Chapter 2
Identification of Dynamical Systems Using Recurrent Neurofuzzy Modeling

2.1 The Recurrent Neurofuzzy Model

Let us consider a nonlinear function $f(x, u)$, where $f : R^{n+m} \to R^n$ is a smooth vector field defined on a compact set $\Psi \subset R^{n+m}$, with input space $u \in U_c \subset R^m$ and state-space $x \in X \subset R^n$. Also, we assume that the dynamic equation which describes the i/o behavior of a system has the following form (Christodoulou et al. 2007; Theodoridis et al. 2009, 2012):

$$\dot{x}(t) = f(x(t), u(t)), \tag{2.1}$$

or in a per-state form:

$$\dot{x}_i(t) = f_i(x(t), u(t)), \tag{2.2}$$

where $f_i(\cdot)$, $i = 1, 2, \ldots, n$, is a continuous function and t denotes the temporal variable. In order to proceed further we have to state the following assumption:

Assumption 1 Notice that since $\Psi \subset \Re^{n+m}$ then Ψ is closed and bounded set. Also, it is noted that even if Ψ is not compact we may assume that there is a time instant T such that $(x(t), u(t))$ remain in a compact subset of Ψ for all $t < T$; i.e. if $\Psi_T := \{(x(t), u(t)) \in \Psi, t < T\}$. The interval Ψ_T represents the time period over which the approximation is to be performed.

We consider that function $f(x, u)$ is approximated by a fuzzy system using appropriate fuzzy rules. In this framework, let Ω_f be defined as the universe of discourse of $(x, u) \in X \cup U \subset R^{n+m}$ belonging to the $(j_1, j_2, \ldots, j_{n+m})$th input fuzzy patch and pointing—through the vector field $f(\cdot)$—to the subset that belongs to the $l_1, l_2, \ldots, l_n$th output fuzzy patch. Also, Ω_{f_i} is a subset of Ω_f containing input pair values associated with f_i. Furthermore, $\Omega^p_{f_i}$, with $p = 1, 2, \ldots, q$ the number of fuzzy partitions of the i-th state variable, is defined as the p-th subregion of Ω_{f_i} such that $\Omega_{f_i} = \bigcup_{p=1}^{q} \Omega^p_{f_i}$.

Y. Boutalis et al., *System Identification and Adaptive Control*,
Advances in Industrial Control, DOI: 10.1007/978-3-319-06364-5_2,

Definition 2.1 According to the above notation the indicator function (IF) connected to $\Omega_{f_i}^p$ is defined as follows:

$$I_i^p(x(t), u(t)) = \begin{cases} \alpha_i^p(x(t), u(t)) & \text{if } (x(t), u(t)) \in \Omega_{f_i}^p \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

where $\alpha_i^p(x(t), u(t))$ denotes the firing strength of the rule.

Then, assuming a standard defuzzification procedure (e.g., centroid of area or weighted average, see Sect. 1.3.1), the functional representation of the fuzzy system that approximates the real one can be written as

$$\hat{f}_i(x(t), u(t)) = \frac{\sum\limits_{p=1}^{q} I_i^p \cdot \bar{x}_{f_i}^p}{\sum\limits_{p=1}^{q} I_i^p}, \tag{2.4}$$

where the summation is carried over all the available fuzzy rules.

Definition 2.2 Using the notation presented in Sect. 1.3.1, we can define as weighted IF (WIF) the following equation:

$$\left(I'\right)_i^p = \frac{I_i^p}{\sum\limits_{p=1}^{q} I_i^p} \tag{2.5}$$

which is the IF defined in (2.3) divided by the sum of all IF participating in the summation of (2.4).

Thus, Eq. (2.4) can be rewritten as

$$\hat{f}_i(x(t), u(t)) = \sum_{p=1}^{q} \left(I'\right)_i^p \cdot \bar{x}_{f_i}^p. \tag{2.6}$$

Based on the fact that functions of high-order neurons are capable of approximating discontinuous functions (Kosmatopoulos and Christodoulou 1996; Christodoulou et al. 2007), we use high-order neural networks (HONNs) to approximate a $\left(I'\right)_i^p$. Thus, we have the following definition:

Definition 2.3 A HONN is defined as:

$$N_i^p(x(t), u(t); w, k) = \sum_{l=1}^{k} w_{f_i}^{pl} \prod_{j \in I_l} \Phi_j^{d_j(l)}, \tag{2.7}$$

where $I_l = \{I_1, I_2, \ldots, I_k\}$ is a collection of k not-ordered subsets of $\{1, 2, \ldots, n + m\}$, $d_j(l)$ are nonnegative integers. Φ_j are the elements of the following vector:

$$\Phi = \left[\Phi_1 \ \ldots \ \Phi_n \ \Phi_{n+1} \ \ldots \ \Phi_{n+m}\right]^T = \left[s(x_1) \ \ldots \ s(x_n) \ s(u_1) \ \ldots s(u_m)\right]^T,$$

where s denotes the sigmoid function defined as:

$$s(x) = \frac{\alpha}{1 + e^{-\beta x}} - \gamma, \tag{2.8}$$

with α, β being positive real numbers and γ being a real number.

Special attention has to be given in the selection of parameters α, β, γ so that $s(x)$ fulfill the persistency of excitation condition $\left(s \in [-\gamma, -\gamma + \alpha] \text{ when } \gamma < 0\right)$ required in some system identification tasks. Also, $w_{f_i}^{pl}$ is the HONN weights with $i = 1, 2, \ldots, n$, $p = 1, 2, \ldots, q$ and $l = 1, 2, \ldots, k$.

Thus, Eq. (2.7) can be written as

$$N_i^p(x(t), u(t); w, k) = \sum_{l=1}^{k} w_{f_i}^{pl} s_l(x(t), u(t)), \tag{2.9}$$

where $s_l(x(t), u(t))$ are high-order terms of sigmoid functions of the state and/or input.

The next lemma (Kosmatopoulos and Christodoulou 1996) states that a HONN of the form in Eq. (2.9) can approximate the weighted indicator function (WIF), $\left(I'\right)_i^p$.

