Research Article

System Identification Using Multilayer Differential Neural Networks: A New Result

J. Humberto Pérez-Cruz,[1] A. Y. Alanis,[1] José de Jesús Rubio,[2] and Jaime Pacheco[2]

[1] *Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, Boulevard Marcelino García Barragán No. 1421, 44430 Guadalajara, JAL, Mexico*

[2] *Sección de Estudios de Posgrado e Investigación, ESIME-UA, IPN, Avenida de las Granjas No. 682, 02250 Santa Catarina, NL, Mexico*

Correspondence should be addressed to J. Humberto Pérez-Cruz, phhhantom2001@yahoo.com.mx

In previous works, a learning law with a dead zone function was developed for multilayer differential neural networks. This scheme requires strictly a priori knowledge of an upper bound for the unmodeled dynamics. In this paper, the learning law is modified in such a way that this condition is relaxed. By this modification, the tuning process is simpler and the dead-zone function is not required anymore. On the basis of this modification and by using a Lyapunov-like analysis, a stronger result is here demonstrated: the exponential convergence of the identification error to a bounded zone. Besides, a value for upper bound of such zone is provided. The workability of this approach is tested by a simulation example.

1. Introduction

During the last four decades system identification has emerged as a powerful and effective alternative to the first principles modeling [1–4]. By using the first approach, a satisfactory mathematical model of a system can be obtained directly from an input and output experimental data set [5]. Ideally no a priori knowledge of the system is necessary since this is considered as a black box. Thus, the time employed to develop such model is reduced significantly with respect to a first principles approach. For the linear case, system identification is a problem well understood and enjoys well-established solutions [6]. However, the nonlinear case is much more challenging. Although some proposals have been presented [7], the class of considered nonlinear systems can result very limited. Due to their capability of handling a more general class of systems and due to advantages such as the fact of not requiring linear in parameters and persistence of excitation assumptions [8],

artificial neural networks (ANNs) have been extensively used in identification of nonlinear systems [9–12]. Their success is based on their capability of providing arbitrarily good approximations to any smooth function [13–15] as well as their massive parallelism and very fast adaptability [16, 17].

An artificial neural network can be simply considered as a nonlinear generic mathematical formula whose parameters are adjusted in order to represent the behavior of a static or dynamic system [18]. These parameters are called weights. Generally speaking, ANN can be classified as feedforward (static) ones, based on the back propagation technique [19] or as recurrent (dynamic) ones [17]. In the first network type, system dynamics is approximated by a static mapping. These networks have two major disadvantages: a slow learning rate and a high sensitivity to training data. The second approach (recurrent ANN) incorporates feedback into its structure. Due to this feature, recurrent neural networks can overcome many problems associated with static ANN, such as global extrema search, and consequently have better approximation properties. Depending on their structure, recurrent neural networks can be classified as discrete-time ones or differential ones.

The first deep insight about the identification of dynamic systems based on neural networks was provided by Narendra and Parthasarathy [20]. However, none-stability analyses of their neuroidentifier were presented. Hunt et al. [21] called attention to determine the convergence, stability and robustness of the algorithms based on neural networks for identification and control. This issue was addressed by Polycarpou and Ioannou [16], Rovithakis and Christodoulou [17], Kosmatopoulos et al. [22], and Yu and Poznyak [23]. Given different structures of continuous-time neural networks, the stability of their algorithms could be proven by using Lyapunov-like analysis. All aforementioned works considered only the case of single-layer networks. However, as it is known, this kind of networks does not necessarily satisfy the property of universal function approximation [24]. And although the activation functions of single-layer neural networks are selected as a basis set in such a way that this property can be guaranteed, the approximation error can never be made smaller than a lower bound [24]. This drawback can be overcome by using multilayer neural networks. Due to this better capability of function approximation, the case multilayer was considered in [25] for feedforward networks and for continuous time recurrent neural networks for first time in [26] and subsequently in [27]. By using Lyapunov-like analysis and a dead-zone function, boundedness for the identification error could be guaranteed in [26]. The following upper bound for the "average" identification error was reported,

$$\limsup_{T \to \infty} \frac{1}{T} \int_0^T \left[1 - \frac{\widetilde{f}_0 + \Upsilon}{\lambda_{\min} \left(P^{-1/2} Q_0 P^{-1/2} \right) \left\| P^{1/2} \Delta_t \right\|} \right]_+ \Delta_t^T Q_0 \Delta_t dt \leq \widetilde{f}_0 + \Upsilon, \qquad (1.1)$$

where Δ_t is the identification error, Q_0 is a positive definite matrix, $\widetilde{f}_0$ is a upper bound for the modeling error, Υ is an upper bound for a deterministic disturbance, and $[\cdot]_+$ is a dead-zone function defined as

$$[z]_+ = \begin{cases} z & z \geq 0, \\ 0 & z < 0. \end{cases} \qquad (1.2)$$

Although, in [28], open-loop analysis based on the passivity method for a multilayer neural network was carried out and certain simplifications were accomplished, the main result about

the aforementioned identification error could not be modified. In [29], the application of the multilayer scheme for control was explored. Since previous works [26–29] are based on this "average" identification error, one could wonder about the real utility of this result. Certainly, boundedness for this kind of error does not guarantee that Δ_t belongs to L_2 or L_∞. Besides, none value for upper bound of identification error norm is provided. Likewise, none information about the speed of the convergence process is presented. Another disadvantage of this approach is that the upper bound for the modeling error $\tilde{f}_0$ must be strictly known a priori in order to implement the learning laws for the weight matrices. In order to avoid these drawbacks, in this paper, we propose to modify the learning laws employed in [26] in such a way that their implementation does not require anymore the knowledge of an upper bound for the modeling error. Besides, on the basis of these new learning laws, a stronger result is here guaranteed: the exponential convergence of the identification error norm to a bounded zone. The workability of the scheme developed in this paper is tested by simulation.

2. Multilayer Neural Identifier

Consider that the nonlinear system to be identified can be represented by

$$\dot{x}_t = f(x_t, u_t, t) + \xi_t, \tag{2.1}$$

where $x_t \in \mathfrak{R}^n$ is the measurable state vector for $t \in \mathfrak{R}^+ := \{t : t \geq 0\}$, $u_t \in \mathfrak{R}^q$ is the control input, $f : \mathfrak{R}^n \times \mathfrak{R}^q \times \mathfrak{R}^+ \rightarrow \mathfrak{R}^n$ is an *unknown* nonlinear vector function which represents the nominal dynamics of the system, and $\xi_t \in \mathfrak{R}^n$ represents a deterministic disturbance. $f(x_t, u_t, t)$ represents a very ample class of systems including affine and nonaffine-in-control nonlinear systems. However, when the control input appears in a nonlinear fashion in the system state equation (2.1), throughout this paper, such nonlinearity with respect to the input is assumed known and represented by $\gamma(\cdot) : \mathfrak{R}^q \rightarrow \mathfrak{R}^s$.

Consider the following parallel structure of multilayer neural network

$$\frac{d}{dt}\hat{x}_t = A\hat{x}_t + W_{1,t}\sigma(V_{1,t}\hat{x}_t) + W_{2,t}\phi(V_{2,t}\hat{x}_t)\gamma(u_t), \tag{2.2}$$

where $\hat{x}_t \in \mathfrak{R}^n$ is the state of the neural network, $u_t \in \mathfrak{R}^q$ is the control input, $A \in \mathfrak{R}^{n \times n}$ is a Hurwitz matrix which can be specified by the designer, the matrices $W_{1,t} \in \mathfrak{R}^{n \times m}$ and $W_{2,t} \in \mathfrak{R}^{n \times r}$ are the weights of output layers, the matrices $V_{1,t} \in \mathfrak{R}^{m \times n}$ and $V_{2,t} \in \mathfrak{R}^{r \times n}$ are the weights of hidden layers, $\sigma(\cdot)$ is the activation vector-function with sigmoidal components, that is, $\sigma(\cdot) := [\sigma_1(\cdot), \ldots, \sigma_m(\cdot)]^T$,

$$\sigma_j(v) := \frac{a_{\sigma j}}{1 + \exp\left(-\sum_{i=1}^m c_{\sigma j,i} v_i\right)} - d_{\sigma j}, \quad \text{for } j = 1, \ldots, m, \tag{2.3}$$

where $a_{\sigma j}$, $c_{\sigma j,i}$, and $d_{\sigma j}$ are positive constants which can be specified by the designer, $\phi(\cdot) : \mathfrak{R}^r \rightarrow \mathfrak{R}^{r \times s}$ is also a sigmoidal function, that is,

$$\phi_{ij}(z) := \frac{a_{\phi ij}}{1 + \exp\left(-\sum_{l=1}^r c_{\phi ij,l} z_l\right)} - d_{\phi ij} \quad \text{for } i = 1, \ldots, r, \ j = 1, \ldots, s, \tag{2.4}$$

where $a_{\phi ij}$, $c_{\phi ij,l}$, and $d_{\phi ij}$ are positive constants which can be specified by the designer, $\gamma(\cdot) : \mathfrak{R}^q \to \mathfrak{R}^s$ represents the nonlinearity with respect to the input—if it exists—which is assumed a priori known for the system (2.1). It is important to mention that m and r, that is, the number of neurons for $\sigma(\cdot)$ and the number of rows for $\phi(\cdot)$, respectively, can be selected by the designer.

The problem of identifying system (2.1) based on the multilayer differential neural network (2.2) consists of, given the measurable state x_t and the input u_t, adjusting on line the weights $W_{1,t}, W_{2,t}, V_{1,t}$, and $V_{2,t}$ by proper learning laws such that the identification error $\Delta_t := \hat{x}_t - x_t$ can be reduced.

