} \tilde{W}(k+j) + \sum_{j=0}^{\tau-1} \text{tr}\{\tilde{V}^T(k+j) \Gamma_v^{-1} \tilde{V}(k+j)\}.$$

Apparently, the Lyapunov function candidate $J(k)$ contains the states of the error dynamics of the system, and the parameter adaptation. It should be note that the future variables, $e(k+1), \dots, e(k+\tau-1)$ and $\tilde{W}(k+1), \dots, \tilde{W}(k+\tau-1)$, are all determined at time instant k as they are independent of current control $u(k)$.

The first difference of $J(k)$ is given as

$$\begin{aligned}\Delta J(k) &= J(k+1) - J(k) \\ &= \frac{1}{g_1} [e^2(k+\tau) - e^2(k)] + \tilde{W}^T(k+\tau) \\ &\quad \times \Gamma_w^{-1} \tilde{W}(k+\tau) - \tilde{W}^T(k) \Gamma_w^{-1} \tilde{W}(k) \\ &\quad + \text{tr}\{\tilde{V}^T(k+\tau) \Gamma_v^{-1} \tilde{V}(k+\tau) \\ &\quad - \tilde{V}^T(k) \Gamma_v^{-1} \tilde{V}(k)\}.\end{aligned}$$

Considering the neural network weight update laws (16) and (17), we have

$$\begin{aligned}\Delta J(k) &= \frac{1}{g_1} [e^2(k+\tau) - e^2(k)] \\ &\quad - \tilde{W}^T(k) \Gamma_w^{-1} \text{Proj}_{\hat{W}}(k+\tau) \\ &\quad - \text{Proj}_{\hat{W}}^T(k+\tau) \Gamma_w^{-1} \tilde{W}(k) \\ &\quad + \text{Proj}_{\hat{W}}^T(k+\tau) \Gamma_w^{-1} \text{Proj}_{\hat{W}}(k+\tau) \\ &\quad + \text{tr}\{-\tilde{V}^T(k) \Gamma_v^{-1} \text{Proj}_{\hat{V}}(k+\tau) \\ &\quad - \text{Proj}_{\hat{V}}^T(k+\tau) \Gamma_v^{-1} \tilde{V}(k) \\ &\quad + \text{Proj}_{\hat{V}}^T(k+\tau) \Gamma_v^{-1} \text{Proj}_{\hat{V}}(k+\tau)\}.\end{aligned}\tag{A.1}$$

Considering the projection algorithms used, there are four possible conditions:

1. All the elements of $\hat{W}(k)$ and $\hat{V}(k)$ are within the known prescribed fictitious bounds;
2. Only some elements of weight vector $\hat{W}(k)$ reach the fictitious bounds, projection algorithm (14) is applied;
3. Only some elements of weight matrix $\hat{V}(k)$ reach the fictitious bounds, projection algorithm (15) is applied;
4. Some elements of both $\hat{W}(k)$ and $\hat{V}(k)$ reach the fictitious bound, projection algorithms (14) and (15) are applied.

We will discuss them one by one in details below.

Condition 1: When all the elements of weight $\hat{W}(k)$ and $\hat{V}(k)$ are within the known prescribed bounds,

Eq. (A.1) becomes

$$\begin{aligned}\Delta J(k) &= \frac{1}{g_1} [e^2(k+\tau) - e^2(k)] - \tilde{W}^T(k) \Gamma_w^{-1} \\ &\quad \times \Gamma_w \eta_w(k+\tau) - (\Gamma_w \eta_w(k+\tau))^T \Gamma_w^{-1} \tilde{W}(k) \\ &\quad + (\Gamma_w \eta_w(k+\tau))^T \Gamma_w^{-1} \Gamma_w \eta_w(k+\tau) \\ &\quad + \text{tr}\{-\tilde{V}^T(k) \Gamma_v^{-1} \Gamma_v \eta_v(k+\tau) \\ &\quad - (\Gamma_v \eta_v(k+\tau))^T \Gamma_v^{-1} \tilde{V}(k) \\ &\quad + (\Gamma_v \eta_v(k+\tau))^T \Gamma_v^{-1} \Gamma_v \eta_v(k+\tau)\} \\ &= \frac{1}{g_1} [e^2(k+\tau) - e^2(k)] - 2\tilde{W}^T(k) \eta_w(k+\tau) \\ &\quad + \eta_w^T(k+\tau) \Gamma_w^T \eta_w(k+\tau) \\ &\quad - 2\text{tr}\{\tilde{V}^T(k) \eta_v(k+\tau)\} \\ &\quad + \text{tr}\{\eta_v^T(k+\tau) \Gamma_v^T \eta_v(k+\tau)\}.\end{aligned}$$