Lemma 2.1 *Consider the WIF* $\left(I'\right)_i^p$ *and the family of HONN's* $N_i^p(x(t), u(t); w, k)$. *Then for any* $\varepsilon_i^p > 0$, *there is a vector of weights w and a number of k high-order connections such that:*

$$\sup_{(x(t),u(t))\in\Psi} \left\{ \left(I'\right)_i^p (x(t), u(t)) - \sum_{l=1}^{k} w_{f_i}^{pl} s_l(x(t), u(t))\right\} < \varepsilon_i^p$$

The magnitude of approximation error $\varepsilon_i^p > 0$ *depends on the choice of the number of high-order terms.*

Under the definition of WIFs and the above lemma, one could rewrite the rules of the fuzzy system as follows:

$$R_i^p : \text{IF } (x(t), u(t)) \in \Omega_{f_i}^p \text{ THEN HONN}_p \text{ is } \left(I'\right)_i^p (t).$$

Following the above analysis and Eq. (2.6), actually we give a weighting value according to the output fuzzy partitioning, as shown in Fig. 2.1, to every HONN that participates to the estimation of $f_i(x, u)$.

As a consequence, we have the following definition:

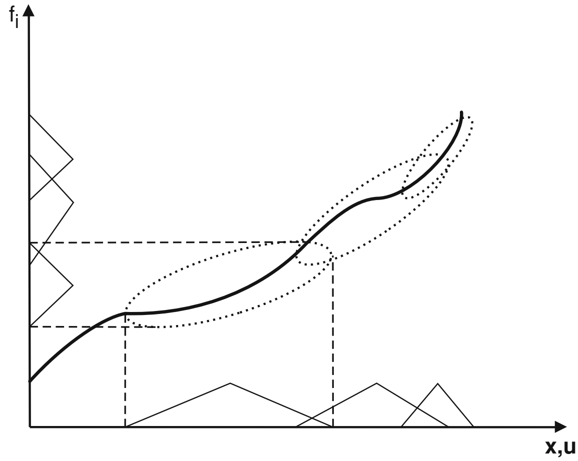

Fig. 2.1 Fuzzy partitioning of the system output

Definition 2.4 The center weighting value (CWV) $\bar{x}^{p}_{f_i}$ which is the p-th fuzzy center of the i-th state variable (or equivalently f_i) influences a HONN by a degree of implementation $\bar{x}^{p}_{f_i}$.

Therefore, rule R^p_i can be equivalently expressed as

R^p_i: IF $(x(t), u(t)) \in \Omega^p_{f_i}$ THEN HONN$_p$ is $\left(I'\right)^p_i(t)$ with CWV $\bar{x}^p_{f_i}$.

Now, we can group the rules that participate in the construction of the i-th state variable output according to the following form:

R_i: IF $(x(t), u(t)) \in \Omega_{f_i}$ THEN HONN$_1$ is $\left(I'\right)^1_i(t)$ with CWV $\bar{x}^1_{f_i}$ and HONN$_2$ is $\left(I'\right)^2_i(t)$ with CWV $\bar{x}^2_{f_i}$ and $\cdots$ and HONN$_q$ is $\left(I'\right)^q_i(t)$ with CWV $\bar{x}^q_{f_i}$.

It then follows easily that, the i-th state variable of the system output is determined as follows:

$$
\begin{aligned}
&R_i : \text{IF } (x(t), u(t)) \in \Omega_{f_i} \text{ THEN} \\
&f_i(x, u) = \left(I'\right)^1_i(t) \cdot \bar{x}^1_{f_i} + \left(I'\right)^2_i(t) \cdot \bar{x}^2_{f_i} + \cdots + \left(I'\right)^q_i(t) \cdot \bar{x}^q_{f_i},
\end{aligned}
$$

where each $\left(I'\right)^l_i, l = 1, \ldots, q$ is replaced by the respective HONN. It is clear that the information about the antecedent partitioning of the rules as well as the number of rules is not necessary to be determined here. Therefore, the rules are not treated here in the classical way of Mamdani or Takagi-Sugeno definition but their consequent parts are determined directly from F-HONNs.

Following the above notation, Eq. (2.6) in conjunction with Eq. (2.9) can be rewritten as

$$
\hat{f}_i(x(t), u(t)) = \sum_{p=1}^{q} \bar{x}^p_{f_i} \cdot \left(\sum_{l=1}^{k} w^{pl}_{f_i} \cdot s_l(x(t), u(t)) \right), \tag{2.10}
$$

or in a more compact form:

$$\dot{\hat{f}} = X_f W_f s_f(x, u). \tag{2.11}$$

An alternative, recurrent NF form of Eq. (2.1) which will be used in the subsequent analysis of this thesis is:

$$\dot{\hat{x}} = A\hat{x} + \hat{f}. \tag{2.12}$$

Considering that f is approximated by the NF model described above, Eq. (2.12) can be rewritten as

$$\dot{\hat{x}} = A\hat{x} + X_f W_f s_f(x, u), \tag{2.13}$$

where A is a $n \times n$ stable matrix, which for simplicity can be taken to be diagonal as $A = \text{diag}[-a_1, -a_2, \ldots, -a_n]$, with $a_i > 0$. Also, X_f is a matrix containing the centers of the partitions of every fuzzy output variable of $f(x, u)$, $s_f(x, u)$ is a vector containing high-order combinations of sigmoid functions of the state x and control input u. Also, W_f is a matrix containing respective neural weights according to (2.9) and (2.10). For notational simplicity we assume that all output fuzzy variables are partitioned to the same number, q, of partitions. Under these specifications X_f is a $n \times n \cdot q$ block diagonal matrix of the form $X_f = \text{diag}(\bar{x}_{f_1}, \bar{x}_{f_2}, \ldots, \bar{x}_{f_n})$, with $\bar{x}_{f_i}$ being a q-dimensional row vector of the form:

$$\bar{x}_{f_i} = \left[\bar{x}^1_{f_i} \; \bar{x}^2_{f_i} \; \cdots \; \bar{x}^q_{f_i}\right],$$

or in a more detailed form:

$$X_f = \begin{bmatrix} \bar{x}^1_{f_1} & \cdots & \bar{x}^q_{f_1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \bar{x}^1_{f_2} & \cdots & \bar{x}^q_{f_2} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \bar{x}^1_{f_n} & \cdots & \bar{x}^q_{f_n} \end{bmatrix}. \tag{2.14}$$