Hereafter, it is considered that the following assumptions are valid;

(A1) System (2.1) satisfies the (uniform on t) Lipschitz condition, that is,

$$\left\| f(x,u,t) - f(z,v,t) \right\| \le L_1 \|x - z\| + L_2 \|u - v\|; \quad x, z \in \mathfrak{R}^n;\ u, v \in \mathfrak{R}^q;\ 0 \le L_1, L_2 < \infty. \quad (2.5)$$

(A2) The differences of functions $\sigma(\cdot)$ and $\phi(\cdot)$ fulfil the generalized Lipschitz conditions

$$\tilde{\sigma}_t^T \Lambda_1 \tilde{\sigma}_t \le \Delta_t^T \Lambda_\sigma \Delta_t, \qquad \gamma^T(u_t)\tilde{\phi}_t^T \Lambda_2 \tilde{\phi}_t \gamma(u_t) \le \Delta_t^T \Lambda_\phi \Delta_t \left\| \gamma(u_t) \right\|^2, \quad (2.6)$$

where

$$\tilde{\sigma}_t := \sigma\left(V_1{}^0 \hat{x}_t\right) - \sigma\left(V_1{}^0 x_t\right), \qquad \tilde{\phi}_t := \phi\left(V_2{}^0 \hat{x}_t\right) - \phi\left(V_2{}^0 x_t\right), \quad (2.7)$$

$\Lambda_1 \in \mathfrak{R}^{m \times m}$, $\Lambda_2 \in \mathfrak{R}^{r \times r}$, $\Lambda_\sigma \in \mathfrak{R}^{n \times n}$, $\Lambda_\phi \in \mathfrak{R}^{n \times n}$ are *known* positive definite matrices, $V_1^0 \in \mathfrak{R}^{m \times n}$ and $V_2^0 \in \mathfrak{R}^{r \times n}$ are constant matrices which can be selected by the designer.

As $\sigma(\cdot)$ and $\phi(\cdot)$ fulfil the Lipschitz conditions and from Lemma A.1 proven in [26] the following is true:

$$\tilde{\sigma}_t' := \sigma(V_{1,t}\hat{x}_t) - \sigma\left(V_1{}^0 \hat{x}_t\right) = D_\sigma \tilde{V}_{1,t}\hat{x}_t + v_\sigma,$$

$$\tilde{\sigma}_t' \gamma(u_t) := \left(\phi(V_{2,t}\hat{x}_t) - \phi\left(V_2^0 \hat{x}_t\right)\right)\gamma(u_t) \quad (2.8)$$

$$= \sum_{i=1}^s \left(\phi_i(V_{2,t}\hat{x}_t) - \phi_i\left(V_2^0\hat{x}_t\right)\right)\gamma_i(u_t) = \sum_{i=1}^s \left(D_{i\phi}\tilde{V}_{2,t}\hat{x}_t + v_{i\phi}\right)\gamma_i(u_t), \quad (2.9)$$

where

$$D_\sigma = \left.\frac{\partial \sigma(Y)}{\partial Y}\right|_{Y=V_1^0\hat{x}_t} \in \mathfrak{R}^{m \times m}, \qquad D_{i\phi} = \left.\frac{\partial \phi_i(Z)}{\partial Z}\right|_{Z=V_2^0\hat{x}_t} \in \mathfrak{R}^{r \times r}, \quad (2.10)$$

$v_\sigma \in \mathfrak{R}^m$ and $v_{i\phi} \in \mathfrak{R}^n$ are unknown vectors but bounded by $\|v_\sigma\|_{\Lambda_1}^2 \le l_1\|\tilde{V}_{1,t}\hat{x}_t\|_{\Lambda_1}^2$, $\|v_{i\phi}\|_{\Lambda_2}^2 \le l_2\|\tilde{V}_{2,t}\hat{x}_t\|_{\Lambda_2}^2$, respectively; $\tilde{V}_{1,t} := V_{1,t} - V_1^0$, $\tilde{V}_{2,t} := V_{2,t} - V_2^0$, l_1 and l_2 are

positive constants which can be defined as $l_1 := 4L_{g,1}^2$, $l_2 := 4L_{g,2}^2$, where $L_{g,1}$ and $L_{g,2}$ are global Lipschitz constants for $\sigma(\cdot)$ and $\phi_i(\cdot)$, respectively.

(A3) The nonlinear function $\gamma(\cdot)$ is such that $\|\gamma(u_t)\|^2 \leq \overline{u}$ where $\overline{u}$ is a known positive constant.

(A4) Unmodeled dynamics $\widetilde{f}_t$ is bounded by

$$\left\|\widetilde{f}_t\right\|_{\Lambda_3}^2 \leq \widetilde{f}_0 + \widetilde{f}_1 \|x_t\|_{\Lambda_3}^2, \tag{2.11}$$

where $\widetilde{f}_0$ and $\widetilde{f}_1$ are known positive constants and $\Lambda_3 \in \Re^{n \times n}$ is a known positive definite matrix and $\widetilde{f}_t$ can be defined as $\widetilde{f}_t := f(x_t, u_t, t) - Ax_t - W_1^0 \sigma(V_1^0 x_t) - W_2^0 \phi(V_2^0 x_t) \gamma(u_t)$; $W_1^0 \in \Re^{n \times m}$ and $W_2^0 \in \Re^{n \times r}$ are constant matrices which can be selected by the designer.

(A5) The deterministic disturbance is bounded, that is, $\|\xi_t\|_{\Lambda_4}^2 \leq \Upsilon$, Λ_4 is a known positive definite matrix.

(A6) The following matrix Riccati equation has a unique, positive definite solution P:

$$A^T P + PA + PRP + Q = 0, \tag{2.12}$$

where

$$R = 2W_1^0 \Lambda_1^{-1} \left(W_1^0\right)^T + 2W_2^0 \Lambda_2^{-1} \left(W_2^0\right)^T + \Lambda_3^{-1} + \Lambda_4^{-1}, \quad Q = \Lambda_\sigma + \overline{u}\Lambda_\phi + Q_0, \tag{2.13}$$

Q_0 is a positive definite matrix which can be selected by the designer.

Remark 2.1. Based on [30, 31], it can be established that the matrix Riccati equation (2.12) has a unique positive definite solution P if the following conditions are satisfied;

(a) The pair $(A, R^{1/2})$ is controllable, and the pair $(Q^{1/2}, A)$ is observable.

(b) The following matrix inequality is fulfilled:

$$\frac{1}{4}\left(A^T R^{-1} - R^{-1} A\right) R \left(A^T R^{-1} - R^{-1} A\right)^T \leq A^T R^{-1} A - Q. \tag{2.14}$$

Both conditions can relatively easily be fulfilled if A is selected as a stable diagonal matrix.

(A7) It exists a bounded control u_t, such that the closed-loop system is quadratic stable, that is, it exists a Lyapunov function $V^0 > 0$ and a positive constant λ such that

$$\frac{\partial V^0}{\partial x} f(x_t, u_t, t) \leq -\lambda \|x_t\|^2. \tag{2.15}$$

Additionally, the inequality $\lambda \geq \widetilde{f}_1 \|\Lambda_3\|$ must be satisfied.

Now, consider the learning law:

$$\dot{W}_{1,t} = -s_t K_1 P \Delta_t \sigma^T (V_{1,t} \hat{x}_t) + s_t K_1 P \Delta_t \hat{x}_t^T \tilde{V}_{1,t}^T D_\sigma^T,$$

$$\dot{W}_{2,t} = -s_t K_2 P \Delta_t \gamma^T (u_t) \phi^T (V_{2,t} \hat{x}_t) + s_t K_2 P \Delta_t \hat{x}_t^T \tilde{V}_{2,t}^T \sum_{i=1}^{s} \left(\gamma_i (u_t) D_{i\phi}^T \right),$$

$$\dot{V}_{1,t} = -s_t K_3 D_\sigma^T W_{1,t}^T P \Delta_t \hat{x}_t^T - s_t \frac{l_1}{2} K_3 \Lambda_1 \tilde{V}_{1,t} \hat{x}_t \hat{x}_t^T,$$

$$\dot{V}_{2,t} = -s_t K_4 \sum_{i=1}^{s} \left(\gamma_i (u_t) D_{i\phi}^T \right) W_{2,t}^T P \Delta_t \hat{x}_t^T - s_t \frac{s l_2 \overline{u}}{2} K_4 \Lambda_2 \tilde{V}_{2,t} \hat{x}_t \hat{x}_t^T, \tag{2.16}$$

where s is the number of columns corresponding to $\phi(\cdot)$, $K_1 \in \mathfrak{R}^{n \times n}$, $K_2 \in \mathfrak{R}^{n \times n}$, $K_3 \in \mathfrak{R}^{m \times m}$, and $K_4 \in \mathfrak{R}^{r \times r}$ are positive definite matrices which are selected by the designer. s_t is a dead-zone function which is defined as

$$s_t := \left[1 - \frac{\mu}{\|P^{1/2} \Delta_t\|} \right]_+, \qquad [z]_+ = \begin{cases} z & z \geq 0, \\ 0 & z < 0, \end{cases} \tag{2.17}$$

$$\mu = \frac{\tilde{f}_0 + \Upsilon}{\lambda_{\min} (P^{-1/2} Q_0 P^{-1/2})}.$$

Based on this learning law, the following result was demonstrated in [26].

Theorem 2.2. *If the assumptions (A1)–(A7) are satisfied and the weight matrices $W_{1,t}$, $W_{2,t} V_{1,t}$, and $V_{2,t}$ of the neural network (2.2) are adjusted by the learning law (2.16), then*

(a) *the identification error and the weights are bounded:*

$$\Delta_t, W_{1,t}, W_{2,t}, V_{1,t}, V_{2,t} \in L_\infty, \tag{2.18}$$

(b) *the identification error Δ_t satisfies the following tracking performance:*

$$\limsup_{T \to \infty} \frac{1}{T} \int_0^T \left[1 - \frac{\tilde{f}_0 + \Upsilon}{\lambda_{\min} (P^{-1/2} Q_0 P^{-1/2}) \|P^{1/2} \Delta_t\|} \right]_+ \Delta_t^T Q_0 \Delta_t \, dt \leq \tilde{f}_0 + \Upsilon. \tag{2.19}$$

In order to prove this result, the following nonnegative function was utilized:

$$V_t := V_0 + \left[\|P^{1/2} \Delta_t\| - \mu \right]_+^2 + \text{tr} \left[\widetilde{W}_{1,t}^T K_1^{-1} \widetilde{W}_{1,t} \right] + \text{tr} \left[\widetilde{W}_{2,t}^T K_2^{-1} \widetilde{W}_{2,t} \right] + \text{tr} \left[\tilde{V}_{1,t}^T K_3^{-1} \tilde{V}_{1,t} \right] \\ + \text{tr} \left[\tilde{V}_{2,t}^T K_4^{-1} \tilde{V}_{2,t} \right], \tag{2.20}$$

where $\widetilde{W}_{1,t} := W_{1,t} - W_1^0$; $\widetilde{W}_{2,t} := W_{2,t} - W_2^0$.