From Eq. (24), we obtain

$$\begin{aligned}\tilde{W}^T(k) \hat{S}(k) &= \frac{e(k+\tau) - \delta_{d_k}}{f_u} \\ &\quad - W^{*T}(\hat{S}(k) - S^*) + \varepsilon_u(z).\end{aligned}$$

Furthermore, considering the adaptive function (12) and (13), noticing that $\text{tr}\{\tilde{V}^T z_l \hat{W}^T \hat{S}'\} = \hat{W}^T \hat{S}' \tilde{V}^T z_l$ and $\text{tr}\{(z_l \hat{W}^T \hat{S}')^T \Gamma_v^T(z_l \hat{W}^T \hat{S}')\} = \lambda_v \|z_l \hat{W}^T \hat{S}'\|_F^2$, then

$$\begin{aligned}\Delta J(k) &= \left[\frac{1}{g_1} - \frac{2}{f_u}\right] e^2(k+\tau) - \frac{1}{g_1} e^2(k) \\ &\quad + 2W^{*T}(\hat{S} - S^*)e(k+\tau) \\ &\quad - 2\left[\varepsilon_u(z) - \frac{\delta_{d_k}}{f_u}\right] e(k+\tau) \\ &\quad + \hat{S}^T \Gamma_w^T \hat{S} e^2(k+\tau) - 2\hat{W}^T \hat{S}' \tilde{V}^T z_l e(k+\tau) \\ &\quad + \lambda_v \|z_l \hat{W}^T \hat{S}'\|_F^2 e^2(k+\tau).\end{aligned}$$

Noticing Assumption 5 and the following facts (please see Appendix B for details):

$$\begin{aligned} -\frac{2}{f_u} &\leq -\frac{2}{g_1}, \\ 2W^{*\text{T}}(\hat{S} - S^*)e(k + \tau) &\leq 4\|W^*\|\sqrt{l}|e(k + \tau)|, \\ -2\left[\varepsilon_u(z) - \frac{\delta_{d_k}}{f_u}\right]e(k + \tau) \\ &\leq 2\left[\varepsilon_l + \frac{\tau g_2 d}{\epsilon}\right]|e(k + \tau)|, \\ \hat{S}^{\text{T}}\Gamma_w^{\text{T}}\hat{S}e^2(k + \tau) &\leq \lambda_w l e^2(k + \tau), \end{aligned}$$

$$\begin{aligned} -2\hat{W}^{\text{T}}\hat{S}'\tilde{V}^{\text{T}}z_l e(k + \tau) &\leq 0.5\hat{w}_m l \tilde{v}_m |e(k + \tau)|, \\ \lambda_v \|z_l \hat{W}^{\text{T}}\hat{S}'\|_F^2 e^2(k + \tau) &\leq 0.0625\lambda_v \hat{w}_m^2 l^2 e^2(k + \tau), \end{aligned}$$

we have

$$\begin{aligned} \Delta J(k) &\leq -\left[\frac{1}{g_1} - \lambda_w l - 0.0625\lambda_v \hat{w}_m^2 l^2\right] \\ &\quad \times e^2(k + \tau) - \frac{1}{g_1} e^2(k) \\ &\quad + \left[4\|W^*\|\sqrt{l} + 2\varepsilon_l + \frac{2\tau g_2 d}{\epsilon}\right. \\ &\quad \left.+ 0.5\hat{w}_m l \tilde{v}_m\right]|e(k + \tau)|. \end{aligned}$$