Also, $s_f(x) = \left[s_1(x) \ldots s_k(x)\right]^T$, where each $s_l(x)$ with $l = 1, 2, \ldots, k$, is a high-order combination of sigmoid functions of the state variables and input signals. Finally, W_f is a $n \cdot q \times k$ matrix with neural weights. W_f assumes the form $W_f = \left[W_{f_1} \; \cdots \; W_{f_n}\right]^T$, where each W_{f_i} is a matrix $\left[w^{pl}_{f_i}\right]_{q \times k}$ and is given as:

$$W_{f_i} = \begin{bmatrix} w^{11}_{f_i} & w^{12}_{f_i} & \cdots & w^{1k}_{f_i} \\ w^{21}_{f_i} & w^{22}_{f_i} & \cdots & w^{2k}_{f_i} \\ \vdots & \vdots & \cdots & \vdots \\ w^{q1}_{f_i} & w^{q2}_{f_i} & \cdots & w^{qk}_{f_i} \end{bmatrix},$$

or in a more detailed form:

$$W_f = \begin{bmatrix} w_{f_1}^{11} & w_{f_1}^{12} & \cdots & w_{f_1}^{1k} \\ w_{f_1}^{21} & w_{f_1}^{22} & \cdots & w_{f_1}^{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{f_1}^{q1} & w_{f_1}^{q2} & \cdots & w_{f_1}^{qk} \\ \vdots & \vdots & \ddots & \vdots \\ w_{f_n}^{11} & w_{f_n}^{12} & \cdots & w_{f_n}^{1k} \\ w_{f_n}^{21} & w_{f_n}^{22} & \cdots & w_{f_n}^{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{f_n}^{q1} & w_{f_n}^{q2} & \cdots & w_{f_n}^{qk} \end{bmatrix}.$$

From the above definitions and Eq. (2.10), it is obvious that the accuracy of the approximation of $f_i(x, u)$ depends on the approximation abilities of HONNs and on an initial estimate of the centers of the output membership functions. These centers can be obtained by experts or by offline techniques based on gathered data. Any other information related to the input membership functions is not necessary because it is replaced by the HONNs.

Figure 2.2 shows the overall scheme of the proposed NF modeling that approximates function $f_i(x, u)$ depending only on measurements of x, u. When these measurements are given as inputs to the NF network (input layer) that includes high-order sigmoidal terms, the output of indicator layer gives the weighted IF outputs that influence the corresponding rules according to output fuzzy center (rule layer). The appropriate summation of all rules at each sampling time instant gives the overall output of the function $f_i(x, u)$ (output layer).

2.2 Approximation Capabilities of the Neurofuzzy Model

The approximation problem consists of determining whether by allowing enough high-order connections and fuzzy centers, there exist weights W_f, such that the F-RHONNs model could approximate the input–output behavior of a complex dynamical system of the form (2.1). In this equation the input u belongs to a class U_c of (piecewise continuous) admissible inputs.

By adding and subtracting Ax, where A is a Hurwitz matrix, (2.2) is rewritten as

$$\dot{x} = Ax + g(x, u) \tag{2.15}$$

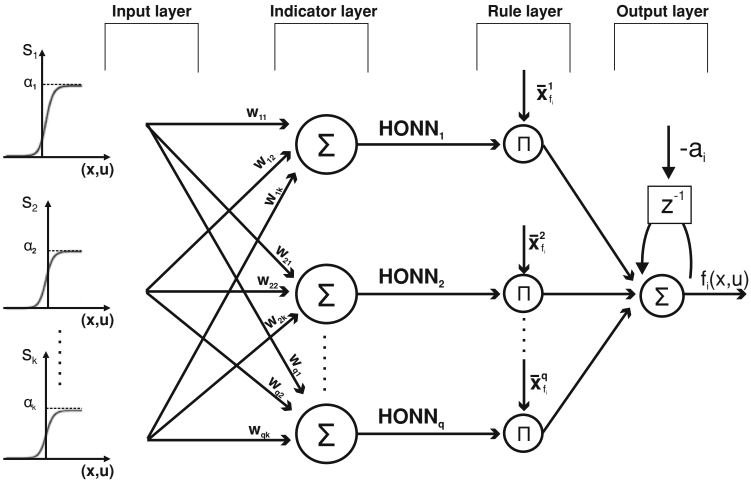

Fig. 2.2 Overall scheme of the proposed NF model that approximates function $f_i(x, u)$ using state measurements x and input signals u

where $g(x, u) := f(x, u) - Ax$.

In order to have a well-posed problem, we will impose the following mild assumptions on the system to be approximated:

Assumption 2 Given a class $U_c \subset R^q$ of admissible inputs, for any $u \in U_c$ and any finite initial condition $x(0)$, the state trajectories are uniformly bounded for any finite $T > 0$. Meaning that we do not allow systems processing trajectories that escape at infinite, in finite time T, T being arbitrarily small. Hence, $|x(T)| < \infty$.

Assumption 3 Functions f_i are continuous with respect to their arguments and satisfy a local Lipschitz condition so that (2.2) has a unique solution for any finite initial condition $x(0)$ and $u \in U_c$, in the sense of Caratheodory (Hale 1969).

