An uniformly stable backpropagation algorithm to train a feedforward neural network

José de Jesús Rubio, Angelov Plamen, Enrique García

***Abstract*— The neural networks are applied to many online fields, but the stability of the neural networks is not always assured and it could damage the instruments causing accidents. There is some research related with the stability of the continuous time neural networks, but there are some systems that are better described in discrete time, for example the population systems, the annual expenses in an industry, the interest earned by the loan of a bank, or the prediction of the distribution of loads received each hour in a warehouse, that is way it is important to consider the stability of the discrete time neural networks. The major contributions of this paper are as follows: 1) a theorem to assure the uniform stability of the general discrete time systems is proposed, 2) it is proven that the backpropagation algorithm with a new time varying rate is uniformly stable for online identification, the identification error converges to a small zone bounded by a uncertainty, 3) it is proven that the weights error is bounded by the initial weights error, i.e. the overfitting is not presented in the proposed algorithm, 4) the backpropagation is applied to predict the distribution of loads that a transelevator receive from a trailer and place in the deposits each hour in a warehouse, the deposits in the warehouse can be reserved in advance using the prediction results, 5) the backpropagation algorithm is compared with the recursive least square algorithm and with the Sugeno fuzzy inference system in the problem of the prediction of the distribution of loads in a warehouse, giving that the first and the second are stable and the third is unstable, and 6) the backpropagation algorithm is compared with the recursive least square algorithm and with the Kalman filter algorithm in an academic example.**

Neural networks, stability, prediction, identification, warehouse.

I. Introduction

The online neural networks can be used in many fields, including nonlinear adaptive control, fault detection, diagnostics, performance analysis of dynamic systems, pattern and image recognition, time-series, identification of nonlinear systems, intelligent agents, modeling, robotic, and mechatronic systems. The stability problem of the neural networks is important for the aforementioned online fields and the stability of the neural networks is not always assured.

There are some researchers who have worked with the stability of continuous time neural networks as are [19], [25], [28], [31], [32], [33], [36], [40], [41].

In [19], they study the approximation and the learning properties of one class of recurrent networks, known as high-order neural networks, and they apply these architectures to the identification of dynamic systems. In [25], the stability conditions of online identification are derived by Lyapunov–Krasovskii approach, which are described by linear matrix inequality. In [28], they present the sufficient conditions for the global asymptotic stability for a kind of recurrent neural network. In [31], they consider the robust stability of neural networks with multiple delays. The paper of [32] is concerned with the global robust exponential stability of a class of interval Cohen–Grossberg neural networks with both multiple time-varying delays and continuously distributed delays. In [33], the static neural network model and a local field neural network model are theoretically compared in terms of their trajectory transformation property, equilibrium correspondence property, nontrivial attractive manifold property, global convergence as well as stability in many different senses. In [36], dynamic multilayer neural networks are used for nonlinear system on-line identification and the passivity approach is applied to access several stability properties of the neuro-identifier. In [40], the passivity-based approach is used to derive stability conditions for dynamic neural networks with different time scales. In [41], the Lyapunov function approach is used to rigorously analyze the convergence of weights, with the use of the backpropagation algorithm, toward minima of the error function. All the works are interesting, but all consider the continuous time neural networks and there are some systems that are better described in discrete time, for example the population systems of some kind of animals [27], or the annual expenses in an industry [5], or the interest earned by the loan of a bank [5], or the prediction of the distribution of loads received each hour in a warehouse, that is way it is important to consider the stability of the discrete time neural networks.

There are some researchers who have worked with the stability of discrete time neural networks as are [24], [29], [35], [37], [38], [39].

In [24], a double dead-zone is used to assure the stability of the identification error in the gradient descent algorithm. In [29], they derive a condition for robust local stability of the multilayer recurrent neural networks. In [35], an input to state stability approach is used to create robust training algorithms for discrete time neural networks. The paper of [37] suggests new learning laws for Mamdani and Takagi–Sugeno–Kang type fuzzy neural networks based on input-to-state stability approach. In [38], the input-to-state stability approach is applied to access robust training algorithms of discrete-time recurrent neural networks. In [39], they modify the backpropagation approach and they employ a time varying rate that is determined from the input output data and the

Rubio and Garcia are with the Instituto Politécnico Nacional - ESIME Azcapotzalco, Sección de Estudios de Posgrado e Investigación, Av. de las Granjas no. 682, Col. Santa Catarina, Del. Azcapotzalco, México D.F., 02550, México, (phone:(+52)55-57296000. Ext.64497; email: jrubioa@ipn.mx)

Angelov is with the Intelligent Systems Research Laboratory, Department of Communication Systems, InfoLab21, South Drive, Lancaster University, Lancaster, LA1 4WA, UK;

model structure and stable learning algorithms for the premise and the consequence parts of the fuzzy rules are proposed. All the works propose new neural network algorithms as [29] or they modify the general backpropagation employing a time varying rate to prove the input-to-state stability as [24], [35], [37], [38], [39], in this paper it is proven that the backpropagation algorithm with a new time varying rate is uniformly stable.

On the other hand, there is some research related with the warehouses as is: [6], [8], [2], [22], [34].

The authors in [6] propose a method for selecting and materializing views, which selects and horizontally fragments a view, recomputes the size of the stored partitioned view while deciding further views to select. In [8], they consider a matrix-based discrete event control approach for a warehouse, the control system is organized in two modules: a dynamic model and a controller. In [2], they focus on the technical challenges of designing and implementing an effective data warehouse for health care information. In [22], they propose, as an extension to the data warehouse model, a knowledge warehouse architecture that will not only facilitate the capturing and coding of knowledge but also enhance the retrieval and sharing of knowledge across the organization. In [34] they propose a new constrained evolutionary algorithm for the maintenance-cost view-selection problem. All the works are interesting, but none uses the neural networks for the prediction of the distribution of loads in a warehouse, in [22], they only mention that it could be made.

In this paper, it is proposed a theorem to assure the uniform stability of the discrete time systems, it is proven that the backpropagation algorithm with a new time varying rate is uniformly stable for online identification, the identification error converges to a small zone bounded by the uncertainty, and the weights error is bounded by the initial weights error; the backpropagation is applied to predict the distribution of loads that a transelevator receive from a trailer and place in the deposits each hour in a warehouse, the deposits in the warehouse can be reserved in advance using the prediction results, the backpropagation algorithm is compared with the recursive least square algorithm and with the Sugeno fuzzy inference system in the problem of the prediction of the distribution of loads inside a warehouse, and the backpropagation algorithm is compared with the recursive least square and with the Kalman filter in an academic example.

This paper is organized as follows. In section II, the theorem that proves the uniformly stability of the discrete time systems is presented. In section III, the general backpropagation to train a feedforward neural network with a hidden layer is presented. In section IV, the uniformly stability of the backpropagation algorithm is proven. In section V, the application of the proposed algorithm is described. In section VI, a brief description of the warehouse is presented. In section VII, the backpropagation algorithm is compared with the recursive least square algorithm, with the Sugeno fuzzy inference system, and with the Kalman filter algorithm in the problem of the prediction of the distribution of loads in a warehouse and in an academic example. Finally, in section VIII, the results and the possible future research are explained.

II. Preliminaries

Let us consider the following discrete-time nonlinear system:

$$x_{k+1} = f\left[x_k, u_k\right] \tag{1}$$

Where $u_k \in \Re^m$ is the input vector, $x_k \in \Re^n$ is the state vector, u_k and x_k are known. f is an unknown nonlinear smooth function $f \in C^\infty$.

Definition 1: The system(1) is said to be uniformly stable if $\forall \epsilon > 0$, $\exists\ \delta = \delta(\epsilon)$ such that:

$$\|x_{k1}\| < \delta \Rightarrow \|x_k\| < \epsilon, \quad \forall k > k_1 \tag{2}$$

If the system has $\delta = \delta(\epsilon, k)$, then the system (1) only is stable.

Now, a basic stability theorem for discrete-time nonlinear systems is given, it is an analogous version of the continuous-time version given by [3] and of the delayed-continuous-time version given by [24].

Theorem 1: Let $L_k(x(k))$ be a Lyapunov function of the discrete-time nonlinear system (1), if it satisfies:

$$\begin{array}{c}\gamma_1\left(\|x_k\|\right) \leq L_k(x_k) \leq \gamma_2\left(\|x_k\|\right) \\ \Delta L_k(x_k) \leq -\gamma_3\left(\|x_k\|\right) + \gamma_3\left(\delta\right)\end{array} \tag{3}$$

Where δ is a positive constant, $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$ are K_∞ functions, and $\gamma_3(\cdot)$ is a K function, then the system (1) is uniformly stable.