3. Exponential Convergence of the Identification Process

Consider that the assumptions (A1)–(A3) and (A5)-(A6) are still valid but the assumption (A4) is slightly modified as follows.

> (B4) In a compact set $\Omega \in \Re^n$, unmodeled dynamics $\widetilde{f}_t$ is bounded by $\|\widetilde{f}_t\|^2_{\Lambda_3} \leq \widetilde{f}_0$ where $\widetilde{f}_0$ is a constant not necessarily a priori known.

Remark 3.1. (B4) is a common assumption in the neural network literature [17, 22]. As mentioned in Section 2, $\widetilde{f}_t$ is given by $\widetilde{f}_t := f(x_t, u_t, t) - Ax_t - W_1^0\sigma(V_1^0 x_t) - W_2^0\phi(V_2^0 x_t)\gamma(u_t)$. Note that $W_1^0\sigma(V_1^0 x_t)$ and $W_2^0\phi(V_2^0 x_t)\gamma(u_t)$ are bounded functions because $\sigma(\cdot)$ and $\phi(\cdot)$ are sigmoidal functions. As x_t belongs to Ω, clearly x_t is also bounded. Therefore, assumption B4 implies implicitly that $f(x_t, u_t, t)$ is a bounded function in a compact set $\Omega \in \Re^n$.

Although certainly assumption (B4) is more restrictive than assumption (A4), from now on, assumption (A7) is not needed anymore.

In this paper, the following modification to the learning law (2.16) is proposed:

$$\dot{W}_{1,t} = -2k_1 P\Delta_t \sigma^T(V_{1,t}\widehat{x}_t) + 2k_1 P\Delta_t \widehat{x}_t^T \widetilde{V}_{1,t}^T D_\sigma^T - \frac{\alpha}{2}\widetilde{W}_{1,t},$$

$$\dot{W}_{2,t} = -2k_2 P\Delta_t \gamma^T(u_t)\phi^T(V_{2,t}\widehat{x}_t) + 2k_2 P\Delta_t \widehat{x}_t^T \widetilde{V}_{2,t}^T \sum_{i=1}^{s}\left(\gamma_i(u_t)D_{i\phi}^T\right) - \frac{\alpha}{2}\widetilde{W}_{2,t},$$

$$\dot{V}_{1,t} = -2k_3 D_\sigma^T W_{1,t}^T P\Delta_t \widehat{x}_t^T - k_3 l_1 \Lambda_1 \widetilde{V}_{1,t}\widehat{x}_t\widehat{x}_t^T - \frac{\alpha}{2}\widetilde{V}_{1,t}, \tag{3.1}$$

$$\dot{V}_{2,t} = -2k_4 \sum_{i=1}^{s}\left(\gamma_i(u_t)D_{i\phi}^T\right)W_{2,t}^T P\Delta_t \widehat{x}_t^T - k_4 s l_2 \overline{u}\Lambda_2 \widetilde{V}_{2,t}\widehat{x}_t\widehat{x}_t^T - \frac{\alpha}{2}\widetilde{V}_{2,t},$$

where k_1, k_2, k_3, and k_4 are positive constants which are selected by the designer; P is the solution of the Riccati equation given by (2.12); $\alpha := \lambda_{\min}(P^{-1/2}Q_0 P^{-1/2})$; s is the number of columns corresponding to $\phi(\cdot)$. By using the constants k_1, k_2, k_3, and k_4 in (3.1) instead of the matrices K_1, K_2, K_3, and K_4 in (2.16), the tuning process of the neural network (2.2) is simplified. Besides, none dead-zone function is now required. Based on the learning law (3.1), the following result is here established.

Theorem 3.2. *If the assumptions (A1)–(A3), (B4), (A5)-(A6) are satisfied and the weight matrices $W_{1,t}$, $W_{2,t}$, $V_{1,t}$, and $V_{2,t}$ of the neural network (2.2) are adjusted by the learning law (3.1), then*

> (a) *the identification error and the weights are bounded:*

$$\Delta_t, W_{1,t}, W_{2,t}, V_{1,t}, V_{2,t} \in L_\infty, \tag{3.2}$$

> (b) *the norm of identification error converges exponentially to a region bounded given by*

$$\lim_{t\to\infty}\|x_t - \widehat{x}_t\| \leq \sqrt{\frac{\widetilde{f}_0 + \Upsilon}{\alpha\lambda_{\min}(P)}}. \tag{3.3}$$

Proof of Theorem 3.2. Before beginning analysis, the dynamics of the identification error Δ_t must be determined. The first derivative of Δ_t is

$$\frac{d\Delta_t}{dt} = \frac{d}{dt}(\hat{x}_t - x_t). \tag{3.4}$$

Note that an alternative representation for (2.1) could be calculated as follows:

$$\dot{x}_t = Ax_t + W_1^0\sigma\left(V_1^0 x_t\right) + W_2^0\phi\left(V_2^0 x_t\right)\gamma(u_t) + \tilde{f}_t + \xi_t. \tag{3.5}$$

Substituting (2.2) and (3.5) into (3.4) yields

$$
\begin{aligned}
\dot{\Delta}_t &= A\hat{x}_t + W_{1,t}\sigma(V_{1,t}\hat{x}_t) + W_{2,t}\phi(V_{2,t}\hat{x}_t)\gamma(u_t) - Ax_t - W_1^0\sigma\left(V_1^0 x_t\right) \\
&\quad - W_2^0\phi\left(V_2^0 x_t\right)\gamma(u_t) - \tilde{f}_t - \xi_t, \\
&= A\Delta_t + W_{1,t}\sigma(V_{1,t}\hat{x}_t) - W_1^0\sigma\left(V_1^0 x_t\right) + W_{2,t}\phi(V_{2,t}\hat{x}_t)\gamma(u_t) \\
&\quad - W_2^0\phi\left(V_2^0 x_t\right)\gamma(u_t) - \tilde{f}_t - \xi_t.
\end{aligned}
\tag{3.6}
$$

Subtracting and adding the terms $W_1^0\sigma(V_{1,t}\hat{x}_t)$, $W_1^0\sigma(V_1^0\hat{x}_t)$, $W_2^0\phi(V_{2,t}\hat{x}_t)\gamma(u_t)$, and $W_2^0\phi(V_2^0\hat{x}_t)\gamma(u_t)$ and considering that $\widetilde{W}_{1,t} := W_{1,t} - W_1^0$, $\widetilde{W}_{2,t} := W_{2,t} - W_2^0$, $\tilde{\sigma}_t := \sigma(V_1^0\hat{x}_t) - \sigma(V_1^0 x_t)$, $\tilde{\phi}_t := \phi(V_2^0\hat{x}_t) - \phi(V_2^0 x_t)$, $\tilde{\sigma}_t' := \sigma(V_{1,t}\hat{x}_t) - \sigma(V_1^0\hat{x}_t)$, $\tilde{\phi}_t'\gamma(u_t) := (\phi(V_{2,t}\hat{x}_t) - \phi(V_2^0\hat{x}_t))\gamma(u_t)$, (3.6) can be expressed as

$$
\begin{aligned}
\dot{\Delta}_t &= A\Delta_t + W_{1,t}\sigma(V_{1,t}\hat{x}_t) - W_1^0\sigma(V_{1,t}\hat{x}_t) + W_1^0\sigma(V_{1,t}\hat{x}_t) - W_1^0\sigma\left(V_1^0\hat{x}_t\right) + W_1^0\sigma\left(V_1^0\hat{x}_t\right) \\
&\quad - W_1^0\sigma\left(V_1^0 x_t\right) + W_{2,t}\phi(V_{2,t}\hat{x}_t)\gamma(u_t) - W_2^0\phi(V_{2,t}\hat{x}_t)\gamma(u_t) + W_2^0\phi(V_{2,t}\hat{x}_t)\gamma(u_t) \\
&\quad - W_2^0\phi\left(V_2^0\hat{x}_t\right)\gamma(u_t) + W_2^0\phi\left(V_2^0\hat{x}_t\right)\gamma(u_t) - W_2^0\phi\left(V_2^0 x_t\right)\gamma(u_t) - \tilde{f}_t - \xi_t \\
&= A\Delta_t + \widetilde{W}_{1,t}\sigma(V_{1,t}\hat{x}_t) + W_1^0\tilde{\sigma}_t' + W_1^0\tilde{\sigma}_t + \widetilde{W}_{2,t}\phi(V_{2,t}\hat{x}_t)\gamma(u_t) + W_2^0\tilde{\phi}_t'\gamma(u_t) \\
&\quad + W_2^0\tilde{\phi}_t\gamma(u_t) - \tilde{f}_t - \xi_t, \\
\dot{\Delta}_t &= A\Delta_t + \widetilde{W}_{1,t}\sigma(V_{1,t}\hat{x}_t) + \widetilde{W}_{2,t}\phi(V_{2,t}\hat{x}_t)\gamma(u_t) + W_1^0\tilde{\sigma}_t + W_2^0\tilde{\phi}_t\gamma(u_t) + W_1^0\tilde{\sigma}_t' \\
&\quad + W_2^0\tilde{\phi}_t'\gamma(u_t) - \tilde{f}_t - \xi_t.
\end{aligned}
\tag{3.7}
$$