Based on the above assumptions, we obtain the following theorem:

Theorem 2.1 *Suppose that the system* (2.1) *and the model* (2.10) *are initially at the same state* $\hat{x}(0) = x(0)$*, then for any* $\varepsilon > 0$ *and any finite* $T > 0$*, there exist integers* k, q*, a matrix* $W_f^\star \in R^{k\times q\times n}$ *and appropriately selected fuzzy output centers* $\bar{x}^p_{f_i}$ *such that the state* $\hat{x}(t)$ *of the F-RHONNs model* (2.10) *with k high-order connections, q fuzzy centers and weight values* $W_f = W_f^\star$ *which satisfies:*

$$\sup_{0\le t\le T} \left|\hat{x}(t) - x(t)\right| \le \varepsilon.$$

Proof Following a procedure similar to the work of Kosmatopoulos et al. (1995), we proceed as follows: By assumption, $(x(t), u(t)) \in \Psi$ for all $t \in [0, T]$, where Ψ is a compact subset of R^{n+m}.

Let $\Psi_e = \{(x,u) \in R^{n+m} : |(x,u) - (x_y, u_y)| \le \varepsilon, (x_y, u_y) \in \Psi\}$. It can be readily seen that Ψ_e is also a compact subset of R^{n+m} and $\Psi \subset \Psi_e$. That is, Ψ_e is ε larger than Ψ, where ε is the required degree of approximation. Since s_f is a continuous function, it satisfies a Lipschitz condition in Ψ_e, i.e., there is a constant l such that for all $(\hat{x}_1, u), (\hat{x}_2, u) \in \Psi_e$:

$$\left|s_f\left(\hat{x}_1, u\right) - s_f\left(\hat{x}_2, u\right)\right| \le l\,|x_1 - x_2|\,. \tag{2.16}$$

In what follows, we show that the function $X_f W_f^{\star} s_f$ satisfies the conditions of Stone–Weirstrass Theorem (Stone 1948; Bishop 1961) and can approximate any continuous function over a compact domain.

The dynamic behavior of F-RHONNs model is described by (2.13). Since $\hat{x}(0) = x(0)$, the state error $e = \hat{x} - x$ satisfies the differential equation

$$\dot{e} = Ae + X_f W_f s_f - g(x,u), \tag{2.17}$$

where $e(0) = 0$.

Therefore, it can be readily shown that if k, q are sufficiently large, then there exist weight values $W_f = W_f^{\star}$ such that $X_f W_f^{\star} s_f(x,u)$ can approximate $g(x,u)$ to any degree of accuracy, for all (x,u) in a compact domain. Hence, there exists $W_f = W_f^{\star}$ such that

$$\sup_{(x,u)\in\Psi_e} \left|X_f W_f^{*} s_f(x,u) - g(x,u)\right| \le \delta, \tag{2.18}$$

where δ is a constant to be designed in the sequel. The solution of (2.17) is

$$\begin{aligned} e(t) &= \int_0^t e^{A(t-\tau)} X_f W_f^{*} s_f(\hat{x}(\tau), u(\tau))\,\mathrm{d}\tau - \int_0^t e^{A(t-\tau)} g(x(\tau), u(\tau))\,\mathrm{d}\tau \\ &= \int_0^t e^{A(t-\tau)} X_f W_f^{*} s_f(\hat{x}(\tau), u(\tau))\,\mathrm{d}\tau - \int_0^t e^{A(t-\tau)} X_f W_f^{*} s_f(x(\tau), u(\tau))\,\mathrm{d}\tau \\ &\quad + \int_0^t e^{A(t-\tau)} X_f W_f^{*} s_f(x(\tau), u(\tau))\,\mathrm{d}\tau - \int_0^t e^{A(t-\tau)} g(x(\tau), u(\tau))\,\mathrm{d}\tau. \end{aligned} \tag{2.19}$$

Since A is a Hurwitz matrix, there exist positive constants c, α such that $\left\|e^{At}\right\| \le ce^{-\alpha t}$ for all $t \ge 0$. Also, let $k = cl\left\|X_f W_f^{\star}\right\|$. Based on the aforementioned definitions of the constants $c, \alpha,\ k,\ \varepsilon$, let δ in (2.18) be chosen as

$$\delta = \frac{\varepsilon\alpha}{2c} e^{-\frac{k}{\alpha}} > 0. \tag{2.20}$$

First consider the case where $(x(t), u(t)) \in \Psi_e$ for all $t \in [0, T]$. Starting from (2.19), taking norms on both sides and using (2.16), (2.18) and (2.20), the following inequalities hold for all $t \in [0, T]$:

$$\begin{aligned}
|e(t)| &\leq \int_0^t \left\| e^{A(t-\tau)} \right\| \left\| X_f W_f^* \right\| \left| s_f(\hat{x}, u) - s_f(x, u) \right| \mathrm{d}\tau \\
&\quad + \int_0^t \left\| e^{A(t-\tau)} \right\| \left| X_f W_f^* s_f(x, u) - g(x, u) \right| \mathrm{d}\tau \\
&\leq \int_0^t e^{-\alpha(t-\tau)} k \left| e(\tau) \right| \mathrm{d}\tau + \int_0^t \delta c e^{-\alpha(t-\tau)} \mathrm{d}\tau \\
&\leq k \int_0^t e^{-\alpha(t-\tau)} \left| e(\tau) \right| \mathrm{d}\tau + \frac{\varepsilon}{2} e^{-\frac{k}{\alpha}}.
\end{aligned} \tag{2.21}$$

Then, using the Bellman–Gronwall Lemma (Hale 1969), we obtain:

$$|e(t)| \leq \frac{\varepsilon}{2} e^{-\frac{k}{\alpha}} \cdot e^{\int_0^t k e^{-\alpha(t-\tau)} \mathrm{d}\tau} \leq \frac{\varepsilon}{2} e^{-\frac{k}{\alpha}} \cdot e^{\frac{k}{\alpha}} \leq \frac{\varepsilon}{2}. \tag{2.22}$$

It should be noted here that the assumption of $\hat{x}(0) = x(0)$ can be easily relaxed without affecting the conclusion of the theorem. In this case, one should consider that an exponentially fast decaying error term is added in ϵ.

The above theorem proves that if sufficiently large number of connections are allowed in F-RHONNs model then it is possible to approximate any dynamical system to any degree of accuracy. This result does not provide us with any constructive method for obtaining the optimal weights $W_f^\star$. In what follows, we consider the learning problem of adjusting the weights adaptively, such that the Neurofuzzy model identifies general dynamic systems.