Proof: First, let us define $\gamma_i : [0,\infty) \to [0,\infty)$, $\|x(k)\| \to y = \gamma_i\left(\|x_k\|\right)$, $i = 1, 2$. So $\forall y \in [0,\infty)$, $\exists\, \|x(k)\| \in [0,\infty)$ such that $y = \gamma_i\left(\|x_k\|\right)$, $\exists \gamma_i^{-1}$ such that $\gamma_i^{-1}\left(\gamma_i\left(\|x_k\|\right)\right) = \|x_k\|$. Two cases are discussed: 1) if $\|x_k\| \geq \delta$, then using the second equation of (3) $\Delta L_k(x_k) \leq 0$ and the system is uniformly stable, 2) if $\|x_k\| < \delta$, the definition of uniform stability (2) is used to prove that the system is uniformly stable. Let us define δ as $\delta = \left(\gamma_2^{-1} \circ \gamma_1\right)(\epsilon) = \gamma_2^{-1}\left[\gamma_1(\epsilon)\right]$. For contradiction, let us suppose that $\exists\ k_3 > k_1$ such that:

$$\|x_{k3}\| > \epsilon \tag{4}$$

Then, $\exists\ k_2 \epsilon \left[k_1, k_3\right)$ such that:

$$\|x_{k2}\| = \delta \tag{5}$$

And $\|x_k\| \geq \delta$, $\forall k \in [k_2, k_3]$, then $\gamma_3\left(\|x_k\|\right) \geq \gamma_3(\delta)$, $\forall k \in [k_2, k_3]$, it gives:

$$-\gamma_3\left(\|x_k\|\right) \leq -\gamma_3\left(\delta\right), \qquad \forall k \in [k_2, k_3] \tag{6}$$

Because γ_3 is non-decreasing, using $k_3 = k$ in the first inequality of (3) $\gamma_1\left(\|x_{k3}\|\right) \leq L_k(x_{k3}) = L_k(x_{k2}) + \Delta_{k_2}^{k_3} L_k(x_k)$. Using the first and the second inequality of (3), gives $\gamma_1\left(\|x_{k3}\|\right) \leq \gamma_2\left(\|x_{k2}\|\right) + \left[-\gamma_3\left(\|x_k\|\right) + \gamma_3\left(\delta\right)\right]\big|_{k_2}^{k_3}$. Using (5) and (6) gives $\gamma_1\left(\|x_{k3}\|\right) \leq \gamma_2(\delta)$, or $\|x_{k3}\| \leq \left(\gamma_1^{-1} \circ \gamma_2\right)(\delta)$. From the definition of δ gives:

$$\|x_{k3}\| \leq \epsilon \tag{7}$$

Where (7) contradicts (4), thus (2) is satisfied and the system is uniformly stable. If $\delta > \epsilon$, then $\exists\ k_3 > k_1$ such that (4) is satisfied, and from this inequality it gives a contradiction, that is:

$$\delta \leq \epsilon, \quad \forall k > k_1 \tag{8}$$

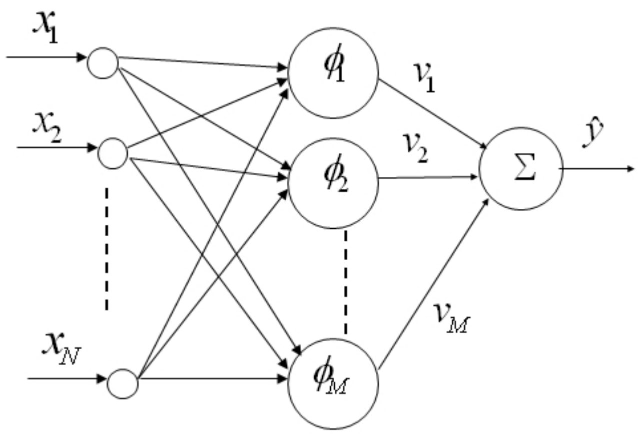


Fig. 1. Architecture of the neural network

Finally, from (2) and the definition of $0 \leq \|x_k\| < \epsilon$ gives:

$$\epsilon > 0 \tag{9}$$

(8) and (9) proves that ϵ is well defined. ■

III. The backpropagation algorithm to train a neural network

Let us consider the following unknown discrete-time nonlinear system:

$$y(k) = f\left[X_k\right] \tag{10}$$

Where $X_k = \left[x_1(k)\ldots, x_i(k), \ldots, x_N(k)\right]^T = \left[y(k-1), \ldots, y(k-n), u\left(k-1\right), \ldots, u\left(k-m\right)\right]^T \in \Re^{N\times 1}$ $(N = n+m)$ is the input vector, $u(k-1) \in \Re$ is the input of the plant, $y(k) \in \Re$ is the output of the plant, and f is an unknown nonlinear function, $f \in C^{\infty}$. The output of the neural network with one hidden layer can be expressed as [10], [11], [14]:

$$\begin{array}{c} \widehat{y}(k) = V_k \Phi_k = \sum\limits_{j=1}^{M} V_{jk}\phi_{jk} \\ \Phi_k = \left[\phi_{1k}, \ldots, \phi_{jk}, \ldots, \phi_{Mk}\right]^T \\ \phi_{jk} = \tanh(\sum\limits_{i=1}^{N} W_{ijk} x_i(k)) \end{array} \tag{11}$$

Where $i = 1, \ldots, N$, $j = 1, \ldots, M$, $X_k \in \Re^{N\times 1}$ is the input vector given by (10), $\widehat{y}(k) \in \Re$ is the output of the neural network, $V_k \in \Re^{1\times M}$ and $W_k \in \Re^{M\times N}$ are the weights of the output and the hidden layer of the neural network, respectively, $W_{ijk} \in \Re$, $x_i(k) \in \Re$, $\Phi_k \in \Re^{M\times 1}$, $\phi_{jk} \in \Re$, $V_{jk} \in \Re$, the Figure 1 shows the feedforward neural network.

IV. Stability of the backpropagation algorithm

The stability of the parameter learning is needed, because this algorithm works online. First, the model is linearized, and later, the stability of the proposed algorithm is analyzed.

According to the Stone Weierstrass theorem [4], the unknown nonlinear function f of (10) is approximated as:

$$\begin{array}{c} y(k) = V_* \Phi_{*k} + \in_f = \sum\limits_{j=1}^{M} V_{j*}\phi_{*jk} + \in_f \\ \Phi_{*k} = \left[\phi_{*1k}, \ldots, \phi_{*jk}, \ldots, \phi_{*Mk}\right]^T \\ \phi_{*jk} = \tanh(\sum\limits_{i=1}^{N} W_{ij*} x_i(k)) \end{array} \tag{12}$$

Where $\Phi_{*k} \in \Re^{M\times 1}$, $\in_f = y(k) - V_* \Phi_{*k} \in \Re$ is the modeling error, $\phi_{*jk} \in \Re$, $V_{j*} \in \Re$, $W_{ij*} \in \Re$, $V_{j*} \in \Re$ and $W_{ij*} \in \Re$ are the optimal parameters that can minimize the modeling error $\in_f$ [19].

First, the network model is linearized, it is used to define the parameters-updating and to prove the stability of the proposed algorithm.

In the case of two independent variables, a function has a Taylor series as follows:

$$\begin{array}{c} f(\omega_1, \omega_2) = f(\omega_{1^0}, \omega_{2^0}) + (\omega_1 - \omega_{1^0})\frac{\partial f(\omega_1,\omega_2)}{\partial \omega_1} \\ + (\omega_2 - \omega_{2^0})\frac{\partial f(\omega_1,\omega_2)}{\partial \omega_2} + R_f \end{array} \tag{13}$$

Where $R_f \in \Re$ is the remainder of the Taylor series. If we let ω_1, and ω_2 correspond to $W_{ijk} \in \Re$, and $V_{jk} \in \Re$, ω_{1^0} and ω_{2^0} correspond to $W_{ij*} \in \Re$ and $V_{j*} \in \Re$, let us define $\widetilde{W}_{ijk} = W_{ijk} - W_{ij*} \in \Re$ and $\widetilde{V}_{jk} = V_{jk} - V_{j*} \in \Re$, then the Taylor series is applied to linearize (11) as follows [24], [25], [26]:

$$V_k\Phi_k = V_*\Phi_{*k} + \sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}\frac{\partial V_k\Phi_k}{\partial W_{ijk}} + \sum_{j=1}^{M}\widetilde{V}_{jk}\frac{\partial V_k\Phi_k}{\partial V_{jk}} + R_f \tag{14}$$

Where $\frac{\partial V_k\Phi_k}{\partial W_{ijk}} \in \Re$ and $\frac{\partial V_k\Phi_k}{\partial V_{jk}} \in \Re$, please note that $V_k\Phi_k = \sum\limits_{j=1}^{M} V_{jk}\phi_{jk}$ and $V_*\Phi_{*k} = \sum\limits_{j=1}^{M} V_{j*}\phi_{*jk}$. As all the parameters are scalars, the Taylor series is well applied. Considering (11) and using the chain rule [15], [24], [25], [26], [30] gives:

$$\begin{array}{c} \frac{\partial V_k\Phi_k}{\partial W_{ijk}} = V_{jk}\frac{\partial \Phi_k}{\partial W_{ijk}} = V_{jk}\frac{\partial \tanh(\sum\limits_{i=1}^{N} W_{ijk}x_i(k))}{\partial W_{ijk}} \\ = V_{jk}\text{sech}^2(\sum\limits_{i=1}^{N} W_{ijk}x_i(k))x_i(k) = \sigma_{ijk} \end{array} \tag{15}$$