In order to begin analysis, the following nonnegative function is selected:

$$
\begin{aligned}
V_t &= \Delta_t^T P\Delta_t + \frac{1}{2k_1}\operatorname{tr}\left\{\widetilde{W}_{1,t}^T\widetilde{W}_{1,t}\right\} + \frac{1}{2k_2}\operatorname{tr}\left\{\widetilde{W}_{2,t}^T\widetilde{W}_{2,t}\right\} \\
&\quad + \frac{1}{2k_3}\operatorname{tr}\left\{\widetilde{V}_{1,t}^T\widetilde{V}_{1,t}\right\} + \frac{1}{2k_4}\operatorname{tr}\left\{\widetilde{V}_{2,t}^T\widetilde{V}_{2,t}\right\},
\end{aligned}
\tag{3.8}
$$

where P is a positive solution for the Riccati matrix equation given by (2.12). The first derivative of V_t is

$$\dot{V}_t = \frac{d}{dt}\left(\Delta_t^T P \Delta_t\right) + \frac{d}{dt}\left(\frac{1}{2k_1}\operatorname{tr}\left\{\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\right\}\right) + \frac{d}{dt}\left(\frac{1}{2k_2}\operatorname{tr}\left\{\widetilde{W}_{2,t}^T \widetilde{W}_{2,t}\right\}\right)$$

$$+ \frac{d}{dt}\left(\frac{1}{2k_3}\operatorname{tr}\left\{\widetilde{V}_{1,t}^T \widetilde{V}_{1,t}\right\}\right) + \frac{d}{dt}\left(\frac{1}{2k_4}\operatorname{tr}\left\{\widetilde{V}_{2,t}^T \widetilde{V}_{2,t}\right\}\right). \tag{3.9}$$

Each term of (3.9) will be calculated separately. For $(d/dt)(\Delta_t^T P \Delta_t)$,

$$\frac{d}{dt}\left(\Delta_t^T P \Delta_t\right) = 2\Delta_t^T P \dot{\Delta}_t. \tag{3.10}$$

substituting (3.7) into (3.10) yields

$$\frac{d}{dt}\left(\Delta_t^T P \Delta_t\right) = 2\Delta_t^T P A \Delta_t + 2\Delta_t^T P \widetilde{W}_{1,t}\sigma(V_{1,t}\widehat{x}_t) + 2\Delta_t^T P \widetilde{W}_{2,t}\phi(V_{2,t}\widehat{x}_t)\gamma(u_t)$$

$$+ 2\Delta_t^T P W_1^0 \widetilde{\sigma}_t + 2\Delta_t^T P W_2^0 \widetilde{\phi}_t \gamma(u_t) + 2\Delta_t^T P W_1^0 \widetilde{\sigma}_t' + 2\Delta_t^T P W_2^0 \widetilde{\phi}_t' \gamma(u_t) \tag{3.11}$$

$$- 2\Delta_t^T P \widetilde{f}_t - 2\Delta_t^T P \xi_t.$$

The terms $2\Delta_t^T P W_1^0 \widetilde{\sigma}_t$, $2\Delta_t^T P W_2^0 \widetilde{\phi}_t \gamma(u_t)$, $-2\Delta_t^T P \widetilde{f}_t$, and $-2\Delta_t^T P \xi_t$ in (3.11) can be bounded using the following matrix inequality proven in [26]:

$$X^T Y + Y^T X \leq X^T \Gamma^{-1} X + Y^T \Gamma Y, \tag{3.12}$$

which is valid for any $X, Y \in \Re^{n \times k}$ and for any positive definite matrix $0 < \Gamma = \Gamma^T \in \Re^{n \times n}$. Thus, for $2\Delta_t^T P W_1^0 \widetilde{\sigma}_t$ and considering assumption (A2),

$$2\Delta_t^T P W_1^0 \widetilde{\sigma}_t = \Delta_t^T P W_1^0 \widetilde{\sigma}_t + \widetilde{\sigma}_t^T \left(W_1^0\right)^T P \Delta_t$$

$$\leq \Delta_t^T P W_1^0 \Lambda_1^{-1}\left(W_1^0\right)^T P \Delta_t + \widetilde{\sigma}_t^T \Lambda_1 \widetilde{\sigma}_t \tag{3.13}$$

$$\leq \Delta_t^T P W_1^0 \Lambda_1^{-1}\left(W_1^0\right)^T P \Delta_t + \Delta_t^T \Lambda_\sigma \Delta_t.$$

For $2\Delta_t^T P W_2^0 \widetilde{\phi}_t \gamma(u_t)$, and considering assumptions (A2) and (A3)

$$2\Delta_t^T P W_2^0 \widetilde{\phi}_t \gamma(u_t) = \Delta_t^T P W_2^0 \widetilde{\phi}_t \gamma(u_t) + \gamma^T(u_t)\widetilde{\phi}_t^T \left(W_2^0\right)^T P \Delta_t$$

$$\leq \Delta_t^T P W_2^0 \Lambda_2^{-1}\left(W_2^0\right)^T P \Delta_t + \gamma^T(u_t)\widetilde{\phi}_t^T \Lambda_2 \widetilde{\phi}_t \gamma(u_t) \tag{3.14}$$

$$\leq \Delta_t^T P W_2^0 \Lambda_2^{-1}\left(W_2^0\right)^T P \Delta_t + \overline{u}\Delta_t^T \Lambda_\phi \Delta_t.$$

By using (3.12) and given assumptions (B4) and (A5), $-2\Delta_t^T P \widetilde{f}_t$ and $-2\Delta_t^T P \xi_t$ can be bounded, respectively, by

$$
\begin{aligned}
-2\Delta_t^T P \widetilde{f}_t &= -\Delta_t^T P \widetilde{f}_t - \widetilde{f}_t^T P \Delta_t \le \Delta_t^T P \Lambda_3^{-1} P \Delta_t + \widetilde{f}_t^T \Lambda_3 \widetilde{f}_t \\
&\le \Delta_t^T P \Lambda_3^{-1} P \Delta_t + \widetilde{f}_0, \\
-2\Delta_t^T P \xi_t &= -\Delta_t^T P \xi_t - \xi_t^T P \Delta_t \le \Delta_t^T P \Lambda_4^{-1} P \Delta_t + \xi_t^T \Lambda_4 \xi_t \\
&\le \Delta_t^T P \Lambda_4^{-1} P \Delta_t + \Upsilon.
\end{aligned}
\tag{3.15}
$$

Considering (2.8), $2\Delta_t^T P W_1^0 \widetilde{\sigma}'_t$ can be developed as

$$
2\Delta_t^T P W_1^0 \widetilde{\sigma}'_t = 2\Delta_t^T P W_1^0 D_\sigma \widetilde{V}_{1,t} \widehat{x}_t + 2\Delta_t^T P W_1^0 \nu_\sigma.
\tag{3.16}
$$

By simultaneously adding and subtracting the term $2\Delta_t^T P W_{1,t} D_\sigma \widetilde{V}_{1,t} \widehat{x}_t$ into the right-hand side of (3.16),

$$
2\Delta_t^T P W_1^0 \widetilde{\sigma}'_t = 2\Delta_t^T P W_{1,t} D_\sigma \widetilde{V}_{1,t} \widehat{x}_t - 2\Delta_t^T P \widetilde{W}_{1,t} D_\sigma \widetilde{V}_{1,t} \widehat{x}_t + 2\Delta_t^T P W_1^0 \nu_\sigma.
\tag{3.17}
$$

By using (3.12) and considering assumption (A2), the term $\Delta_t^T P W_1^0 \nu_\sigma$ can be bounded as

$$
\begin{aligned}
2\Delta_t^T P W_1^0 \nu_\sigma &= \Delta_t^T P W_1^0 \nu_\sigma + \nu_\sigma^T \left(W_1^0\right)^T P \Delta_t \le \Delta_t^T P W_1^0 \Lambda_1^{-1} \left(W_1^0\right)^T P \Delta_t + \nu_\sigma^T \Lambda_1 \nu_\sigma \\
&\le \Delta_t^T P W_1^0 \Lambda_1^{-1} \left(W_1^0\right)^T P \Delta_t + l_1 \left\| \widetilde{V}_{1,t} \widehat{x}_t \right\|_{\Lambda_1}^2.
\end{aligned}
\tag{3.18}
$$

And consequently, $2\Delta_t^T P W_1^0 \widetilde{\sigma}'_t$ is bounded by

$$
2\Delta_t^T P W_1^0 \widetilde{\sigma}'_t \le 2\Delta_t^T P W_{1,t} D_\sigma \widetilde{V}_{1,t} \widehat{x}_t - 2\Delta_t^T P \widetilde{W}_{1,t} D_\sigma \widetilde{V}_{1,t} \widehat{x}_t + \Delta_t^T P W_1^0 \Lambda_1^{-1} \left(W_1^0\right)^T P \Delta_t + l_1 \left\| \widetilde{V}_{1,t} \widehat{x}_t \right\|_{\Lambda_1}^2.
\tag{3.19}
$$

For $2\Delta_t^T P W_2^0 \widetilde{\phi}'_t \gamma(u_t)$ and considering (2.9),

$$
\begin{aligned}
2\Delta_t^T P W_2^0 \widetilde{\phi}'_t \gamma(u_t) &= 2\Delta_t^T P W_2^0 \sum_{i=1}^s \left(D_{i\phi} \widetilde{V}_{2,t} \widehat{x}_t + \nu_{i\phi} \right) \gamma_i(u_t) \\
&= 2\Delta_t^T P W_2^0 \sum_{i=1}^s D_{i\phi} \widetilde{V}_{2,t} \widehat{x}_t \gamma_i(u_t) + 2\Delta_t^T P W_2^0 \sum_{i=1}^s \nu_{i\phi} \gamma_i(u_t).
\end{aligned}
\tag{3.20}
$$

Adding and subtracting the term $2\Delta_t^T P W_{2,t} \sum_{i=1}^{s} D_{i\phi} \tilde{V}_{2,t} \hat{x}_t \gamma_i(u_t)$ into the right-hand side of (3.20),

$$
\begin{aligned}
2\Delta_t^T P W_2^0 \tilde{\phi}'_t \gamma(u_t) = {} & 2\Delta_t^T P W_{2,t} \sum_{i=1}^{s} D_{i\phi} \tilde{V}_{2,t} \hat{x}_t \gamma_i(u_t) - 2\Delta_t^T P \widetilde{W}_{2,t} \sum_{i=1}^{s} D_{i\phi} \tilde{V}_{2,t} \hat{x}_t \gamma_i(u_t) \\
& + 2\Delta_t^T P W_2^0 \sum_{i=1}^{s} \nu_{i\phi} \gamma_i(u_t).
\end{aligned}
\tag{3.21}
$$