2.3 Learning Algorithms for Parameter Identification

We proceed now to develop weight updating laws assuming that the unknown system is modeled exactly by an F-RHONNs architecture of the form (2.13). In the next section, we extend this analysis to cover the case where there exists a nonzero mismatch between the system and F-RHONNs model with optimal weight values, that is, we assume the existence of modeling errors.

Following the standard practice in system identification algorithms, we will assume that the input $u(t)$ and the state $x(t)$ remain bounded for all $t \geq 0$. Based on the definition of $s_f(x, u)$, X_f as given by (2.8), (2.14) this implies that $s_f(x, u)$, X_f are also bounded. In the sections that follow, we present different approaches for estimating the unknown parameters $\left(w_{f_i}^{pl}\right)^\star$ of F-RHONNs model.

2.3.1 Simple Gradient Descent

In developing this identification scheme, we start again from the differential equation that describes the unknown system with no modeling error which is given by

$$\dot{x}_i = -a_i x_i + \bar{x}_{f_i} W_{f_i}^\star s_f(x, u). \tag{2.23}$$

Based on (2.23), the identifier is now chosen as

$$\dot{\hat{x}}_i = -a_i \hat{x}_i + \bar{x}_{f_i} W_{f_i} s_f(x, u), \tag{2.24}$$

where W_{f_i} is again the estimate of the unknown optimal weight matrix $W_{f_i}^\star$. In this case, the state error $e_i = \hat{x}_i - x_i$ satisfies

$$\dot{e}_i = -a_i e_i + \bar{x}_{f_i} \tilde{W}_{f_i} s_f(x, u), \tag{2.25}$$

where $\tilde{W}_{f_i} = W_{f_i} - W_{f_i}^\star$.

The next theorem gives the error F-RHONNs model with the gradient method for adjusting the weights.

Theorem 2.2 *Consider the error F-RHONNs model given by* (2.25) *whose weights are adjusted according to equation*

$$\dot{W}_{f_i} = -\bar{x}_{f_i}^T e_i s_f^T P_i. \tag{2.26}$$

Then for $i = 1, 2, \ldots, n$, *the following properties are guaranteed:*

1. $e_i, \tilde{W}_{f_i} \in L_\infty, \quad e_i \in L_2$,
2. $\lim_{t\to\infty} e_i(t) = 0$,
3. $\lim_{t\to\infty} \dot{W}_{f_i}(t) = 0$.

Proof (1) Consider the Lyapunov candidate function:

$$V = \frac{1}{2}\sum_{i=1}^{n} e_i^2 + \frac{1}{2}\sum_{i=1}^{n} \mathrm{tr}\left\{\tilde{W}_{f_i} P_i^{-1} \tilde{W}_{f_i}^T\right\}. \tag{2.27}$$

Taking the time derivatives of the Lyapunov function candidate (2.27) and after substituting Eq. (2.25) we obtain

$$\begin{aligned}
\dot{V} &= \sum_{i=1}^{n} e_i \dot{e}_i + \sum_{i=1}^{n} \operatorname{tr}\left\{ \dot{W}_{f_i} P_i^{-1} \tilde{W}_{f_i}^T \right\} \\
&= -\sum_{i=1}^{n} a_i \left| e_i \right|^2 + \sum_{i=1}^{n} \left(e_i \bar{x}_{f_i} \tilde{W}_{f_i} s_f + \operatorname{tr}\left\{ \dot{W}_{f_i} P_i^{-1} \tilde{W}_{f_i}^T \right\} \right) \\
&= -\sum_{i=1}^{n} a_i \left| e_i \right|^2 .
\end{aligned} \tag{2.28}$$

Considering that in deriving (2.28) we assumed that

$$\operatorname{tr}\left\{ \dot{W}_{f_i} P_i^{-1} \tilde{W}_{f_i}^T \right\} = -e_i \bar{x}_{f_i} \tilde{W}_{f_i} s_f ,$$

and using matrix trace properties we result in Eq. (2.26).
Thus, $\dot{V}$ is negative semidefinite. Since $\dot{V} \leq 0$, we conclude that $V \in L_\infty$, which implies that e_i, $\tilde{W}_{f_i} \in L_\infty$. Furthermore, $W_{f_i} = \tilde{W}_{f_i} + W_{f_i}^*$ is also bounded. Since V is a nonincreasing function of time and bounded from below, the $\lim_{t\to\infty} V = V_\infty$ exists; therefore, by integrating $\dot{V}$ from 0 to ∞ we have

$$\int_0^\infty \sum_{i=1}^{n} a_i \left| e_i \right|^2 dt \leq \left[V(0) - V_\infty \right] < \infty,$$

which implies that $e_i \in L_2$.

(2) Since $e_i \in L_2 \cap L_\infty$, using Barbalat's Lemma we conclude that $\lim_{t\to\infty} e_i(t) = 0$.

(3) Finally, using the boundedness of $\bar{x}_{f_i}$, $s_f(x, u)$ and the convergence of $e_i(t)$ to zero, we have that $\dot{W}_{f_i}$ also converges to zero (Ioannou and Fidan 2006).

Remark 2.1 The above theorem does not imply that the weight estimation error $\tilde{W}_{f_i} = W_{f_i} - W_{f_i}^*$ converges to zero. In order to achieve convergence of the weights to their correct value, the additional assumption of persistent excitation needs to be imposed on the vector $s_f(x, u)$ because $\bar{x}_{f_i}$ satisfies this condition by definition. In particular, $s_f \in R^k$ is said to be *persistently exciting* if there exist positive scalars β_1, β_2 and T such that for all $t \geq 0$:

$$\beta_1 I \leq \int_t^{t+T} s_f(\tau) s_f^T(\tau)\, \mathrm{d}\tau \leq \beta_2 I, \tag{2.29}$$

where I is the $k \times k$ identity matrix.