Where $\sigma_{ijk} = V_{jk}\text{sech}^2(\sum\limits_{i=1}^{N} W_{ijk}x_i(k))x_i(k) \in \Re$ because $V_{jk} \in \Re$, $\text{sech}^2(\sum\limits_{i=1}^{N} W_{ijk}x_i(k)) \in \Re$ and $x_i(k) \in \Re$.

$$\frac{\partial V_k\Phi_k}{\partial V_{jk}} = \frac{\partial \sum\limits_{j=1}^{M} V_{jk}\phi_{jk}}{\partial V_{jk}} = \phi_{jk} \tag{16}$$

Where $\phi_{jk} = \tanh(\sum\limits_{i=1}^{N} W_{ijk}x_i(k)) \in \Re$. Substituting $\frac{\partial V_k\Phi_k}{\partial W_{ijk}}$

of (15), and $\frac{\partial V_k \Phi_k}{\partial V_{jk}}$ of (16) into (14) gives:

$$\begin{aligned} V_k \Phi_k = V_* \Phi_{*k} + \sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk}\sigma_{ijk} \\ + \sum_{j=1}^{M} \widetilde{V}_{jk}\phi_{jk} + R_f \end{aligned} \tag{17}$$

Let us define the identification error $e(k) \in \Re$ as follows:

$$e(k) = \widehat{y}(k) - y(k) \tag{18}$$

Where $y(k)$ and $\widehat{y}(k)$ are defined in (10) and (11), respectively. Substituting (11), (12), and (18) into (17) gives:

$$e\,(k) = \sum_{j=1}^{M} \widetilde{V}_{jk}\phi_{jk} + \sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk}\sigma_{ijk} + \mu(k) \tag{19}$$

Where $\mu(k) = R_f - \in_f$.

From (19), it is obtained that:

$$\sum_{j=1}^{M} \widetilde{V}_{jk}\phi_{jk} + \sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk}\sigma_{ijk} = e\,(k) - \mu(k) \tag{20}$$

The proposed backpropagation algorithm uses a new time varying rate as follows:

$$\begin{aligned} V_{jk+1} = V_{jk} - \alpha_k \phi_{jk} e(k) \\ W_{ijk+1} = W_{ijk} - \alpha_k \sigma_{ijk} e(k) \end{aligned} \tag{21}$$

Where the new time varying rate α_k is:

$$\alpha_k = \frac{\alpha_0}{2\left(\frac{1}{2} + \sum_{j=1}^{M}\phi_{jk}^2 + \sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2\right)}$$

Where $i = 1, \ldots, N$, $j = 1, \ldots, M$, $\sigma_{ijk} = V_{jk}\text{sech}^2(\sum_{i=1}^{N} W_{ijk}x_i(k))x_i(k) \in \Re$ is defined in (15), $\phi_{jk} = \tanh(\sum_{i=1}^{N} W_{ijk}x_i(k)) \in \Re$ is defined in (11) and used in (16), $e(k)$ is defined in (18), $0 < \alpha_0 \le 1 \in \Re$, so $0 < \alpha_k \in \Re$, it is assumed that the uncertainty is bounded [13], [18], [19], [20], [21], [24], [25], [26], [37] where $\overline{\mu}$ is the upper bound of the uncertainty $\mu(k)$, $|\mu(k)| < \overline{\mu}$.

Remark 1: Please note that $e(k) = \widehat{y}(k) - y(k) = \sum_{j=1}^{M} V_{jk}\phi_{jk} - y(k)$ used in (21) is well defined because V_{jk}, ϕ_{jk}, and $y(k)$ are known.

The following theorem gives the stability of the proposed backpropagation algorithm.

Theorem 2: The backpropagation algorithm (11), (18), and (21) applied for the identification of the nonlinear system (10) is uniformly stable and the upper bound of the average identification error $e_p^2(k)$ satisfies:

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{k=2}^{T} e_p^2(k) \le \alpha_0 \overline{\mu}^2 \tag{22}$$

Where $e_p^2(k) = \frac{\alpha_k}{2} e^2(k-1)$, $0 < \alpha_0 \le 1 \in \Re$ and $0 < \alpha_k \in \Re$ are defined in (21), $e(k)$ is defined in (18), $\overline{\mu}$ is the upper bound of the uncertainty $\mu(k)$, $|\mu(k)| < \overline{\mu}$.

Proof: Let us define the following Lyapunov function:

$$L_k = \frac{1}{2}\alpha_k e^2(k-1) + \sum_{j=1}^{M} \widetilde{V}_{jk}^2 + \sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk}^2 \tag{23}$$

Where $\widetilde{V}_{jk}$ and $\widetilde{W}_{ijk}$ are defined in (13). Then ΔL_k is:

$$\begin{aligned} \Delta L_k = \tfrac{1}{2}\alpha_k e^2(k) + \sum_{j=1}^{M} \widetilde{V}_{jk+1}^2 + \sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk+1}^2 \\ -\tfrac{1}{2}\alpha_k e^2(k-1) - \sum_{j=1}^{M} \widetilde{V}_{jk}^2 - \sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk}^2 \end{aligned} \tag{24}$$

Now, the weights error can be rewritten as follows:

$$\begin{aligned} \sum_{j=1}^{M} \widetilde{V}_{jk+1}^2 = \sum_{j=1}^{M} \widetilde{V}_{jk}^2 \\ -2\alpha_k e(k)\sum_{j=1}^{M} \widetilde{V}_{jk}\phi_{jk} + \alpha_k^2 e^2(k)\sum_{j=1}^{M}\phi_{jk}^2 \\ \sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk+1}^2 = \sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk}^2 \\ -2\alpha_k e(k)\sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk}\sigma_{ijk} + \alpha_k^2 e^2(k)\sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2 \end{aligned} \tag{25}$$

It will be proven that (25) is true. Let us consider the $\widetilde{V}_{jk}$ case, substituting $\widetilde{V}_{jk+1}$ of (21) into $\sum_{j=1}^{M} \widetilde{V}_{jk+1}^2$ gives:

$$\begin{aligned} \sum_{j=1}^{M} \widetilde{V}_{jk+1}^2 = \sum_{j=1}^{M} \left[\widetilde{V}_{jk} - \alpha_k \phi_{jk} e(k)\right]^2 \\ = \sum_{j=1}^{M} \widetilde{V}_{jk}^2 - 2\alpha_k e(k)\sum_{j=1}^{M} \widetilde{V}_{jk}\phi_{jk} + \alpha_k^2 e^2(k)\sum_{j=1}^{M}\phi_{jk}^2 \end{aligned}$$

Then for $\widetilde{V}_{jk}$ of (25) is true. Let us consider the $\widetilde{W}_{ijk}$ case, substituting $\widetilde{W}_{ijk+1}$ of (21) into $\sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk+1}^2$ gives:

$$\begin{aligned} \sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk+1}^2 = \sum_{j=1}^{M}\sum_{i=1}^{N} \left[\widetilde{W}_{ijk} - \alpha_k \sigma_{ijk} e(k)\right]^2 = \sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk}^2 \\ -2\alpha_k e(k)\sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk}\sigma_{ijk} + \alpha_k^2 e^2(k)\sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2 \end{aligned}$$

Then for $\widetilde{W}_{ijk}$ of (25) is true. Thus (25) is true. Substituting (25) into (24) gives:

$$\begin{aligned} \Delta L_k = -2\alpha_k e(k)\sum_{j=1}^{M} \widetilde{V}_{jk}\phi_{jk} + \alpha_k^2 e^2(k)\sum_{j=1}^{M}\phi_{jk}^2 \\ -2\alpha_k e(k)\sum_{j=1}^{M}\sum_{i=1}^{N} \widetilde{W}_{ijk}\sigma_{ijk} + \alpha_k^2 e^2(k)\sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2 \\ +\tfrac{1}{2}\alpha_k e^2(k) - \tfrac{1}{2}\alpha_k e^2(k-1) \end{aligned} \tag{26}$$

(26) can be rewritten as:

$$\begin{array}{l}\Delta L_k = -2\alpha_k e(k)\left[\sum\limits_{j=1}^{M}\widetilde{V}_{jk}\phi_{jk} + \sum\limits_{j=1}^{M}\sum\limits_{i=1}^{N}\widetilde{W}_{ijk}\sigma_{ijk}\right]\\ \quad +\alpha_k^2 e^2(k)\left[\sum\limits_{j=1}^{M}\phi_{jk}^2 + \sum\limits_{j=1}^{M}\sum\limits_{i=1}^{N}\sigma_{ijk}^2\right]\\ \quad +\frac{1}{2}\alpha_k e^2(k) - \frac{1}{2}\alpha_k e^2(k-1)\end{array} \tag{27}$$