By using (3.12), $2\Delta_t^T P W_2^0 \sum_{i=1}^{s} \nu_{i\phi} \gamma_i(u_t)$ can be bounded by

$$
\begin{aligned}
2\Delta_t^T P W_2^0 \sum_{i=1}^{s} \nu_{i\phi} \gamma_i(u_t) = {} & \Delta_t^T P W_2^0 \sum_{i=1}^{s} \nu_{i\phi} \gamma_i(u_t) + \left(\sum_{i=1}^{s} \nu_{i\phi} \gamma_i(u_t) \right)^T \left(W_2^0 \right)^T P \Delta_t \\
\leq {} & \Delta_t^T P W_2^0 \Lambda_2^{-1} \left(W_2^0 \right)^T P \Delta_t + \left(\sum_{i=1}^{s} \nu_{i\phi} \gamma_i(u_t) \right)^T \Lambda_2 \sum_{i=1}^{s} \nu_{i\phi} \gamma_i(u_t),
\end{aligned}
\tag{3.22}
$$

but considering that

$$
\left(\sum_{i=1}^{s} \nu_{i\phi} \gamma_i(u_t) \right)^T \Lambda_2 \sum_{i=1}^{s} \nu_{i\phi} \gamma_i(u_t) \leq s \sum_{i=1}^{s} \gamma_i^2(u_t) \nu_{i\phi}^T \Lambda_2 \nu_{i\phi}
\tag{3.23}
$$

and from assumptions (A2) and (A3), the following can be concluded:

$$
2\Delta_t^T P W_2^0 \sum_{i=1}^{s} \nu_{i\phi} \gamma_i(u_t) \leq \Delta_t^T P W_2^0 \Lambda_2^{-1} \left(W_2^0 \right)^T P \Delta_t + s l_2 \bar{u} \left\| \tilde{V}_{2,t} \hat{x}_t \right\|_{\Lambda_2}^2.
\tag{3.24}
$$

Thus, $2\Delta_t^T P W_2^0 \tilde{\phi}'_t \gamma(u_t)$ is bounded by

$$
\begin{aligned}
2\Delta_t^T P W_2^0 \tilde{\phi}'_t \gamma(u_t) \leq {} & 2\Delta_t^T P W_{2,t} \sum_{i=1}^{s} D_{i\phi} \tilde{V}_{2,t} \hat{x}_t \gamma_i(u_t) - 2\Delta_t^T P \widetilde{W}_{2,t} \sum_{i=1}^{s} D_{i\phi} \tilde{V}_{2,t} \hat{x}_t \gamma_i(u_t) \\
& + \Delta_t^T P W_2^0 \Lambda_2^{-1} \left(W_2^0 \right)^T P \Delta_t + s l_2 \bar{u} \left\| \tilde{V}_{2,t} \hat{x}_t \right\|_{\Lambda_2}^2.
\end{aligned}
\tag{3.25}
$$

Consequently, given (3.13), (3.14), (3.15), (3.19), and (3.25), $(d/dt)(\Delta_t^T P \Delta_t)$ can be bounded as

$$
\begin{aligned}
\frac{d}{dt}\left(\Delta_t^T P \Delta_t\right) \leq & \Delta_t^T \left(A^T P + P A\right)\Delta_t + 2\Delta_t^T P \widetilde{W}_{1,t}\sigma(V_{1,t}\hat{x}_t) + 2\Delta_t^T P \widetilde{W}_{2,t}\phi(V_{2,t}\hat{x}_t)\gamma(u_t) \\
& + \Delta_t^T P W_1^0 \Lambda_1^{-1}\left(W_1^0\right)^T P \Delta_t + \Delta_t^T \Lambda_\sigma \Delta_t + \Delta_t^T P W_2^0 \Lambda_2^{-1}\left(W_2^0\right)^T P \Delta_t \\
& + \overline{u}\Delta_t^T \Lambda_\phi \Delta_t + \Delta_t^T P \Lambda_3^{-1} P \Delta_t + \widetilde{f}_0 + \Delta_t^T P \Lambda_4^{-1} P \Delta_t + \Upsilon \\
& + 2\Delta_t^T P W_{1,t} D_\sigma \widetilde{V}_{1,t}\hat{x}_t - 2\Delta_t^T P \widetilde{W}_{1,t} D_\sigma \widetilde{V}_{1,t}\hat{x}_t + \Delta_t^T P W_1^0 \Lambda_1^{-1}\left(W_1^0\right)^T P \Delta_t \\
& + l_1 \left\|\widetilde{V}_{1,t}\hat{x}_t\right\|_{\Lambda_1}^2 + 2\Delta_t^T P W_{2,t}\sum_{i=1}^{s} D_{i\phi}\widetilde{V}_{2,t}\hat{x}_t\gamma_i(u_t) \\
& - 2\Delta_t^T P \widetilde{W}_{2,t}\sum_{i=1}^{s} D_{i\phi}\widetilde{V}_{2,t}\hat{x}_t\gamma_i(u_t) + \Delta_t^T P W_2^0 \Lambda_2^{-1}\left(W_2^0\right)^T P \Delta_t \\
& + s l_2 \overline{u}\left\|\widetilde{V}_{2,t}\hat{x}_t\right\|_{\Lambda_2}^2.
\end{aligned}
\tag{3.26}
$$

With respect to $(d/dt)((1/2k_1)\operatorname{tr}\{\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\})$, using several properties of the trace of a matrix,

$$
\begin{aligned}
\frac{d}{dt}\left(\frac{1}{2k_1}\operatorname{tr}\left\{\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\right\}\right) &= \frac{1}{2k_1}\operatorname{tr}\left\{\frac{d}{dt}\left(\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\right)\right\} = \frac{1}{2k_1}\operatorname{tr}\left\{\dot{\widetilde{W}}_{1,t}^T \widetilde{W}_{1,t} + \widetilde{W}_{1,t}^T \dot{\widetilde{W}}_{1,t}\right\} \\
&= \frac{1}{2k_1}\left(\operatorname{tr}\left\{\dot{\widetilde{W}}_{1,t}^T \widetilde{W}_{1,t}\right\} + \operatorname{tr}\left\{\widetilde{W}_{1,t}^T \dot{\widetilde{W}}_{1,t}\right\}\right) = \frac{1}{k_1}\operatorname{tr}\left\{\dot{\widetilde{W}}_{1,t}^T \widetilde{W}_{1,t}\right\}.
\end{aligned}
\tag{3.27}
$$

As $\widetilde{W}_{1,t} := W_{1,t} - W_1^0$, the derivative of $W_{1,t}$ is clearly $\dot{\widetilde{W}}_{1,t} = \dot{W}_{1,t}$. However, $\dot{W}_{1,t}$ is given by the learning law (3.1). Therefore, by substituting (3.1) into $\dot{\widetilde{W}}_{1,t} = \dot{W}_{1,t}$ and the corresponding expression into the right-hand side of (3.27), it is possible to obtain

$$
\begin{aligned}
\frac{d}{dt}\left(\frac{1}{2k_1}\operatorname{tr}\left\{\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\right\}\right) &= \frac{1}{k_1}\operatorname{tr}\left\{-2k_1\sigma(V_{1,t}\hat{x}_t)\Delta_t^T P \widetilde{W}_{1,t} + 2k_1 D_\sigma \widetilde{V}_{1,t}\hat{x}_t \Delta_t^T P \widetilde{W}_{1,t} - \frac{\alpha}{2}\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\right\} \\
&= -2\operatorname{tr}\left\{\sigma(V_{1,t}\hat{x}_t)\Delta_t^T P \widetilde{W}_{1,t}\right\} + 2\operatorname{tr}\left\{D_\sigma \widetilde{V}_{1,t}\hat{x}_t \Delta_t^T P \widetilde{W}_{1,t}\right\} - \frac{\alpha}{2k_1}\operatorname{tr}\left\{\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\right\} \\
&= -2\operatorname{tr}\left\{\Delta_t^T P \widetilde{W}_{1,t}\sigma(V_{1,t}\hat{x}_t)\right\} + 2\operatorname{tr}\left\{\Delta_t^T P \widetilde{W}_{1,t} D_\sigma \widetilde{V}_{1,t}\hat{x}_t\right\} - \frac{\alpha}{2k_1}\operatorname{tr}\left\{\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\right\} \\
&= -2\Delta_t^T P \widetilde{W}_{1,t}\sigma(V_{1,t}\hat{x}_t) + 2\Delta_t^T P \widetilde{W}_{1,t} D_\sigma \widetilde{V}_{1,t}\hat{x}_t - \frac{\alpha}{2k_1}\operatorname{tr}\left\{\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\right\}.
\end{aligned}
\tag{3.28}
$$