If we choose the parameters to satisfy the following condition:

$$\frac{1}{g_1} - \lambda_w l - 0.0625\lambda_v \hat{w}_m^2 l^2 > 0 \quad (\text{A.2})$$

and define positive constants

$$\alpha = \frac{1}{g_1} - \lambda_w l - 0.0625\lambda_v \hat{w}_m^2 l^2,$$

$$\beta = 4\|W^*\|\sqrt{l} + 2\varepsilon_l + 2\frac{\tau g_2 d}{\epsilon} + 0.5\hat{w}_m l \tilde{v}_m,$$

then

$$\begin{aligned} \Delta J(k) &\leq -\alpha|e(k + \tau)|^2 + \beta|e(k + \tau)| - \frac{1}{g_1}e^2(k) \\ &\leq -\alpha|e(k + \tau)|^2 + \beta|e(k + \tau)| \\ &= -\alpha|e(k + \tau)|\left(|e(k + \tau)| - \frac{\beta}{\alpha}\right). \end{aligned}$$

Define compact set

$$\Omega_e \triangleq \left\{e(k) \mid |e(k)| < \frac{\beta}{\alpha}\right\},$$

we can see that once $|e(k + \tau)|$ is out of the compact set Ω_e , $\Delta J(k) < 0$. That means $e(k + \tau)$ will converge to the compact set denoted by Ω_e .

Now it still remains to show that the weight estimates $\hat{W}(k)$ and $\hat{V}(k)$ are bounded. Considering the projection algorithms we used, it is obvious that $\hat{W}(k)$ and $\hat{V}(k)$ are bounded in compact sets.

Finally, for all $k \geq 0$, $J(k)$ is bounded. As $k \rightarrow +\infty$, we have

$$\begin{aligned} J(\infty) &= \frac{1}{g_1} \sum_{j=0}^{\tau-1} e^2(\infty + j) \\ &\quad + \sum_{j=0}^{\tau-1} \tilde{W}^{\text{T}}(\infty + j)\Gamma_w^{-1}\tilde{W}(\infty + j) \\ &\quad + \sum_{j=0}^{\tau-1} \text{tr}\{\tilde{V}^{\text{T}}(\infty + \tau)\Gamma_v^{-1}\tilde{V}(\infty + \tau)\}. \end{aligned}$$

Because we have proved that $e(k)$ is bounded, $\tilde{W}(k)$ and $\tilde{V}(k)$ are all bounded by the projection algorithms, we obtain $J(\infty) < \infty$, that is to say $J(k)$ is also bounded.

Condition 2: When only some elements of $\hat{W}(k)$ reach the fictitious bounds, Eq. (A.1) becomes

$$\begin{aligned} \Delta J(k) &= \frac{1}{g_1}[e^2(k + \tau) - e^2(k)] \\ &\quad - \tilde{W}^{\text{T}}(k)\Gamma_w^{-1}\text{Proj}_{\hat{W}}(k + \tau) \\ &\quad - \text{Proj}_{\hat{W}}^{\text{T}}(k + \tau)\Gamma_w^{-1}\tilde{W}(k) \\ &\quad + \text{Proj}_{\hat{W}}^{\text{T}}(k + \tau)\Gamma_w^{-1}\text{Proj}_{\hat{W}}(k + \tau) \\ &\quad + \text{tr}\{-\tilde{V}^{\text{T}}(k)\Gamma_v^{-1}\Gamma_v\eta_v(k + \tau) \\ &\quad - (\Gamma_v\eta_v(k + \tau))^{\text{T}}\Gamma_v^{-1}\tilde{V}(k) \\ &\quad + (\Gamma_v\eta_v(k + \tau))^{\text{T}}\Gamma_v^{-1}\Gamma_v\eta_v(k + \tau)\}. \end{aligned}$$

Adding and subtracting $-\tilde{W}^{\text{T}}(k)\eta_w(k + \tau) - \eta_w^{\text{T}}(k + \tau)\tilde{W}(k) + \eta_w^{\text{T}}(k + \tau)\Gamma_w^{\text{T}}\eta_w(k + \tau)$ to the right-hand side of the above equation, we obtain

$$\begin{aligned} \Delta J(k) &= \frac{1}{g_1}[e^2(k + \tau) - e^2(k)] \\ &\quad - \tilde{W}^{\text{T}}(k)\eta_w(k + \tau) - \eta_w^{\text{T}}(k + \tau)\tilde{W}(k) \end{aligned}$$