Substituting (20) in the first term of the equation (27) gives:

$$\begin{array}{l}\Delta L_k = -2\alpha_k e(k)\left[e(k) - \mu(k)\right]\\ \quad +\alpha_k^2 e^2(k)\left[\sum\limits_{j=1}^{M}\phi_{jk}^2 + \sum\limits_{j=1}^{M}\sum\limits_{i=1}^{N}\sigma_{ijk}^2\right]\\ \quad +\frac{1}{2}\alpha_k e^2(k) - \frac{1}{2}\alpha_k e^2(k-1)\\ \Delta L_k = -2\alpha_k e^2(k) + 2\alpha_k e(k)\mu(k)\\ \quad +\alpha_k^2 e^2(k)\left[\sum\limits_{j=1}^{M}\phi_{jk}^2 + \sum\limits_{j=1}^{M}\sum\limits_{i=1}^{N}\sigma_{ijk}^2\right]\\ \quad +\frac{1}{2}\alpha_k e^2(k) - \frac{1}{2}\alpha_k e^2(k-1)\end{array} \tag{28}$$

Substituting α_k of (21) into the term $\alpha_k^2 e^2(k)\left[\sum\limits_{j=1}^{M}\phi_{jk}^2 + \sum\limits_{j=1}^{M}\sum\limits_{i=1}^{N}\sigma_{ijk}^2\right]$ and considering $\alpha_0 \leq 1$ gives:

$$\begin{array}{l}\alpha_k^2 e^2(k)\left[\sum\limits_{j=1}^{M}\phi_{jk}^2 + \sum\limits_{j=1}^{M}\sum\limits_{i=1}^{N}\sigma_{ijk}^2\right]\\ = \dfrac{\alpha_0\left[\sum\limits_{j=1}^{M}\phi_{jk}^2+\sum\limits_{j=1}^{M}\sum\limits_{i=1}^{N}\sigma_{ijk}^2\right]}{2\left(\frac{1}{2}+\sum\limits_{j=1}^{M}\phi_{jk}^2+\sum\limits_{j=1}^{M}\sum\limits_{i=1}^{N}\sigma_{ijk}^2\right)}\alpha_k e^2(k)\\ \leq \frac{\alpha_0}{2}\alpha_k e^2(k) \leq \frac{1}{2}\alpha_k e^2(k)\end{array} \tag{29}$$

Considering that $2\alpha_k e(k)\mu(k) \leq \alpha_k e^2(k) + \alpha_k\mu^2(k)$, and considering (29) in (28) gives:

$$\begin{array}{l}\Delta L_k \leq -2\alpha_k e^2(k) + \alpha_k e^2(k) + \alpha_k\mu^2(k)\\ +\frac{1}{2}\alpha_k e^2(k) + \frac{1}{2}\alpha_k e^2(k) - \frac{1}{2}\alpha_k e^2(k-1)\end{array}$$

$$\Delta L_k \leq -\frac{1}{2}\alpha_k e^2(k-1) + \alpha_k\mu^2(k) \tag{30}$$

From (21), $\alpha_k = \dfrac{\alpha_0}{2\left(\frac{1}{2}+\sum\limits_{j=1}^{M}\phi_{jk}^2+\sum\limits_{j=1}^{M}\sum\limits_{i=1}^{N}\sigma_{ijk}^2\right)} \leq \alpha_0$, and considering that $|\mu(k)| \leq \overline{\mu}$ in (30) gives:

$$\Delta L_k \leq -\frac{1}{2}\alpha_k e^2(k-1) + \alpha_0\overline{\mu}^2 \tag{31}$$

It is known that there exits K_∞ functions $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$ such that:

$$\begin{array}{l}\gamma_1\left(\alpha_k e^2(k-1), \widetilde{V}_{jk}^2, \widetilde{W}_{ijk}^2\right) \leq L_k\\ L_k \leq \gamma_2\left(\alpha_k e^2(k-1), \widetilde{V}_{jk}^2, \widetilde{W}_{ijk}^2\right)\end{array} \tag{32}$$

Using the fact that (32) satisfies the first condition of (3), (31) satisfies the second condition of (3), thus using the theorem of the preliminaries, it is known that the identification error of the neural network applied for the identification of a nonlinear system is uniformly stable. So L_k is bounded, i.e. the identification error is bounded. Using (31) and using $e_p^2(k)$ defined in (22) gives:

$$\Delta L_k \leq -e_p^2(k) + \alpha_0\overline{\mu}^2 \tag{33}$$

Summarizing (33) from 2 to T, gives:

$$\sum_{k=2}^{T}\left(e_p^2(k) - \alpha_0\overline{\mu}^2\right) \leq L_1 - L_T \tag{34}$$

Since $L_T > 0$ is bounded:

$$\begin{array}{l}\frac{1}{T}\sum\limits_{k=2}^{T}e_p^2(k) \leq \alpha_0\overline{\mu}^2 + \frac{1}{T}L_1\\ \limsup\limits_{T\to\infty}\frac{1}{T}\sum\limits_{k=2}^{T}e_p^2(k) \leq \alpha_0\overline{\mu}^2\end{array} \tag{35}$$

(22) is established. ■

Remark 2: There are two conditions to apply this algorithm for nonlinear systems, the first one is that the nonlinear system may have the form described by equation (10), the second one is that the uncertainty $\mu(k)$ may be bounded.

Remark 3: The value of the parameter $\overline{\mu}$ is unimportant, because this parameter is not used in the algorithm. The bound of $\mu(k)$ is needed to guarantee the stability of the algorithm, but it is not used in the backpropagation algorithm (11), (18), (21).

Remark 4: The fact that $\mu(k)$ is bounded has been used for other authors in some earlier studies as are [18], [19], [20] and [21] in continuous time systems and [13], [24], [25], [26], and [37] in discrete time systems.

The following theorem proves that the weights of the proposed backpropagation algorithm are bounded.

Theorem 3: When the average error $e_p^2(k)$ is bigger than the uncertainty $\alpha_0\overline{\mu}^2$, the weights error is bounded by the initial weights error as follows:

$$\begin{array}{l}e_p^2(k) \geq \alpha_0\overline{\mu}^2\\ \Longrightarrow \sum\limits_{j=1}^{M}\widetilde{V}_{jk+1}^2 + \sum\limits_{j=1}^{M}\sum\limits_{i=1}^{N}\widetilde{W}_{ijk+1}^2 \leq \sum\limits_{j=1}^{M}\widetilde{V}_{j1}^2 + \sum\limits_{j=1}^{M}\sum\limits_{i=1}^{N}\widetilde{W}_{ij1}^2\end{array} \tag{36}$$

Where $i = 1, \ldots, N$, $j = 1, \ldots, M$, $\widetilde{V}_{jk}$ and $\widetilde{W}_{ijk}$ is defined in (13), $\widetilde{V}_{j1}$ and $\widetilde{W}_{ij1}$ is the initial weights error, $e_p^2(k) = \frac{\alpha_k}{2}e^2(k)$, V_{jk+1}, W_{ijk+1}, $0 < \alpha_0 \leq 1 \in \Re$, and $0 < \alpha_k \in \Re$ are defined in (21), $e(k)$ is defined in (18), $\overline{\mu}$ is the upper bound of the uncertainty $\mu(k)$, $|\mu(k)| < \overline{\mu}$.

Proof: From (25), the weights are written as follows:

$$\begin{aligned}
&\sum_{j=1}^{M}\widetilde{V}_{jk+1}^{2}=\sum_{j=1}^{M}\widetilde{V}_{jk}^{2} \\
&-2\alpha_k e(k)\sum_{j=1}^{M}\widetilde{V}_{jk}\phi_{jk}+\alpha_k^2 e^2(k)\sum_{j=1}^{M}\phi_{jk}^2 \\
&\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk+1}^{2}=\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}^{2} \\
&-2\alpha_k e(k)\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}\sigma_{ijk}+\alpha_k^2 e^2(k)\sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2
\end{aligned} \tag{37}$$

Adding $\sum_{j=1}^{M}\widetilde{V}_{jk+1}^{2}$ with $\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk+1}^{2}$ of (37) gives:

$$\begin{aligned}
&\sum_{j=1}^{M}\widetilde{V}_{jk+1}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk+1}^{2}=\sum_{j=1}^{M}\widetilde{V}_{jk}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}^{2} \\
&-2\alpha_k e(k)\sum_{j=1}^{M}\widetilde{V}_{jk}\phi_{jk}+\alpha_k^2 e^2(k)\sum_{j=1}^{M}\phi_{jk}^2 \\
&-2\alpha_k e(k)\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}\sigma_{ijk}+\alpha_k^2 e^2(k)\sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2
\end{aligned} \tag{38}$$