This can be achieved if the constant γ in Eq. (2.8) is selected such that: $s_f \cdot s_f^T > 0$.

2.3.2 Pure Least Squares

The basic idea behind least squares (LS) method is to fit a mathematical model to a sequence of observed data by minimizing the sum of the squares of the difference between the observed and computed data. This way, any noise or inaccuracies in the observed data are expected to have less effect on the accuracy of the mathematical model (Ioannou and Fidan 2006).

The method is simple to apply and analyze in the case where the unknown parameters appear in a linear form, such as in Eq. (2.23). The pure LS algorithm can be thought as a gradient algorithm with a time-varying learning rate and could be written as follows:

$$
\begin{aligned}
\dot{W}_{f_i} &= -\frac{\bar{x}_{f_i}^T e_i z_i^T P_i}{\left|\bar{x}_{f_i}\right|^2}, \quad W_{f_i}(0) = W_{f0}, \\
\dot{P}_i &= -\frac{P_i z_i z_i^T P_i}{n_s^2}, \quad P_i(0) = P_0,
\end{aligned}
\tag{2.30}
$$

where P_i is the gain matrix which is positive definite and $n_s^2 \geq 1$ is a normalization signal designed to guarantee that $\frac{z_i}{n_s}$ is bounded, with z_i defined in the following lemma 2.2. The property of n_s is used to establish the boundedness of the estimated parameters even when z_i is not guaranteed to be bounded. A straightforward choice for n_s is $n_s^2 = 1 + \alpha z_i^T z_i$, $\alpha > 0$. If z_i is bounded, we can take $\alpha = 0$. The following lemma is useful in the development of the adaptive identification algorithm, which is presented in this section.

Lemma 2.2 *The system described by* Eq. (2.23) *can be expressed as*

$$\dot{z}_i = -a_i z_i + s_f, \;\; z_i(0) = 0, \tag{2.31}$$

$$x_i = \bar{x}_{f_i} W_{f_i}^{\star} z_i + e^{-a_i t} x_i(0). \tag{2.32}$$

Proof From (2.31) after integrating we have

$$z_i(t) = \int_0^t e^{-a_i(t-\tau)} s_f\left(x(\tau), u(\tau)\right) \mathrm{d}\tau,$$

therefore,

$$\bar{x}_{f_i} W_{f_i}^{\star} z_i + e^{-a_i t} x_i(0) = e^{-a_i t} x_i(0) + \int_0^t e^{-a_i(t-\tau)} \bar{x}_{f_i} W_{f_i}^{\star} s_f\left(x(\tau), u(\tau)\right) \mathrm{d}\tau. \tag{2.33}$$

Using (2.32), the right-hand side of (2.33) is equal to $x_i(t)$ and this concludes the proof.

Using the above lemma the dynamical system is described by the following equation:

$$x_i = \bar{x}_{f_i} W^{\star}_{f_i} z_i + \varepsilon_i, \tag{2.34}$$

where $\varepsilon_i = e^{-a_i t} x_i(0)$ is an exponentially decaying term that appears when a nonzero initial state is applied. After ignoring the exponentially decaying term ε_i (Rovithakis and Christodoulou 2000), the F-RHONNs model can be written as

$$\hat{x}_i = \bar{x}_{f_i} W_{f_i} z_i. \tag{2.35}$$

The state error equation $e_i = \hat{x}_i - x_i$, after substituting (2.34), (2.35) becomes

$$e_i = \bar{x}_{f_i} \tilde{W}_{f_i} z_i - \varepsilon_i. \tag{2.36}$$

The cost function $J(W_{f_i})$ is chosen as

$$J(W_{f_i}) = \frac{\sum_{i=1}^{n} e_i^2}{2} = \frac{\sum_{i=1}^{n} \left[\left(\bar{x}_{f_i} W_{f_i} z_i - \bar{x}_{f_i} W^{\star}_{f_i} z_i \right) - \varepsilon_i \right]^2}{2}. \tag{2.37}$$

If we use the LS method described by (2.30) and (2.31), a problem that may be encountered in the application of the LS's algorithm is that P_i may become arbitrarily small and thus slow down adaptation in some directions. Therefore, we can use one of various modifications that prevent $P_i(t)$ from going to zero as follows: if the smallest eigenvalue of $P_i(t)$ becomes smaller than ρ_1 then $P_i(t)$ is reset to $P_i(t) = \rho_0 I$, where $\rho_0 \geq \rho_1 > 0$ are some design constants (Rovithakis and Christodoulou 2000).

Theorem 2.3 *The pure LS algorithm given by* (2.30), (2.31) *guarantees the following properties:*

1. $e_i, \dot{W}_{f_i} \in L_2 \cap L_\infty, \quad W_{f_i}, P_i \in L_\infty$.
2. $\lim_{t \to \infty} e_i(t) = 0$,
3. $\lim_{t \to \infty} \dot{W}_{f_i}(t) = 0$.

Proof (1) From (2.31) we have that $\dot{P}_i \leq 0$, i.e., $P_i(t) \leq P_0$. Because $P_i(t)$ is nonincreasing and bounded from below (i.e., $P_i(t) = P_i^T(t) \geq 0$) it has a limit, i.e.,

$$\lim_{t \to \infty} P_i(t) = \bar{P}_i,$$

where $\bar{P}_i = \bar{P}_i^T \geq 0$ is a constant positive definite diagonal matrix and thus $P_i \in L_\infty$.