Proceeding in a similar way for

$$\frac{d}{dt}\left(\frac{1}{2k_2}\,\mathrm{tr}\left\{\widetilde{W}_{2,t}^T\widetilde{W}_{2,t}\right\}\right), \qquad \frac{d}{dt}\left(\frac{1}{2k_3}\,\mathrm{tr}\left\{\widetilde{V}_{1,t}^T\widetilde{V}_{1,t}\right\}\right), \qquad \frac{d}{dt}\left(\frac{1}{2k_4}\,\mathrm{tr}\left\{\widetilde{V}_{2,t}^T\widetilde{V}_{2,t}\right\}\right), \qquad (3.29)$$

it is possible to obtain

$$\frac{d}{dt}\left(\frac{1}{2k_2}\,\mathrm{tr}\left\{\widetilde{W}_{2,t}^T\widetilde{W}_{2,t}\right\}\right) = -2\Delta_t^T P\widetilde{W}_{2,t}\phi(V_{2,t}\widehat{x}_t)\gamma(u_t) + 2\Delta_t^T P\widetilde{W}_{2,t}\sum_{i=1}^{s}D_{i\phi}\widetilde{V}_{2,t}\widehat{x}_t\gamma_i(u_t)$$

$$-\frac{\alpha}{2k_2}tr\left\{\widetilde{W}_{2,t}^T\widetilde{W}_{2,t}\right\},$$

$$\frac{d}{dt}\left(\frac{1}{2k_3}\,\mathrm{tr}\left\{\widetilde{V}_{1,t}^T\widetilde{V}_{1,t}\right\}\right) = -2\Delta_t^T PW_{1,t}D_\sigma\widetilde{V}_{1,t}\widehat{x}_t - l_1\left\|\widetilde{V}_{1,t}\widehat{x}_t\right\|_{\Lambda_1}^2 - \frac{\alpha}{2k_3}\,\mathrm{tr}\left\{\widetilde{V}_{1,t}^T\widetilde{V}_{1,t}\right\},$$

$$\frac{d}{dt}\left(\frac{1}{2k_4}\,\mathrm{tr}\left\{\widetilde{V}_{2,t}^T\widetilde{V}_{2,t}\right\}\right) = -2\Delta_t^T PW_{2,t}\sum_{i=1}^{s}D_{i\phi}\widetilde{V}_{2,t}\widehat{x}_t\gamma_i(u_t) - sl_2\overline{u}\left\|\widetilde{V}_{2,t}\widehat{x}_t\right\|_{\Lambda_2}^2 - \frac{\alpha}{2k_4}\,\mathrm{tr}\left\{\widetilde{V}_{2,t}^T\widetilde{V}_{2,t}\right\}.$$

$$(3.30)$$

By substituting (3.26), (3.28), and (3.30) into (3.9), the following bound for $\dot{V}_t$ can be determined:

$$\dot{V}_t \le \Delta_t^T\left(A^TP + PA\right)\Delta_t + 2\Delta_t^T P\widetilde{W}_{1,t}\sigma(V_{1,t}\widehat{x}_t) + 2\Delta_t^T P\widetilde{W}_{2,t}\phi(V_{2,t}\widehat{x}_t)\gamma(u_t)$$

$$+ \Delta_t^T PW_1^0\Lambda_1^{-1}\left(W_1^0\right)^T P\Delta_t + \Delta_t^T\Lambda_\sigma\Delta_t + \Delta_t^T PW_2^0\Lambda_2^{-1}\left(W_2^0\right)^T P\Delta_t$$

$$+ \overline{u}\Delta_t^T\Lambda_\phi\Delta_t + \Delta_t^T P\Lambda_3^{-1}P\Delta_t + \widetilde{f}_0 + \widetilde{f}_1 x_t^T\Lambda_3 x_t + \Delta_t^T P\Lambda_4^{-1}P\Delta_t + \Upsilon$$

$$+ 2\Delta_t^T PW_{1,t}D_\sigma\widetilde{V}_{1,t}\widehat{x}_t - 2\Delta_t^T P\widetilde{W}_{1,t}D_\sigma\widetilde{V}_{1,t}\widehat{x}_t + \Delta_t^T PW_1^0\Lambda_1^{-1}\left(W_1^0\right)^T P\Delta_t$$

$$+ l_1\left\|\widetilde{V}_{1,t}\widehat{x}_t\right\|_{\Lambda_1}^2 + 2\Delta_t^T PW_{2,t}\sum_{i=1}^{s}D_{i\phi}\widetilde{V}_{2,t}\widehat{x}_t\gamma_i(u_t) - 2\Delta_t^T P\widetilde{W}_{2,t}\sum_{i=1}^{s}D_{i\phi}\widetilde{V}_{2,t}\widehat{x}_t\gamma_i(u_t)$$

$$+ \Delta_t^T PW_2^0\Lambda_2^{-1}\left(W_2^0\right)^T P\Delta_t + sl_2\overline{u}\left\|\widetilde{V}_{2,t}\widehat{x}_t\right\|_{\Lambda_2}^2 - 2\Delta_t^T P\widetilde{W}_{1,t}\sigma(V_{1,t}\widehat{x}_t) \qquad (3.31)$$

$$+ 2\Delta_t^T P\widetilde{W}_{1,t}D_\sigma\widetilde{V}_{1,t}\widehat{x}_t - \frac{\alpha}{2k_1}\,\mathrm{tr}\left\{\widetilde{W}_{1,t}^T\widetilde{W}_{1,t}\right\} - 2\Delta_t^T P\widetilde{W}_{2,t}\phi(V_{2,t}\widehat{x}_t)\gamma(u_t)$$

$$+ 2\Delta_t^T P\widetilde{W}_{2,t}\sum_{i=1}^{s}D_{i\phi}\widetilde{V}_{2,t}\widehat{x}_t\gamma_i(u_t) - \frac{\alpha}{2k_2}tr\left\{\widetilde{W}_{2,t}^T\widetilde{W}_{2,t}\right\} - 2\Delta_t^T PW_{1,t}D_\sigma\widetilde{V}_{1,t}\widehat{x}_t$$

$$- l_1\left\|\widetilde{V}_{1,t}\widehat{x}_t\right\|_{\Lambda_1}^2 - \frac{\alpha}{2k_3}\,\mathrm{tr}\left\{\widetilde{V}_{1,t}^T\widetilde{V}_{1,t}\right\} - 2\Delta_t^T PW_{2,t}\sum_{i=1}^{s}D_{i\phi}\widetilde{V}_{2,t}\widehat{x}_t\gamma_i(u_t)$$

$$- sl_2\overline{u}\left\|\widetilde{V}_{2,t}\widehat{x}_t\right\|_{\Lambda_2}^2 - \frac{\alpha}{2k_4}\,\mathrm{tr}\left\{\widetilde{V}_{2,t}^T\widetilde{V}_{2,t}\right\}.$$

Simplifying like terms

$$
\begin{aligned}
\dot{V}_t \leq{}& \Delta_t^T\left(A^T P + P A\right)\Delta_t + 2\Delta_t^T P W_1^0 \Lambda_1^{-1}\left(W_1^0\right)^T P \Delta_t + 2\Delta_t^T P W_2^0 \Lambda_2^{-1}\left(W_2^0\right)^T P \Delta_t \\
&+ \Delta_t^T P \Lambda_3^{-1} P \Delta_t + \Delta_t^T P \Lambda_4^{-1} P \Delta_t + \Delta_t^T \Lambda_\sigma \Delta_t + \overline{u}\Delta_t^T \Lambda_\phi \Delta_t + \tilde{f}_0 + \Upsilon \\
&- \frac{\alpha}{2k_1}\operatorname{tr}\left\{\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\right\} - \frac{\alpha}{2k_2}\operatorname{tr}\left\{\widetilde{W}_{2,t}^T \widetilde{W}_{2,t}\right\} - \frac{\alpha}{2k_3}\operatorname{tr}\left\{\widetilde{V}_{1,t}^T \widetilde{V}_{1,t}\right\} - \frac{\alpha}{2k_4}\operatorname{tr}\left\{\widetilde{V}_{2,t}^T \widetilde{V}_{2,t}\right\}.
\end{aligned}
\tag{3.32}
$$

Adding and subtracting $\Delta_t^T Q_0 \Delta_t$ into the right-hand side of the last inequality yields the expression $A^T P + P A + P R P + Q$. That is

$$
\begin{aligned}
\dot{V}_t \leq{}& -\Delta_t^T Q_0 \Delta_t + \Delta_t^T\left(A^T P + P A + P R P + Q\right)\Delta_t + \tilde{f}_0 + \Upsilon - \frac{\alpha}{2k_1}\operatorname{tr}\left\{\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\right\} \\
&- \frac{\alpha}{2k_2}\operatorname{tr}\left\{\widetilde{W}_{2,t}^T \widetilde{W}_{2,t}\right\} - \frac{\alpha}{2k_3}\operatorname{tr}\left\{\widetilde{V}_{1,t}^T \widetilde{V}_{1,t}\right\} - \frac{\alpha}{2k_4}\operatorname{tr}\left\{\widetilde{V}_{2,t}^T \widetilde{V}_{2,t}\right\}.
\end{aligned}
\tag{3.33}
$$

However, the expression $A^T P + P A + P R P + Q$ is, in accordance with the assumption (A6), equal to zero. Therefore,

$$
\begin{aligned}
\dot{V}_t \leq{}& -\Delta_t^T Q_0 \Delta_t - \frac{\alpha}{2k_1}\operatorname{tr}\left\{\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\right\} - \frac{\alpha}{2k_2}\operatorname{tr}\left\{\widetilde{W}_{2,t}^T \widetilde{W}_{2,t}\right\} - \frac{\alpha}{2k_3}\operatorname{tr}\left\{\widetilde{V}_{1,t}^T \widetilde{V}_{1,t}\right\} \\
&- \frac{\alpha}{2k_4}\operatorname{tr}\left\{\widetilde{V}_{2,t}^T \widetilde{V}_{2,t}\right\} + \tilde{f}_0 + \Upsilon.
\end{aligned}
\tag{3.34}
$$

Now, considering that $\Delta_t^T Q_0 \Delta_t = \Delta_t^T P^{1/2} P^{-1/2} Q_0 P^{-1/2} P^{1/2} \Delta_t$ and using Rayleigh's inequality, the following can be obtained:

$$
\Delta_t^T Q_0 \Delta_t \geq \lambda_{\min}\left(P^{-1/2} Q_0 P^{-1/2}\right)\Delta_t^T P \Delta_t
\tag{3.35}
$$

or alternatively

$$
-\Delta_t^T Q_0 \Delta_t \leq -\lambda_{\min}\left(P^{-1/2} Q_0 P^{-1/2}\right)\Delta_t^T P \Delta_t.
\tag{3.36}
$$

In view of (3.36), it is possible to establish that

$$
\begin{aligned}
\dot{V}_t \leq{}& -\lambda_{\min}\left(P^{-1/2} Q_0 P^{-1/2}\right)\Delta_t^T P \Delta_t - \frac{\alpha}{2k_1}\operatorname{tr}\left\{\widetilde{W}_{1,t}^T \widetilde{W}_{1,t}\right\} - \frac{\alpha}{2k_2}\operatorname{tr}\left\{\widetilde{W}_{2,t}^T \widetilde{W}_{2,t}\right\} \\
&- \frac{\alpha}{2k_3}\operatorname{tr}\left\{\widetilde{V}_{1,t}^T \widetilde{V}_{1,t}\right\} - \frac{\alpha}{2k_4}\operatorname{tr}\left\{\widetilde{V}_{2,t}^T \widetilde{V}_{2,t}\right\} + \tilde{f}_0 + \Upsilon.
\end{aligned}
\tag{3.37}
$$