(38) can be rewritten as:

$$\begin{aligned}
&\sum_{j=1}^{M}\widetilde{V}_{jk+1}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk+1}^{2}=\sum_{j=1}^{M}\widetilde{V}_{jk}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}^{2} \\
&-2\alpha_k e(k)\left[\sum_{j=1}^{M}\widetilde{V}_{jk}\phi_{jk}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}\sigma_{ijk}\right] \\
&+\alpha_k^2 e^2(k)\left[\sum_{j=1}^{M}\phi_{jk}^2+\sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2\right]
\end{aligned} \tag{39}$$

Substituting (20) in the second term of the equation (39) gives:

$$\begin{aligned}
&\sum_{j=1}^{M}\widetilde{V}_{jk+1}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk+1}^{2}=\sum_{j=1}^{M}\widetilde{V}_{jk}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}^{2} \\
&-2\alpha_k e(k)\left[e\left(k\right)-\mu(k)\right] \\
&+\alpha_k^2 e^2(k)\left[\sum_{j=1}^{M}\phi_{jk}^2+\sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2\right]
\end{aligned}$$

$$\begin{aligned}
&\sum_{j=1}^{M}\widetilde{V}_{jk+1}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk+1}^{2}=\sum_{j=1}^{M}\widetilde{V}_{jk}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}^{2} \\
&-2\alpha_k e^2(k)+2\alpha_k e(k)\mu(k) \\
&+\alpha_k^2 e^2(k)\left[\sum_{j=1}^{M}\phi_{jk}^2+\sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2\right]
\end{aligned} \tag{40}$$

Substituting α_k of (21) into the last term of (40) and considering $\alpha_0 \le 1$ gives:

$$\begin{aligned}
&\alpha_k^2 e^2(k)\left[\sum_{j=1}^{M}\phi_{jk}^2+\sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2\right] \\
&=\frac{\alpha_0\left[\sum_{j=1}^{M}\phi_{jk}^2+\sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2\right]}{2\left(\frac{1}{2}+\sum_{j=1}^{M}\phi_{jk}^2+\sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2\right)}\alpha_k e^2(k) \\
&\le \frac{\alpha_0}{2}\alpha_k e^2(k)\le \frac{1}{2}\alpha_k e^2(k)
\end{aligned} \tag{41}$$

Considering that $2\alpha_k e(k)\mu(k) \le \alpha_k e^2(k)+\alpha_k\mu^2(k)$, and considering (41) in (40) gives:

$$\begin{aligned}
&\sum_{j=1}^{M}\widetilde{V}_{jk+1}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk+1}^{2}\le\sum_{j=1}^{M}\widetilde{V}_{jk}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}^{2} \\
&-2\alpha_k e^2(k)+\alpha_k e^2(k)+\alpha_k\mu^2(k)+\tfrac{1}{2}\alpha_k e^2(k) \\
&\sum_{j=1}^{M}\widetilde{V}_{jk+1}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk+1}^{2} \\
&\le\sum_{j=1}^{M}\widetilde{V}_{jk}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}^{2}-\tfrac{1}{2}\alpha_k e^2(k)+\alpha_k\mu^2(k)
\end{aligned} \tag{42}$$

From (21), $\alpha_k = \frac{\alpha_0}{2\left(\frac{1}{2}+\sum_{j=1}^{M}\phi_{jk}^2+\sum_{j=1}^{M}\sum_{i=1}^{N}\sigma_{ijk}^2\right)} \le \alpha_0$ and considering that $|\mu(k)| \le \overline{\mu}$ in (42) gives:

$$\begin{aligned}
&\sum_{j=1}^{M}\widetilde{V}_{jk+1}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk+1}^{2} \\
&\le\sum_{j=1}^{M}\widetilde{V}_{jk}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}^{2}-\tfrac{1}{2}\alpha_k e^2(k)+\alpha_0\overline{\mu}^2
\end{aligned} \tag{43}$$

Considering $e_p^2(k)=\frac{\alpha_k}{2}e^2(k)$ gives:

$$\begin{aligned}
&e_p^2(k)\ge\alpha_0\overline{\mu}^2 \\
&\Rightarrow\sum_{j=1}^{M}\widetilde{V}_{jk+1}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk+1}^{2}\le\sum_{j=1}^{M}\widetilde{V}_{jk}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}^{2}
\end{aligned}$$

Considering that $e_p^2(k)\ge\alpha_0\overline{\mu}^2$ for $k\in[1,\,k]$ is true, so:

$$\begin{aligned}
&\sum_{j=1}^{M}\widetilde{V}_{jk+1}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk+1}^{2}\le\sum_{j=1}^{M}\widetilde{V}_{jk}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ijk}^{2} \\
&\le\ldots\le\sum_{j=1}^{M}\widetilde{V}_{j1}^{2}+\sum_{j=1}^{M}\sum_{i=1}^{N}\widetilde{W}_{ij1}^{2}
\end{aligned}$$

Thus (36) is true. ■

Remark 5: From the Theorem 2 the average identification error $e_p^2(k)$ of the backpropagation algorithm is bounded and from the Theorem 3 the weights error $\widetilde{V}_{jk+1}^2$ and $\widetilde{W}_{ijk+1}^2$ is bounded, i.e. the proposed backpropagation algorithm to train a feedforward neural network is uniformly stable in the presence of unknown and bounded uncertainties and the overfitting mentioned in [14] and [35] is not presented. In addition, the identification error converges to a small zone bounded by the uncertainty $\overline{\mu}$.

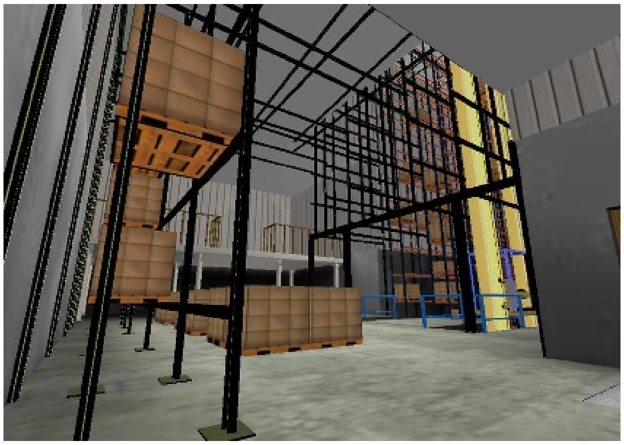


Fig. 2. The authomatic warehouse

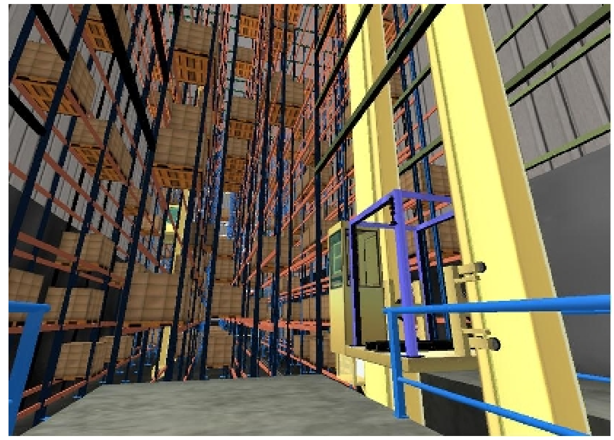


Fig. 3. The transelevator inside of the warehouse

V. THE PROPOSED ALGORITHM

The proposed algorithm is as follows:

1) Obtain the output of the nonlinear system $y(k)$ with equation (10), note that nonlinear system may have the structure with equation (10), the parameter N is selected according to this nonlinear system.

2) Select the following parameters: V_1 and W_1 are selected as random number between 0 and 1. M is selected as an entire number and α_0 is selected with positive values smaller or equal to 1; obtain the output of the neural network $\widehat{y}(1)$ with equation (11).

3) For each iteration k, obtain the output of the neural network $\widehat{y}(k)$ with equation (11), obtain the identification error $e(k)$ with equation (18), and update the parameters V_{jk+1} and W_{ijk+1} with equation (21).

4) Note that the behavior of the algorithm could be improved by changing the values of M or α_0.

Remark 6: The proposed neural network has one hidden layer. Some earlier results as [4], [19], and [30] mention that there is a result where the feedforward neural network with one hidden layer is enough to approximate any nonlinear system.

Fig. 4. The transelevator with a load

VI. THE WAREHOUSE

An automatic warehouse has elements used to make easy the work of moving loads from one place to another one in an automatic way. The loads are some objects inside of boxes that are saved in the warehouse until they are sent to the costumers. The deposits are the place where the loads are placed. The Figure 2 shows the automatic warehouse in gray color, the deposits in black color and the loads in brown color.

A transelevator moves inside of the warehouse. This transelevator can be used to move some load from one place to another one in the warehouse, for example, from the floor to the deposit, from the deposit to the floor, from one deposit to another one, or from a trailer to the deposits. The Figure 3 shows a transelevator inside of the warehouse in yellow color and the Figure 4 shows the same transelevator moving a load.