Let us now consider the Lyapunov candidate function

$$V = \frac{1}{2} \sum_{i=1}^{n} \left[\left(\bar{x}_{f_i} \tilde{W}_{f_i} \right) P_i^{-1} \left(\bar{x}_{f_i} \tilde{W}_{f_i} \right)^T + \int_t^{\infty} \varepsilon_i^2 (\tau) \, \mathrm{d}\tau \right]. \tag{2.38}$$

Taking the time derivatives of the Lyapunov function candidate (2.38) and considering the equations, $\frac{\mathrm{d}}{\mathrm{d}t} \left(P^{-1} \right) = -P^{-1} \dot{P} P^{-1}$, (2.30), (2.31), (2.36) we obtain

$$\begin{aligned}
\dot{V} &= \sum_{i=1}^{n} \left(\bar{x}_{f_i} \dot{W}_{f_i} P_i^{-1} \tilde{W}_{f_i}^T \bar{x}_{f_i}^T \right) + \frac{1}{2} \sum_{i=1}^{n} \left[\left(\bar{x}_{f_i} \tilde{W}_{f_i} \right) \dot{P}_i^{-1} \left(\bar{x}_{f_i} \tilde{W}_{f_i} \right)^T - \varepsilon_i^2 \right] \\
&= -\sum_{i=1}^{n} \left(\bar{x}_{f_i} \tilde{W}_{f_i} z_i e_i \right) + \frac{1}{2} \sum_{i=1}^{n} \left[\frac{\left| \bar{x}_{f_i} \tilde{W}_{f_i} z_i \right|^2}{n_s^2} - \varepsilon_i^2 \right] \\
&= -\sum_{i=1}^{n} \left[e_i \left(e_i + \varepsilon_i \right) \right] + \frac{1}{2} \sum_{i=1}^{n} \left(\frac{(e_i + \varepsilon_i)^2}{n_s^2} - \varepsilon_i^2 \right) \\
&= -\frac{1}{2} \sum_{i=1}^{n} \left[e_i^2 + (e_i + \varepsilon_i)^2 \right] + \frac{1}{2} \sum_{i=1}^{n} \frac{(e_i + \varepsilon_i)^2}{n_s^2} \\
&\leq -\frac{1}{2} \sum_{i=1}^{n} {e_i}^2 \leq 0.
\end{aligned} \tag{2.39}$$

Equation (2.39) implies that $V \in L_\infty$, and therefore $\tilde{W}_{f_i} \in L_\infty$. Then, Eq. (2.36) in conjunction with the boundedness of z_i, gives $e_i \in L_\infty$. Furthermore, $W_{f_i} = \tilde{W}_{f_i} + W_{f_i}^*$ is also bounded. Since V is a nonincreasing function of time and bounded from below, the $\lim_{t \to \infty} V = V_\infty$ exists; therefore, by integrating $\dot{V}$ from 0 to ∞ we have

$$\frac{1}{2} \int_0^{\infty} \sum_{i=1}^{n} {e_i}^2 \leq \left[V(0) - V_\infty \right] < \infty,$$

which implies that $e_i \in L_2$. From (2.30) we have

$$\left\| \dot{W}_{f_i} \right\| \leq \frac{|e_i| \, |z_i| \, \| P_i \|}{\left| \bar{x}_{f_i} \right|}. \tag{2.40}$$

Since $\bar{x}_{f_i}, P_i, z_i, e_i \in L_\infty$, and $e_i \in L_2$, we have $\dot{W}_{f_i} \in L_2 \cap L_\infty$.

(2) Since $e_i \in L_2 \cap L_\infty$, using Barbalat's Lemma we conclude that $\lim_{t \to \infty} e_i(t) = 0$.

(3) Finally, using the boundness of $\bar{x}_{f_i}$, $s_f(x, u)$ and the convergence of $e_i(t)$ to zero, we have that $\dot{W}_{f_i}$ also converges to zero (Ioannou and Fidan 2006).

Also, if a persistency of excitation condition such as Remark 2.1 is valid then $W_{f_i}(t) \to W_{f_i}^\star$ as $t \to \infty$.

2.4 Robust Learning Algorithms

Due to an insufficient number of high-order terms or fuzzy output centers in the F-RHONNs model, we have to deal with unmodeled dynamics, noises, disturbances, and other frequently encountered uncertainties. In such cases, if standard adaptive laws are used for updating the weights, then the presence of modeling error in problems related to learning in dynamic environments may cause the adjusted weight values (and consequently the estimation error e_i) to drift to infinity. Examples of such behavior can be found in the adaptive control literature of linear systems (Ioannou and Fidan 2006).

In this section, we modify the weight updating laws to avoid the parameter drift phenomenon. To formulate the problem we note that by adding and subtracting $-a_i x_i + \sum_{l=1}^{k} \bar{x}_{f_i} W_{f_i}^{l\star} s_l(x,u)$, the dynamic behavior of each state of the system (2.2) can be expressed by the following differential equation:

$$\dot{x}_i = -a_i x_i + \sum_{l=1}^{k} \bar{x}_{f_i} W_{f_i}^{l\star} s_l(x,u) + \mu_i(t), \tag{2.41}$$

where the modeling error $\mu_i(t)$ is given by

$$\mu_i(t) = f_i\left(x\left(t\right), u\left(t\right)\right) + a_i x_i\left(t\right) - \sum_{l=1}^{k} \bar{x}_{f_i} W_{f_i}^{l\star} s_l\left(x\left(t\right), u\left(t\right)\right). \tag{2.42}$$

The unknown optimal weight matrix $W_{f_i}^{l\star}$ is defined as the value of the weight vector $W_{f_i}^{l}$ that minimizes the L_∞-norm difference between $f_i\left(x,u\right) + a_i x_i$ and $\sum_{l=1}^{k} \bar{x}_{f_i} W_{f_i}^{l} s_l(x,u)$ for all $(x,u) \in \Psi \subset R^{n+m}$, subject to the constraint that $\left|x_{f_i} \cdot W_{f_i}^{l}\right| \leq \rho_l$, where ρ_l is a large design constant.

The region Ψ denotes the smallest compact subset of R^{n+m} that includes all the values that (x,u) can take, i.e., $\left(x\left(t\right), u\left(t\right)\right) \in \Psi$ for all $t \geq 0$. Since by assumption $u(t)$ is uniformly bounded and the dynamical system to be identified is bounded input bounded output (BIBO) stable, the existence of such Ψ is ensured.