$$\begin{aligned}
& + \eta_w^T(k + \tau) \Gamma_w^T \eta_w(k + \tau) \\
& + \text{tr} \{ -\tilde{V}^T(k) \Gamma_v^{-1} \Gamma_v \eta_v(k + \tau) \\
& \quad - (\Gamma_v \eta_v(k + \tau))^T \Gamma_v^{-1} \tilde{V}(k) \\
& \quad + (\Gamma_v \eta_v(k + \tau))^T \Gamma_v^{-1} \Gamma_v \eta_v(k + \tau) \} \\
& - \tilde{W}^T(k) \Gamma_w^{-1} \text{Proj}_{\hat{W}}(k + \tau) \\
& - \text{Proj}_{\hat{W}}^T(k + \tau) \Gamma_w^{-1} \tilde{W}(k) \\
& + \text{Proj}_{\hat{W}}^T(k + \tau) \Gamma_w^{-1} \text{Proj}_{\hat{W}}(k + \tau) \\
& + \tilde{W}^T(k) \eta_w(k + \tau) + \eta_w^T(k + \tau) \tilde{W}(k) \\
& - \eta_w^T(k + \tau) \Gamma_w^T \eta_w(k + \tau) \\
& = \frac{1}{g_1} [e^2(k + \tau) - e^2(k)] - \tilde{W}^T(k) \eta_w(k + \tau) \\
& - \eta_w^T(k + \tau) \tilde{W}(k) \\
& + \eta_w^T(k + \tau) \Gamma_w^T \eta_w(k + \tau) \\
& + \text{tr} \{ -\tilde{V}^T(k) \Gamma_v^{-1} \Gamma_v \eta_v(k + \tau) \\
& \quad - (\Gamma_v \eta_v(k + \tau))^T \Gamma_v^{-1} \tilde{V}(k) \\
& \quad + (\Gamma_v \eta_v(k + \tau))^T \Gamma_v^{-1} \Gamma_v \eta_v(k + \tau) \} \\
& - \tilde{W}^T(k) (\Gamma_w^{-1} \text{Proj}_{\hat{W}}(k + \tau) - \eta_w(k + \tau)) \\
& - (\text{Proj}_{\hat{W}}^T(k + \tau) \Gamma_w^{-1} - \eta_w^T(k + \tau)) \tilde{W}(k) \\
& + \text{Proj}_{\hat{W}}^T(k + \tau) \Gamma_w^{-1} \text{Proj}_{\hat{W}}(k + \tau) \\
& - \eta_w^T(k + \tau) \Gamma_w^T \eta_w(k + \tau).
\end{aligned}$$

Noticing Lemma 4, using Eqs. (18) and (20), we have

$$\begin{aligned}
\Delta J(k) & \leq \frac{1}{g_1} [e^2(k + \tau) - e^2(k)] - \tilde{W}^T(k) \eta_w(k + \tau) \\
& - \eta_w^T(k + \tau) \tilde{W}(k) + \eta_w^T(k + \tau) \Gamma_w^T \eta_w(k + \tau) \\
& + \text{tr} \{ -\tilde{V}^T(k) \Gamma_v^{-1} \Gamma_v \eta_v(k + \tau) \\
& \quad - (\Gamma_v \eta_v(k + \tau))^T \Gamma_v^{-1} \tilde{V}(k) \\
& \quad + (\Gamma_v \eta_v(k + \tau))^T \Gamma_v^{-1} \Gamma_v \eta_v(k + \tau) \}
\end{aligned}$$

which is the same as we discussed in Condition 1. Thus, we obtain the same stability results.

Following the same procedure in the proof of Condition 2, by noting Lemma 4, using Eqs. (18)–

(21), Conditions 3 and 4 also can be transformed to Condition 1.

We can see that Conditions 2–4 can be transformed to Condition 1, in which we have proved that the tracking error to be bounded in a compact set, then Theorem 1 holds. \square