The Figure 5 shows the trailer with the loads that are saved in the warehouse. The transelevator takes the loads from the trailer and place them in the deposits.

Fig. 5. The trailer with loads for the warehouse

In this paper, the main prediction problem in the warehouse is the distribution of the loads that the transelevator receive from the trailer and place in the deposits each hour inside the warehouse, the deposits in the warehouse can be reserved in advance using the prediction results.

VII. Simulations

In this section, two examples are considered. In the first example, the backpropagation algorithm is applied for the prediction of the distribution of loads inside a warehouse, the proposed algorithm is compared with the recursive least square algorithm given by [9] and used by [1] and [17] and with the Sugeno fuzzy inference system given by [14] and [30]. In the second example, the backpropagation algorithm is applied in an academic problem, the proposed algorithm is compared with the recursive least square algorithm given by [9] and used by [1] and [17] and with the Kalman filter algorithm given by [9] and [10] and used by [25].

The root mean square error (RMSE) [16] is used and it is given as follows:

$$RMSE = \left(\frac{1}{N}\sum_{k=1}^{N} e^2(k)\right)^{\frac{1}{2}} \tag{44}$$

Example 1: In this example, the backpropagation is applied for the prediction of the distribution of loads that the transelevator receive from the trailer and place in the deposits each hour in the warehouse, there are 3 kind of loads received by the transelevator inside the warehouse, the 3 kind of loads are denoted as A, B and C, the 3 kind of loads are received in the warehouse each hour, the number of loads of the kind A received each hour in the warehouse can vary from 4 to 5, the number of loads of the kind B received each hour in the warehouse can vary from 3 to 4 and the number of loads of the kind C received each hour in the warehouse can vary from 1 to 3. The data of 1800 hours are used for the training and the data of the least 200 hours are used for the testing, the prediction is obtained with 200 hours in advance. 3 neural networks are used for the training and the same neural networks are used for the testing, $B(k)$ and $C(k)$ are the inputs and $A(k+200)$ is the output for the training of the first neural network, $A(k)$ and $C(k)$ are the inputs and $B(k+200)$ is the output for the training of the second neural network, $A(k)$ and $B(k)$ are the inputs and $C(k+200)$ is the output for the training of the third neural network. Similar inputs are used for the testing of the three neural networks, and the outputs are not used for the testing. In this prediction example, the backpropagation algorithm is given as (11), (18), and (21) changing $y(k)$ by $y(k+200)$ and changing $e(k)$ by $e(k+200)$ [9]. The parameters of the backpropagation algorithm are $N=2$, $M=5$, V_{j1} and W_{ij1} are random number between 0 and 1, and $\alpha_0 = 1$. The backpropagation algorithm is compared with the recursive least square algorithm given by [9] and used by [1] and [17] with parameters $P_1 = cI \in \Re^{2\times 2}$, where $c = 1$, V_1 are random number between 0 and 1 and is compared with the Sugeno fuzzy inference system given by [14] and [30] with parameters $M=2$, m_1, σ_1 and v_1 are random number between 0 and 1. The training results are shown in the Figure

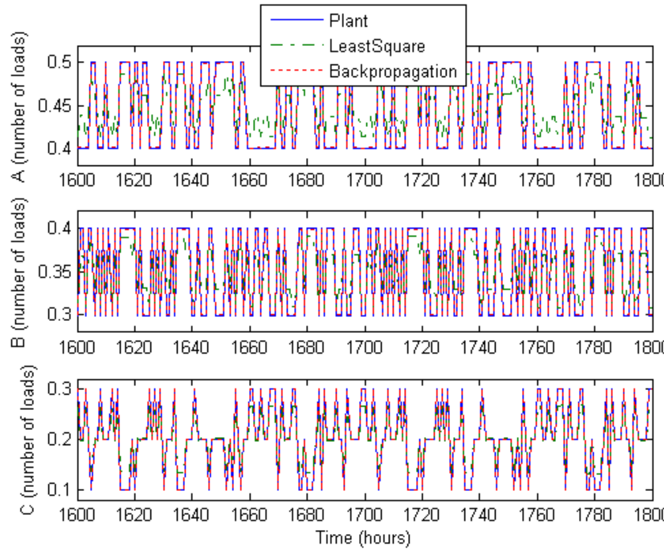


Fig. 6. Training results for example 1

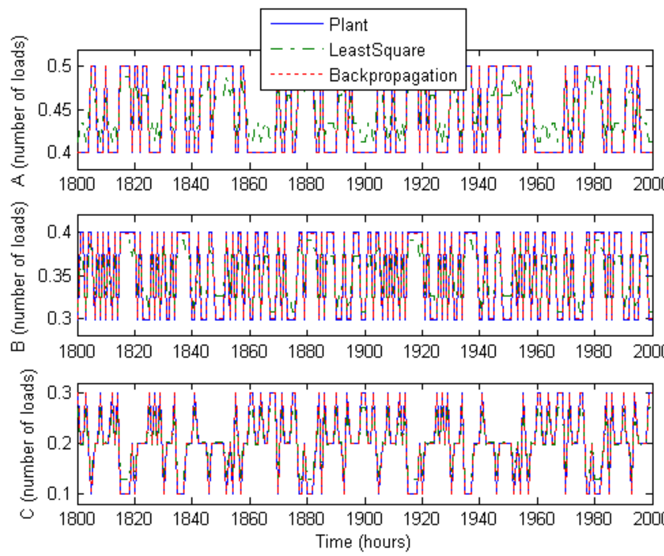


Fig. 7. Testing results for example 1

6 and the testing results are shown in the Figure 7, the Table 1 shows the training and the testing RMSE results using (44). The Figure 8 shows that in this example not all the algorithms are stable because the Sugeno fuzzy inference system is not stable and it is reported in the Table 1.

Table 1: Results for Example 1

Methods	Training RMSE	Testing RMSE
Recursive Least Square	0.0717	0.0121
Backpropagation	0.0321	3.2561×10^{-5}
Sugeno Fuzzy Inference	NAN	$-$

From the Table 1, it can be seen that the backpropagation algorithm achieves better accuracy when compared with the recursive least square because the training RMSE and the testing RMSE are smaller for the backpropagation algorithm. From

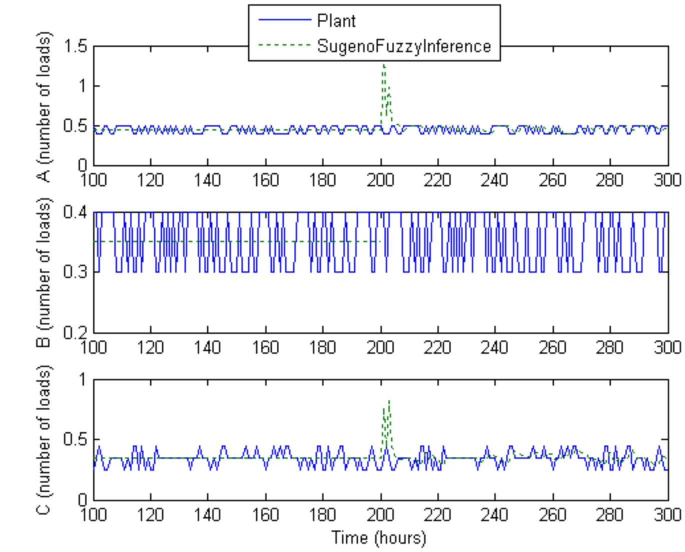


Fig. 8. Training for the Sugeno fuzzy inference system

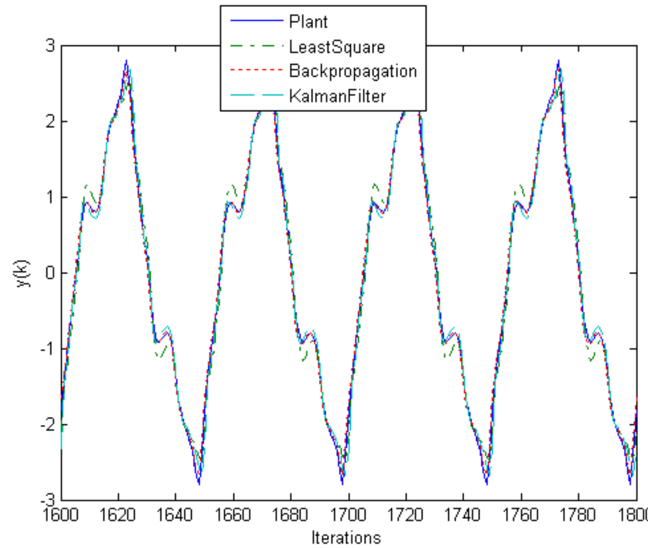


Fig. 10. Training results for example 2

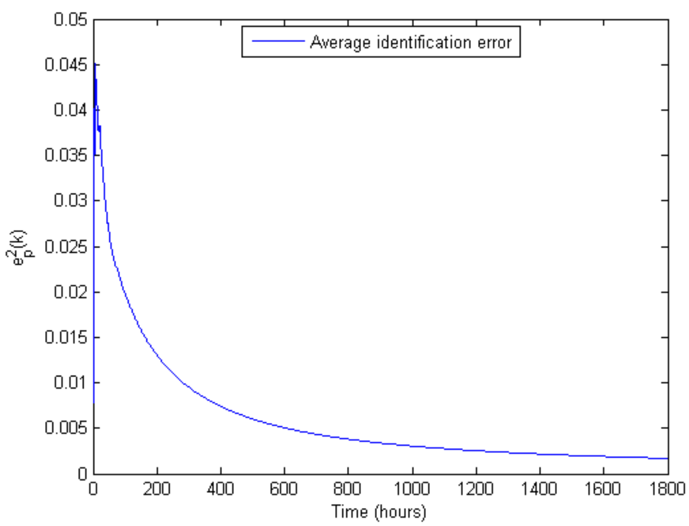


Fig. 9. Average identification error for example 1

the Figures 6 and 7, it can be seen that the backpropagation improves the recursive least square because the signal of the first one follows better the signal of the plant than the signal of the second one. From the Figure 8, the Sugeno fuzzy inference system is unstable for this prediction example, that is way it is important to guarantee the stability of the algorithms. Thus the backpropagation is good for the prediction problems.