As $\alpha := \lambda_{\min}(P^{-1/2} Q_0 P^{-1/2})$, finally the following bound for $\dot{V}_t$ can be concluded:

$$
\dot{V}_t \leq -\alpha V_t + \tilde{f}_0 + \Upsilon.
\tag{3.38}
$$

Equation (3.38) can be rewritten in the following form

$$\dot{V}_t + \alpha V_t \leq \tilde{f}_0 + \Upsilon. \tag{3.39}$$

Multiplying both sides of the last inequality by $\exp(\alpha t)$, it is possible to obtain

$$\exp(\alpha t)\dot{V}_t + \alpha \exp(\alpha t)V_t \leq \left(\tilde{f}_0 + \Upsilon\right)\exp(\alpha t). \tag{3.40}$$

The left-hand side of (3.40) can be rewritten as

$$\frac{d}{dt}(\exp(\alpha t)V_t) \leq \exp(\alpha t)\left(\tilde{f}_0 + \Upsilon\right) \tag{3.41}$$

or equivalently as

$$d(\exp(\alpha t)V_t) \leq \exp(\alpha t)\left(\tilde{f}_0 + \Upsilon\right)dt. \tag{3.42}$$

Integrating both sides of the last inequality yields

$$\exp(\alpha t)V_t - V_0 \leq \int_0^t \left(\tilde{f}_0 + \Upsilon\right)\exp(\alpha \tau)d\tau. \tag{3.43}$$

Adding V_0 to both sides of the inequality,

$$\exp(\alpha t)V_t \leq V_0 + \int_0^t \left(\tilde{f}_0 + \Upsilon\right)\exp(\alpha \tau)d\tau. \tag{3.44}$$

Multiplying both sides of the inequality (3.44) by $\exp(-\alpha t)$, the following can be obtained:

$$V_t \leq \exp(-\alpha t)V_0 + \exp(-\alpha t)\int_0^t \left(\tilde{f}_0 + \Upsilon\right)\exp(\alpha \tau)d\tau. \tag{3.45}$$

and, consequently,

$$V_t \leq V_0 \exp(-\alpha t) + \frac{\tilde{f}_0 + \Upsilon}{\alpha}(1 - \exp(-\alpha t)). \tag{3.46}$$

As P and Q_0 are positive definite matrices, then α is always a positive scalar and therefore V_t is an upperly bounded function. However, in reference to (3.8), V_t is also a nonnegative function. Consequently, $\Delta_t, W_{1,t}, W_{2,t}, V_{1,t}, V_{2,t} \in L_\infty$, and, thus, the first part of the Theorem 3.2 has been proven. With respect to the final part of this theorem, from (3.8), it is evident that $\Delta_t^T P \Delta_t \leq V_t$. Besides, from Rayleigh's inequality, $\lambda_{\min}(P)\Delta_t^T \Delta_t \leq \Delta_t^T P \Delta_t$.

Consequently, $\lambda_{\min}(P)\Delta_t^T\Delta_t \le V_t$. Nonetheless, in accordance with (3.46), V_t is bounded by $V_0\exp(-\alpha t) + ((\tilde{f}_0 + \Upsilon)/\alpha)(1 - \exp(-\alpha t))$. This means that

$$\|\Delta_t\| \le \sqrt{\frac{V_0\exp(-\alpha t) + \left(\left(\tilde{f}_0 + \Upsilon\right)/\alpha\right)\left(1 - \exp(-\alpha t)\right)}{\lambda_{\min}(P)}}. \tag{3.47}$$

Finally, taking the limit as $t \to \infty$ of the last inequality, the last part of Theorem 3.2 has been proven. $\qquad\square$

Remark 3.3. Based on the results presented in [32, 33], and, from the inequality (3.38), uniform stability for the identification error can be guaranteed.

Remark 3.4. Although, in [34], the asymptotic convergence of the identification error to zero is proven for multilayer neural networks, the considered class of nonlinear systems is much more restrictive than in this work.

Remark 3.5. In [35], the main idea behind Theorem 3.2 was utilized but only for the single layer case. In this paper, the generalization for the multilayer case is presented for first time.

4. Tuning of the Multilayer Identifier

In this section, some details about the selection of the parameters for the neural identifier are presented. In first place, it is important to mention that the positive definite matrices $\Lambda_1 \in \Re^{m\times m}, \Lambda_2 \in \Re^{r\times r}, \Lambda_\sigma \in \Re^{n\times n}, \Lambda_\phi \in \Re^{n\times n}, \Lambda_3 \in \Re^{n\times n}$, and $\Lambda_4 \in \Re^{n\times n}$ presented throughout assumptions (A2)–(A5) are known a priori. In fact, their selection can be very free. Although, in many cases, identity matrices can be enough, the corresponding freedom of selection can be used to satisfy the conditions specified in Remark 2.1.

Other important design decision is related to the proper number of elements m or neurons for $\sigma(\cdot)$. A good point of departure is to select $m = n$ where n is the dimension of the state vector x_t. Normally, this selection is enough in order to produce adequate results. In other case, m should be selected such as $m > n$. With respect to $\phi(\cdot)$, for simplicity, a first attempt could be to set the elements of this matrix as zeroes except for the main diagonal.

Another very important question which must be taken into account is the following: how should the weights $W_1{}^0, W_2{}^0, V_1{}^0, V_2{}^0$ be selected? Ideally, these weights should be chosen in such a way that the modelling error or unmodeled dynamics $\tilde{f}_t$ can be minimized. Likewise, the design process must consider the solution of the Riccati equation (2.12). In order to guarantee the existence of a unique positive definite solution P for (2.12), the conditions specified in Remark 2.1 must be satisfied. However, these conditions could not be fulfilled for the optimal weights. Consequently, different values for $W_1{}^0$ and $W_2{}^0$ could be tested until a solution for (2.12) can be found. At the same time, the designer should be aware of that as $W_1{}^0, W_2{}^0, V_1{}^0, V_2{}^0$ take values increasingly different from the optimal ones, the upper bound for unmodeled dynamics in assumption B4 becomes greater. With respect to the initial values for $W_{1,t}, W_{2,t}, V_{1,t}$, and $V_{2,t}$, some authors, for example [26], simply select $W_{1,0} = W_1{}^0, W_{2,0} = W_2{}^0, V_{1,0} = V_1{}^0$, and $V_{2,0} = V_2{}^0$.

Finally, a last recommendation, in order to achieve a proper performance of the neural network, the variables in the identification process should be normalized. In this context, normalization means to divide each variable by its corresponding maximum value.

5. Numerical Example

In this section, a very simple but illustrative example is presented in order to clarify the tuning process of the neural identifier and compare the advantages of the scheme developed in this paper with respect to the results of previous works [26–29].
Consider the following first-order nonlinear system:

$$\dot{x}_t = -x_t + 1.9\sin(x_t) + 2.2\cos(x_t) - 3\sin(x_t)\cos(x_t) + x_t u_t, \tag{5.1}$$

with the initial condition $x_0 = 0.7$ and the input given by $u_t = \sin(t)$.

For simplicity, ξ_t is assumed equal to zero. It is very important to note that (5.1) is only used as a data generator since apart from the assumptions (A1)–(A3), (B4), (A5)-(A6), none previous knowledge about the unknown system (5.1) is required to satisfactorily carry out the identification process.

The parameters of the neural network (2.2) and the learning laws (3.1) are selected as

$$\sigma(z) = \frac{2}{1 + e^{-2z}} - 1, \qquad \phi(y) = \frac{1}{1 + e^{-0.5y}} - 0.5,$$

$$D_\sigma = \left.\frac{\partial\sigma(z)}{\partial z}\right|_{z=V_1^0 \hat{x}_t} = \frac{4e^{-2V_1^0 \hat{x}_t}}{\left(1 + e^{-2V_1^0 \hat{x}_t}\right)^2}, \qquad D_{i\phi} = \left.\frac{\partial\phi_i(y)}{\partial y}\right|_{y=V_2^0 \hat{x}_t} = \frac{0.5e^{-0.5V_2^0 \hat{x}_t}}{\left(1 + e^{-0.5V_2^0 \hat{x}_t}\right)^2}, \tag{5.2}$$

$$A = -5, \quad \Lambda_1 = 1, \quad \Lambda_2 = 1, \quad \Lambda_3 = 1, \quad \Lambda_\sigma = 1,$$

$$\Lambda_\phi = 1, \quad \bar{u} = 1, \quad Q_0 = 2, \quad W_1^0 = 1.5, \tag{5.3}$$

$$W_2^0 = 0.5, \quad V_1^0 = 0.9, \quad V_2^0 = 1.1.$$

Note that Riccati equation (2.12) becomes a simple second-order algebraic equation for this case:

$$RP^2 + 2AP + Q = 0. \tag{5.4}$$

As $R = 2W_1^0 \Lambda_1^{-1}(W_1^0)^T + 2W_2^0 \Lambda_2^{-1}(W_2^0)^T + \Lambda_3^{-1} + \Lambda_4^{-1}$, $Q = \Lambda_\sigma + \bar{u}\Lambda_\phi + Q_0$, and given the previous values for these parameters, (5.4) has the following solution: $P = 1$. The rest of the parameters for the neural identifier are selected as $\alpha = 2$, $l_1 = 4$, $l_2 = 0.0625$, $k_1 = 500$, $k_2 = 400$, $k_3 = 600$, $k_4 = 800$. The initial condition for the neural identifier is selected as $\hat{x}_0 = 0.1$.