Appendix B. Explanation of facts used

- $-2/f_u \leq -2/g_1$
Noticing Assumption 3, we obtain the above inequality.
- $2W^{*T}(\hat{S} - S^*)e(k + \tau) \leq 4\|W^*\|\sqrt{l}|e(k + \tau)|$
Because every element of \hat{S} and S^* is less than 1, then $(\hat{S} - S^*) \leq \|\hat{S}\| + \|S^*\| = 2\sqrt{l}$. The above inequality holds.
- $-2[\varepsilon_u(z) - \delta_{d_k}/f_u]e(k + \tau) \leq 2[\varepsilon_l + \tau g_2 d/\varepsilon]|e(k + \tau)|$
Because $\varepsilon_u(z) \leq \varepsilon_l$, $\delta_{d_k} \leq \tau g_2 d$ and $\varepsilon < |f_u|$, the above inequality holds.
- $\hat{S}^T \Gamma_w \hat{S} e^2(k + \tau) \leq \lambda_w l e^2(k + \tau)$
Because $\Gamma_w = \lambda_w I$ is a positive diagonal matrix, $\hat{S}^T \Gamma_w \hat{S} = \lambda_w \hat{S}^T \hat{S}$. Noticing every element of \hat{S} is less than 1, thus the inner product of \hat{S} must be less than l .
- $-2\hat{W}^T \hat{S}' \tilde{V}^T z_l e(k + \tau) \leq 0.5 \hat{w}_m l \tilde{v}_m |e(k + \tau)|$
Noticing Lemma 3, we have $\|\hat{W}\| \leq \hat{w}_m$ and $\|\tilde{V}\|_F \leq \tilde{v}_m$. By definition, $\|z_l\| = 1$. Noticing Eq. (7), we have $\|\hat{S}'\|_F \leq 0.25l$. Thus, the above inequality holds.
- $\lambda_v \|z_l \hat{W}^T \hat{S}'\|_F^2 e^2(k + \tau) \leq 0.0625 \lambda_v \hat{w}_m^2 l^2 e^2(k + \tau)$
Noticing $\|z_l \hat{W}^T \hat{S}'\|_F \leq \|z_l\| \|\hat{W}\| \|\hat{S}'\|_F \leq 0.25l \hat{w}_m$, the above inequality holds.

References

- [1] S.A. Billings, W.S.F. Voon, A prediction-error and stepwise regression algorithm for nonlinear systems, *Internat. J. Control* 44 (1986) 803–822.
- [2] S. Chen, S.A. Billings, Representations of non-linear systems: the narmax model, *Internat. J. Control* 49 (3) (1989) 1013–1032.
- [3] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signals Systems* 4 (1989) 303–312.
- [4] S.S. Ge, C.C. Hang, T.H. Lee, T. Zhang, *Stable Adaptive Neural Network Control*, Kluwer Academic, Boston, MA, 2001.

- [5] S.S. Ge, T.H. Lee, G.Y. Li, J. Zhang, Adaptive NN control for a class of discrete-time nonlinear systems, *Internat. J. Control* 76 (4) (2003) 334–354.
- [6] S.S. Ge, G.Y. Li, T.H. Lee, Adaptive NN control for a class of strict-feedback discrete-time nonlinear systems, *Automatica* 39 (5) (2003) 807–819.
- [7] J.Q. Gong, B. Yao, Neural network adaptive robust control of nonlinear systems in semi-strict feedback form, *Automatica* 37 (8) (2001) 1149–1160.
- [8] G.C. Goodwin, D.Q. Mayne, A parameter estimation perspective of continuous time model reference adaptive control, *Automatica* 23 (1) (1987) 57–70.
- [9] I. Kanellakopoulos, A discrete-time adaptive nonlinear system, *IEEE Trans. Automat. Control* 39 (11) (1994) 2362–2365.
- [10] E.B. Kosmatopoulos, M.M. Polycarpou, M.A. Christodoulou, P.A. Ioannou, High-order neural network structures for identification of dynamical systems, *IEEE Trans. Neural Networks* 6 (2) (1995) 422–431.
- [11] A.U. Levin, K.S. Narendra, Control of nonlinear dynamical systems using neural networks: controllability and stabilization, *IEEE Trans. Neural Networks* 4 (2) (1993) 192–206.
- [12] A.U. Levin, K.S. Narendra, Control of nonlinear dynamical systems using neural networks-part ii: observability, identification, and control, *IEEE Trans. Neural Networks* 7 (1) (1996) 30–42.
- [13] F.L. Lewis, S. Jagannathan, A. Yesildirek, *Neural Network Control of Robot Manipulators and Nonlinear Systems*, Taylor & Francis, London, 1999.
- [14] T. Poggio, F. Girosi, Networks for approximation and learning, *Proc. IEEE* 78 (1996) 1481–1497.
- [15] N. Sadegh, A perception network for functional identification and control of nonlinear systems, *IEEE Trans. Neural Networks* 4 (6) (1993) 982–988.
- [16] S.O. Sastry, M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness*, Prentice-Hall, Englewood Cliff, NJ, 1989.