The Figure 9 shows the average of the identification error for the modified backpropagation algorithm. From this Figure, it can be observed that the average of the identification error $\limsup_{T\to\infty} \frac{1}{T}\sum_{k=2}^{T} e_p^2(k)$ decrease and it will converge to a value smaller than the upper bound of the uncertainty $\alpha_0\overline{\mu}^2$, as stated in the Theorem 2.

The simulation of the weights error for the Theorem 3 cannot be obtained because the optimal weights which can minimize the modeling error are unknown [19].

Example 2: Let us consider the nonlinear system given in an earlier study [30]:

$$y(k) = 0.3y(k-1) + 0.6y(k-2) + f(u(k-1)) \qquad (45)$$

With $f(u(k-1)) = 0.6\sin(\pi u(k-1)) + 0.3\sin(3\pi u(k-1)) + 0.1\sin(5\pi u(k-1))$, the input is $u(k-1) = \sin(8\pi(k-1)/200)$. In this example, the backpropagation algorithm given as (11), (18), and (21) is used for the identification of the nonlinear plant (45). The parameters of the backpropagation algorithm are $N = 2$, $M = 3$, V_{j1} and W_{ij1} are random number between 0 and 1, and $\alpha_0 = 0.25$. The backpropagation algorithm is compared with the recursive least square algorithm given by [9] and used by [1] and [17] with parameters $P_1 = cI \in \Re^{2\times 2}$, where $c = 1$, V_1 are random number between 0 and 1, and is compared with the Kalman filter algorithm given by [9] and [10] and used by [25] with parameters $P_1 = cI \in \Re^{2\times 2}$, where $c = 1$, $R_1 = 0.1$, $R_2 = 1$, V_1 are random number between 0 and 1. The training results are shown in the Figure 10 and the testing results are shown in the Figure 11, using (44) the Table 2 shows the training and the testing RMSE results.

Table 2: Results for Example 2

Methods	Training RMSE	Testing RMSE
Recursive Least Square	0.0714	0.0183
Kalman Filter	0.0520	0.0283
Backpropagation	0.0413	0.0132

From the Table 2, it can be seen that the backpropagation algorithm achieves better accuracy when compared with the recursive least square and the Kalman filter because the training RMSE and the testing RMSE are smaller for the backpropagation algorithm. From the Figures 10 and 11, it can be seen that the backpropagation improves the recursive least square and the Kalman filter because the signal of the

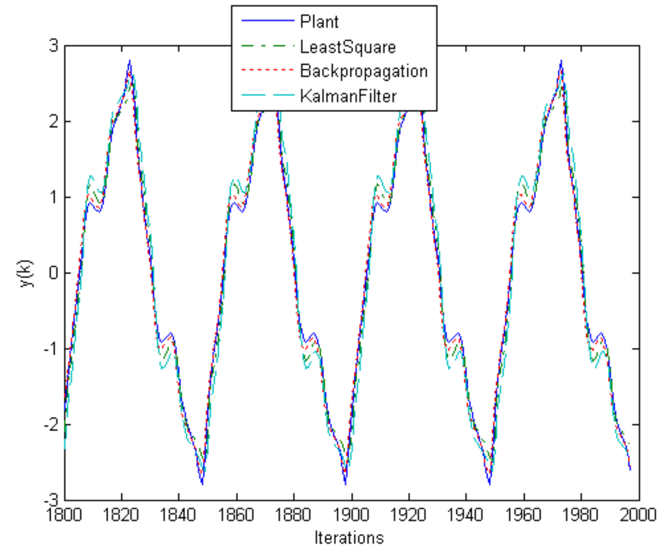


Fig. 11. Testing results for example 2

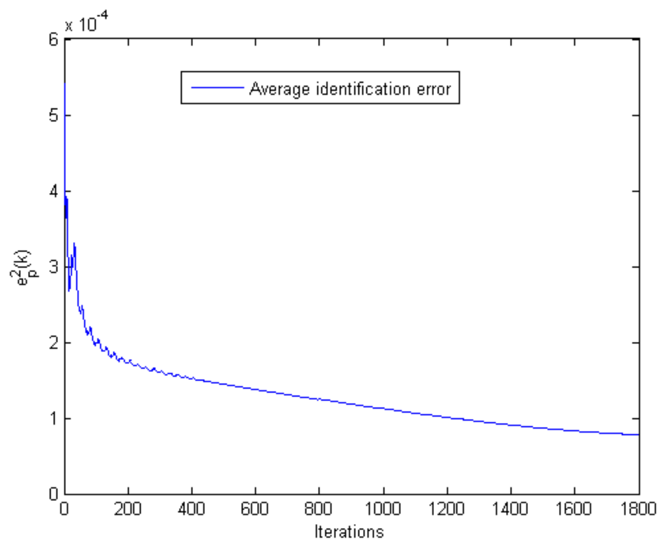


Fig. 12. Average identification error for example 2

first follows better the signal of the plant than the signal of the second and the third. Thus, the backpropagation is good for the identification problems.

The Figure 12 shows the average of the identification error for the modified backpropagation algorithm. From this Figure, it can be observed that the average of the identification error $\limsup_{T\to\infty}\frac{1}{T}\sum_{k=2}^{T}e_p^2(k)$ decrease and it will converge to a value smaller than the upper bound of the uncertainty $\alpha_0\overline{\mu}^2$, as stated in the Theorem 2.

The simulation of the weights error for the Theorem 3 cannot be obtained because the optimal weights which can minimize the modeling error are unknown [19].

VIII. Conclusion

In this paper, it was proposed a theorem to assure the uniform stability of discrete time systems, it was proven that the backpropagation algorithm with a new time varying rate is uniformly stable for online identification, the identification error converges to a small zone bounded by the uncertainty, and the weights error are bounded by their initial weights error. The backpropagation algorithm was compared with the recursive least square algorithm and with the Sugeno fuzzy inference system in the problem of the prediction of the distribution of loads each hour inside a warehouse and the backpropagation algorithm was compared with the recursive least square and with the Kalman filter in an academic example. From the Tables 1 and 2, it can be seen that the backpropagation algorithm achieved better accuracy when compared with the recursive least square algorithm and with the Kalman filter algorithm, in addition, the Sugeno fuzzy inference system was unstable. From the Figures 6, 7, 10, and 11, it can be seen that the backpropagation algorithm improves the recursive least square algorithm and with the Kalman filter algorithm, from the Figure 8, it can be seen the Sugeno fuzzy inference system is unstable in this example. From the simulation results, the backpropagation is good for the prediction and the identification problems. As a future work, an stable algorithm for the radial basis function will be proposed, a new algorithm for the feedforward neural network that guarantees asymptotic stability will be proposed, a method to find the number of neurons in the hidden layer online will be proposed, and the proposed algorithms will be applied for other real problems.

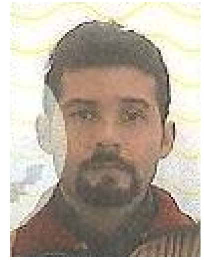
IX. Acknowledgements

The authors thank the Secretaria de Investigación y Posgrado, the Comisión de Operación y Fomento de Actividades Académicas del IPN, and the Consejo Nacional de Ciencia y Tecnologia for their help in this research.