The results of the identification process are displayed in Figures 1 and 2. In Figure 1, the state x_t of the nonlinear system (5.1) is represented by solid line whereas the state $\hat{x}_t$ is represented by dashed line. Both states were obtained by using Simulink with the numerical method ode23s. In order to appreciate better the quality of the identification process, the
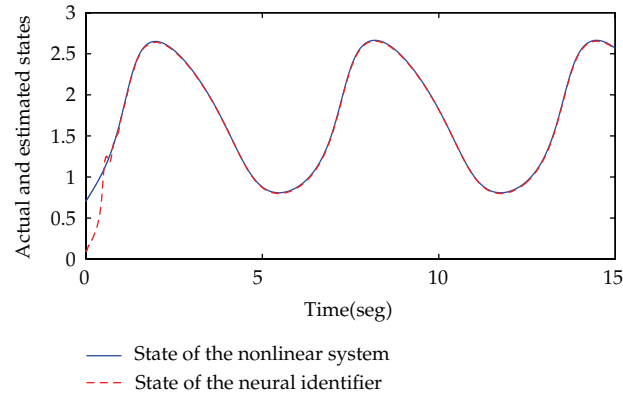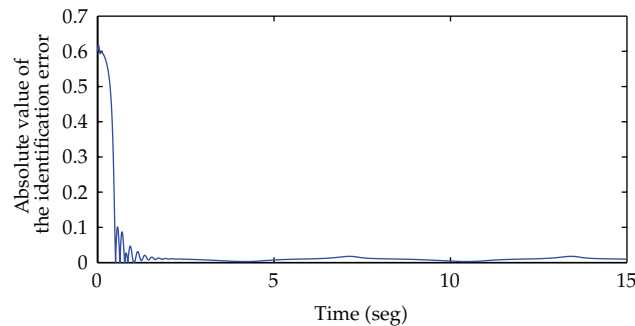
Figure 1: Identification process.

Figure 2: Identification error evolution.

absolute value of the identification error $\Delta_t := \widehat{x}_t - x_t$ is showed in Figure 2. Clearly, the new learning laws proposed in this paper exhibit a satisfactory behavior.

Now, which is the practical advantage of this method with respect to previous works [26–29]? The determination of $\widetilde{f}_0$ and $\widetilde{f}_1$ more still (parameters associated with assumption A.4) can result difficult. Besides, assuming that $\widetilde{f}_1$ is equal to zero, $\widetilde{f}_0$ can result excessively large inclusive for simple systems. For example, for system (5.1) and the values selected for the parameters of the identifier, $\widetilde{f}_0$ can approximately be estimated as 140. This implies that the learning laws (2.16) are activated only when $|\Delta_t| \geq 70$ due to the dead-zone function s_t. Thus, although the results presented in works [26–29] are technically right, on these conditions, that is, $\widetilde{f}_0 \approx 70$, the performance of the identifier results completely unsatisfactory from a practical point of view since the corresponding identification error is very high. To avoid this situation, it is necessary to be very careful with the selection of weights W_1^0, W_2^0, V_1^0, and V_2^0 in order to minimize the unmodeled dynamics $\widetilde{f}_t$. However, with these optimal weights, the matrix Riccati equation could have no solution. This dilemma is overcome by means of the learning laws (3.1) developed in this paper. In fact, as can be appreciated, a priory knowledge of $\widetilde{f}_0$ is not required anymore for the proper implementation of (3.1).

6. Conclusions

In this paper, a modification of a learning law for multilayer differential neural networks is proposed. By this modification, the dead-zone function is not required anymore and a stronger result is here guaranteed: the exponential convergence of the identification error norm to a bounded zone. This result is thoroughly proven. First, the dynamics of the identification error is determined. Next, a proper nonnegative function is proposed. A bound for the first derivative of such function is established. This bound is formed by the negative of the original nonnegative function multiplied by a constant parameter α plus a constant term. Thus, the convergence of the identification error to a bounded zone can be guaranteed. Apart from the theoretical importance of this result, from a practical point of view, the learning law here proposed is easier to implement and tune. A numerical example confirms the efficiency of this approach.

Acknowledgments

References

[1] G. Kerschen, K. Worden, A. F. Vakakis, and J. C. Golinval, "Past, present and future of nonlinear system identification in structural dynamics," *Mechanical Systems and Signal Processing*, vol. 20, no. 3, pp. 505–592, 2006.

[2] I. J. Umoh and T. Ogunfunmi, "An adaptive nonlinear filter for system identification," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, Article ID 859698, 2009.

[3] E. Gandino and S. Marchesiello, "Identification of a Duffing oscillator under different types of excitation," *Mathematical Problems in Engineering*, vol. 2010, Article ID 695025, 15 pages, 2010.

[4] J. Fei and Y. Yang, "System identification of MEMS vibratory gyroscope sensor," *Mathematical Problems in Engineering*, vol. 2011, Article ID 829432, 12 pages, 2011.

[5] J. Sjöberg, Q. Zhang, L. Ljung et al., "Nonlinear black-box modeling in system identification: a unified overview," *Automatica*, vol. 31, no. 12, pp. 1691–1724, 1995.

[6] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 1999.

[7] O. Nelles, *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*, Springer, Berlin, Germany, 2001.

[8] J. Sarangapani, *Neural Network Control of Nonlinear Discrete-Time Systems*, CRC Press, Taylor & Francis, Boca Raton, Fla, USA.

[9] M. Ibnkahla, "Nonlinear system identification using neural networks trained with natural gradient descent," *Eurasip Journal on Applied Signal Processing*, vol. 2003, no. 12, pp. 1229–1237, 2003.

[10] T. Korkobi, M. Djemel, and M. Chtourou, "Stability analysis of neural networks-based system identification," *Modelling and Simulation in Engineering*, vol. 2008, Article ID 343940, 8 pages, 2008.

[11] M. R. Zakerzadeh, M. Firouzi, H. Sayyaadi, and S. B. Shouraki, "Hysteresis nonlinearity identification using new Preisach model-based artificial neural network approach," *Journal of Applied Mathematics*, vol. 2011, Article ID 458768, 22 pages, 2011.

[12] J. Mubiru, "Using artificial neural networks to predict direct solar irradiation," *Advances in Artificial Neural Systems*, vol. 2011, Article ID 142054, 6 pages, 2011.

[13] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems*, vol. 2, no. 4, pp. 303–314, 1989.

[14] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[15] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Computation*, vol. 3, pp. 246–257, 1991.

[16] M. M. Polycarpou and P. A. Ioannou, "Identification and control of nonlinear systems using neural network models: design and stability analysis," Tech. Rep. 91-09-01, Dept. Elect. Eng. Syst., Univ. Southern Calif., Los Angeles, Calif, USA, 1991.

[17] G. A. Rovithakis and M. A. Christodoulou, "Adaptive control of unknown plants using dynamical neural networks," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 3, pp. 400–412, 1994.

[18] M. Norgaard, O. Ravn, N. K. Poulsen, and L. K. Hansen, *Neural Networks for Modelling and Control of Dynamic Systems*, Springer, Berlin, Germany, 2000.

[19] S. Haykin, *Neural Networks: A Comprehensive Foundation*, IEEE Press, New York, NY, USA, 1994.

[20] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 4–27, 1990.

[21] K. J. Hunt, D. Sbarbaro, R. Żbikowski, and P. J. Gawthrop, "Neural networks for control systems—a survey," *Automatica*, vol. 28, no. 6, pp. 1083–1112, 1992.

[22] E. B. Kosmatopoulos, M. M. Polycarpou, M. A. Christodoulou, and P. A. Ioannou, "High-order neural network structures for identification of dynamical systems," *IEEE Transactions on Neural Networks*, vol. 6, no. 2, pp. 422–431, 1995.

[23] W. Yu and A. Poznyak, "Indirect adaptive control via parallel dynamic neural networks," *IEE Proceedings: Control Theory and Applications*, vol. 146, no. 1, pp. 25–30, 1999.

[24] F. L. Lewis, S. Jagannathan, and A. Yesildirek, *Neural Network Control of Robot Manipulators and Nonlinear Systems*, Taylor & Francis, Boca Raton, Fla, USA, 1999.

[25] F. L. Lewis, K. Liu, and A. Yesildirek, "Neural net robot controller with guaranteed tracking performance," *IEEE Transactions on Neural Networks*, vol. 6, no. 3, pp. 703–715, 1995.

[26] A. S. Poznyak, E. N. Sanchez, and W. Yu, *Dynamic Neural Networks for Nonlinear Control: Identification, State Estimation and Trajectory Tracking*, World Scientific, River Edge, NJ, USA, 2001.

[27] X. Li and W. Yu, "Dynamic system identification via recurrent multilayer perceptrons," *Information Sciences*, vol. 147, no. 1–4, pp. 45–63, 2002.

[28] W. Yu, "Passivity analysis for dynamic multilayer neuro identifier," *IEEE Transactions on Circuits and Systems. I*, vol. 50, no. 1, pp. 173–178, 2003.

[29] W. Yu, "Multiple recurrent neural networks for stable adaptive control," *Neurocomputing*, vol. 70, no. 1–3, pp. 430–444, 2006.

[30] J. C. Willems, "Least squares stationary optimal control and the algebraic Riccati equation," *IEEE Transactions on Automatic Control*, vol. 16, pp. 621–634, 1971.

[31] R. Carmona and W. Yu, "Robust adaptive control via neural linearization and compensation," *Journal of Control Science and Engineering*, vol. 2012, Article ID 867178, 9 pages, 2012.

[32] J. de Jesús Rubio and W. Yu, "Stability analysis of nonlinear system identification via delayed neural networks," *IEEE Transactions on Circuits and Systems II*, vol. 54, no. 2, pp. 161–165, 2007.

[33] J. D. J. Rubio, P. Angelov, and J. Pacheco, "Uniformly stable backpropagation algorithm to train a feedforward neural network," *IEEE Transactions on Neural Networks*, vol. 22, no. 3, pp. 356–366, 2011.

[34] H. Dinh, S. Bhasin, and W. E. Dixon, "Dynamic neural network-based robust identification and control of a class of nonlinear systems," in *Proceedings of the 49th IEEE Conference on Decision and Control (CDC '10)*, pp. 5536–5541, Atlanta, Ga, USA, December 2010.

[35] J. Humberto Pérez-Cruz and A. Poznyak, "Control of nuclear research reactors based on a generalized Hopfield neural network," *Intelligent Automation and Soft Computing*, vol. 16, no. 1, pp. 39–60, 2010.