References

[1] P. P. Angelov, D. P. Filev, A approach to Online Identification of Takagi-Sugeno Fuzzy Models, *IEEE Transactions on Systems, Man and Cybernetics*, Vol.32, No.1, 484-498, 2004.
[2] D. J. Berndt, A. R. Hevner, J. Studnicki, The Catch data warehouse: support for community health care decision-making, *Decision Support Systems*, Vol. 35, 367-384, 2003
[3] C. I. Byrnes, A. Isidori and J. C. Willems, Passivity, feedback equivalence, and the global estabilization of minimum phase nonlinear systems, *IEEE Transactions on Automatic Control*, Vol.36, pp. 1228-1240, 1991.
[4] G. Cybenco, Approximation by superposition of sigmoidal activation function, *Math. Control Sig. Syst.*, 2, 303-314, 1989.
[5] Eronini I, Umez Eronini E., System dinamics and control, ISBN 970-686-041-X, 1998.
[6] C.I. Ezeife, Selecting and materializing horizontally partitioned warehouse views, *Data & Knowledge Engineering*, Vol. 36, 185-210, 2001.
[7] B. Egardt, Stability of Adaptive Controlers, *Springer-Verlag Berlin*, *Lecture Notes in Control and Information Sciences*, Vol.20, 1979.
[8] V. Giordano, J. Bing, D. Naso, F. Lewis, Integrated Supervisory and Operational Control of a Warehouse With a Matrix-Based Approach, *IEEE Transactions on Automation Science and Engineering*, Vol. 5, No. 1, 53-70, 2008
[9] G. C. Goodwin and K. S. Sin, *Adaptive filtering prediction and control*, Prentice Hall, Englewood Cliffs, NJ07632, 1984.
[10] S. Haykin, *Neural Networks- A Comprehensive Foundation*, Macmillan College Publ. Co., New York, 1994.

[11] J. R. Hilera, V. J. Martines, *Redes Neuronales Artificiales, Fundamentos, Modelos y Aplicaciones*, Addison Wesley Iberoamericana, U.S.A., 1995.
[12] P. A. Ioannou and J. Sun, *Robust Adaptive Control*, Prentice-Hall, Inc, Upper Saddle River: NJ, 1996.
[13] S. Jagannathan, Control of a Class of Nonlinear Discrete-Time Systems Using Multilayer Neural Networks, IEEE Transactions on Neural Networks, Vol. 12, No. 5, 1113-1120, 2001.
[14] J. S. R. Jang, C. T. Sun, *Neuro-Fuzzy and soft computing*, Prentice Hall, NJ, 1996.
[15] C.F. Juang and C.T. Lin, An Online Self Constructing Neural Fuzzy Inference Network and Its Applications, *IEEE Transactions on Fuzzy Systems*, Vol.6, No.1, 12-32, 1998.
[16] N. K. Kasabov, Evolving Fuzzy Neural Networks for Supervised/Unsupervised Online Knowledge-Based Learning, *IEEE Transactions on Systems, Man and Cybernetics*, Vol.31, No.6, 902-918, 2001.
[17] N. K. Kasabov, Q. Song, DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and Its Application for Time-Series Prediction, *IEEE Transactions on Fuzzy Systems*, Vol.10, No.2, 144-154, 2002.
[18] Y. H. Kim, F. L. Lewis, Optimal Design of CMAC Neural-Network Controller for Robot Manipulators, *IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews*, Vol.30, No.1, pp. 22-30, 2000.
[19] E.B. Kosmatopoulos, M.M. Polycarpou, M.A. Christodoulou, P. A. Ioannou, High-order neural network structures for identification of dynamic systems, *IEEE Transactions on Neural Networks*, Vol.6, No.2, 422-431, 1995.
[20] C. Kwan, F. L. Lewis, D. M. Dawson, Robust Neural Network Control of Rigid-Link Electrically Driven Robots, *IEEE Transactions on Neural Networks*, Vol.9, No.8, 591-588, 1998.
[21] G. Loreto, R. Garrido, Stable Neurovisual Servoing for Robot Manipulators, *IEEE Transactions on Neural Networks*, Vol.17, No.4, 953-965, 2006.
[22] H. R. Nemati, D. M. Steiger, L. S. Iyer, R. T. Herschel, Knowledge warehouse: an architectural integration of knowledge management, decision support, artif icial intelligence and data warehousing, *Decision Support Systems*, Vol. 33, 143-161, 2002.
[23] G. V. Puskorius and L. A. Feldkamp, Neurocontrol of Nonlinear Dynamical Systems with Kalman Filter Trained Recurrent Networks, *IEEE Transactions on Neural Networks*, Vol.5, No.2, 279-297, March 1994.
[24] J. J. Rubio, W. Yu, Stability Analysis of Nonlinear System Identification via Delayed Neural Networks, IEEE Transactions on Circuits and Systems-II, Vol. 54, No. 2, 161-165, 2007.
[25] J. J. Rubio and W. Yu, Nonlinear system identification with recurrent neural networks and dead-zone Kalman filter algorithm, *Neurocomputing*, Vol.70, 2460-2466, 2007.
[26] J. J. Rubio, SOFMLS, Online Self-Organizing Fuzzy Modified Least-Squares Network, *IEEE Trancactions on Fuzzy Systems*, Vol. 17, No. 6, 1296-1309, 2009.
[27] J. J. Rubio, Stability Analysis for an Online Evolving Neuro-Fuzzy Recurrent Network, *Evolving Intelligent Systems: Methodology and Applications*, John Wiley and Sons - IEEE Press, ISBN: 978-0-470-28719-4, 173-199, 2010.
[28] J. A. K. Suykens, J. Vandewalle, B. D. Moor, Lu´re systems with multilayer perceptron and recurrent neural networks: absolute stability and dissipativity, *IEEE Trancactions on Automatic Control*, Vol. 44, No. 4, 770-774, 1999.
[29] J. A. K. Suykens, B. D. Moor, J. Vandewalle, Robust local stability of multilayer recurrent neural networks, *IEEE Trancactions on Neural Networks*, Vol. 11, No. 1, 222-229, 2000.
[30] L. X. Wang, *A Course in Fuzzy Systems and Control*, Prentice Hall, Englewood Cliffs, NJ, 1997.
[31] Z. Wang, H. Zhang, W. Yu, Robust exponential stability analysis of neural networks with multiple time delays, *Neurocomputing*, Vol. 70, 2534-2543, 2007.
[32] Z. Wang, H. Zhang, W. Yu, Robust Stability of Cohen-Grosberg Neural Networks via State Transmission Matrix, *IEEE Transactions on Neural Networks*, Vol. 20, No. 1, 169-174, 2009.
[33] Z. B. Xu, H. Qiao, J. Peng, B. Zhang, A comparative study of two modeling approaches in nerual networks, *Neural Networks*, Vol. 17, 73-85, 2004.
[34] J. Xu, X. Yao, C.-H. Choi, G. Gou, Materialized View Selection as Constrained Evolutionary Optimization, *IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews*, Vol. 33, No. 4, ,458-467, 2003
[35] W. Yu, X. Li, Discrete-time neuro-identification without robust identification, *IEE Proccedings-Control Theory and Applications,* Vol.150, No.3, 311-316, 2003.
[36] W. Yu, Passivity Analysis for Dynamic Multilayer Neuro Identifier, *IEEE Transactions on Circuits and Systems I: Fundamental Theory an Applications*, Vol. 50, No. 1, 2003.
[37] W. Yu and X. Li, Fuzzy Identification Using Fuzzy Neural Networks with Stable Learning Algorithms, *IEEE Transactions on Fuzzy Systems*, Vol.12, No.3, 411-420, 2004.
[38] W. Yu, Nonlinear system identification, using discrete-time recurrent neural networkswith stable learning algorithms, *Information Sciences*, Vol. 158, 131-147, 2004.
[39] W. Yu, M. A. Moreno, F. Ortiz, System identification using hierarchical fuzzy neural networks with stable learning algorithm, *Journal of Intelligent & Fuzzy Systems*, Vol. 18, 171-183, 2007.
[40] W. Yu, X. Li, Passivity Analysis of Dynamic Neural Networks, with Different Time-scales, *Neural Processing Letters*, Vol. 25, 143-155, 2007.
[41] X. Yu, M. Onder, O. Kaynak, A general Backpropagation Algorithm for Feedforward Neural Networks Learning, *IEEE Transactions on Neural Networks*, Vol. 13, No. 1, 2002.



José de Jesús Rubio was born in México City in 1979. He received the B.S. degree from the Instituto Politécnico Nacional in México in 2001. He received the MS in automatic control form the CINVESTAV IPN in México in 2004, and the Ph.D. in automatic control from the CINVESTAV IPN in México in 2007. He was a full time professor in the Autonomous Metropolitan University - Mexico City from 2006 to 2008. Since 2008, he is a full time professor of the Sección de Estudios de Posgrado e Investigación - Instituto Politécnico Nacional – ESIME Azcapotzalco. He has published 18 papers in international journals, 8 chapters in international books, and he has presented 26 papers in International Conferences. He is a member of the IEEE AFS Adaptive Fuzzy Systems. He is part of the editorial board of the journal Evolving Systems. His research interests are primarily focused on neural networks, stability, evolving intelligent systems, nonlinear and adaptive control systems, neural-fuzzy systems, mechatronic, robotic, delayed systems and modeling.