In particular, for $i = 1, 2, \ldots, n$, the optimal weight vector $W_{f_i}^{l\star}$ is defined as

$$W_{f_i}^{l\star} := \arg \min_{\left|\bar{x}_{f_i} W_{f_i}^{l}\right| \leq \rho_l} \left\{ \sup_{(x,u)\in\Psi} \left| f_i(x,u) + a_i x_i - \sum_{l=1}^{k} \bar{x}_{f_i} W_{f_i}^{l} s_l(x,u) \right| \right\}.$$

The formulation developed above follows the methodology of Kosmatopoulos et al. (1995) closely. Using this formulation, we now have a system of the form (2.41)

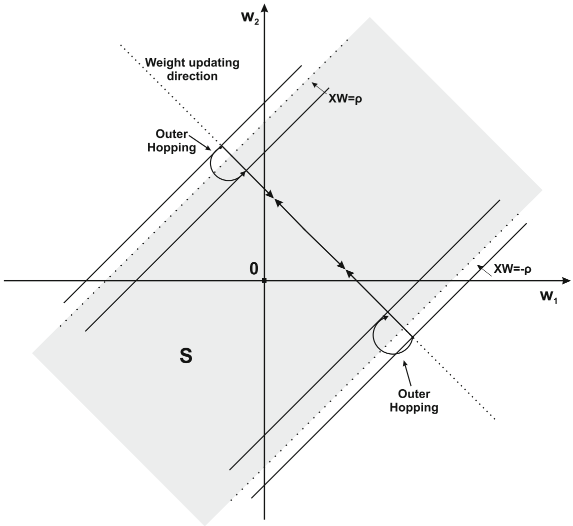

Fig. 2.3 Pictorial representation of outer parameter hopping during the identification procedure

instead of (2.23). It is noted that since $x(t)$ and $u(t)$ are bounded, the modeling error $\mu_i(t)$ is also bounded, i.e., $\sup_{t\geq 0}|\mu_i(t)| \leq \bar{\mu}_i$ for some finite constant $\bar{\mu}_i$.

Now, it is also of practical use to ensure that $\sum_{l=1}^{k} \bar{x}_{f_i} W_{f_i}^{l} s_l(x,u)$ does not approach even temporarily infinity because in this case the method may become algorithmically unstable. To avoid this situation we have to ensure that $\left|\bar{x}_{f_i}\cdot W_{f_i}^{l}\right| < \rho_l$, with ρ_l being a design parameter determining an external limit for $\bar{x}_{f_i}\cdot W_{f_i}^{l}$. We note that, since $\bar{x}_{f_i}$ and $W_{f_i}^{l}$ are row and column vectors, respectively, and since $\bar{x}_{f_i}$ has constant values, their product is linear in respect to the elements of $W_{f_i}^{l}$ and $\bar{x}_{f_i}\cdot W_{f_i}^{l}$ can describe a hyperplane. In the sequel, we consider the forbidden hyperplanes being defined by the equation $\left|\bar{x}_{f_i}\cdot W_{f_i}^{l}\right| = \rho_l$. When the weight vector reaches one of the forbidden hyperplanes $\left|\bar{x}_{f_i}\cdot W_{f_i}^{l}\right| = \rho_l$ and the direction of updating is toward the forbidden hyperplane, a *parameter hopping* is introduced that moves the weights inside the restricted area. A more analytical and general description of the novel method of *parameter hopping* is presented in Chap. 3 Sect. 3.2.2.

The above procedure is depicted in Fig. 2.3, in a simplified two-dimensional representation. The magnitude of hopping is $-\frac{\kappa_l P_i^l\left(\bar{x}_{f_i} W_{f_i}^{l}(\bar{x}_{f_i})^T\right)}{\mathrm{tr}\{(\bar{x}_{f_i})^T\bar{x}_{f_i}\}}$ being determined by following the vectorial proof given in Chap. 3 (where $b = W_{f_i}^{l}$ and our plane is described by equation $\bar{x}_{f_i}\cdot W_{f_i}^{l} = \rho_l$, with $\bar{x}_{f_i}$ the normal to it), with κ_l a positive constant (such as, $0 < \kappa_l P_i^l < 1$) decided appropriately from the designer and P_i^l is the l-th element of the gain matrix P_i.

In what follows, we develop a robust learning algorithm based on the F-RHONNs identifier employing the parameter hopping. Hence, the identifier is chosen as in

(2.24) where $W^l_{f_i}$ is the estimate of the unknown optimal weight matrix $W^{l\star}_{f_i}$. Using (2.24), (2.41), the state error $e_i = \hat{x}_i - x_i$ satisfies:

$$\dot{e}_i = -a_i e_i + \bar{x}_{f_i} \tilde{W}_{f_i} s_f(x,u) - \mu_i(t), \tag{2.43}$$

or in a more detailed form:

$$\dot{e}_i = -a_i e_i + \sum_{l=1}^{k} \bar{x}_{f_i} \tilde{W}^l_{f_i} s_l - \mu_i\,(t)\,. \tag{2.44}$$

Owing to the presence of the modeling error $\mu_i(t)$, the learning law given by (2.26) is modified by performing *parameter hopping*, when $\bar{x}_{f_i} \cdot W^l_{f_i}$ reaches the outer forbidden planes as depicted in Fig. 2.3. $\bar{x}_{f_i} \cdot W^l_{f_i}$ is confined in space $S = \left\{\bar{x}_{f_i} \cdot W^l_{f_i} : \left|\bar{x}_{f_i} \cdot W^l_{f_i}\right| \le \rho_l\right\}$, lying between these hyperplanes. The weight updating law for $W^l_{f_i}$ can now be expressed as

$$\dot{W}^l_{f_i} = -\left(\bar{x}_{f_i}\right)^T e_i s_l P^l_i - \frac{\sigma_l \kappa_l P^l_i \left(\bar{x}_{f_i} W^l_{f_i} \left(\bar{x}_{f_i}\right)^T\right)}{\operatorname{tr}\left\{\left(\bar{x}_{f_i}\right)^T \bar{x}_{f_i}\right\}}, \tag{2.45}$$

with

$$\sigma_l = \begin{cases} 0 & \text{if } \bar{x}_{f_i} W^l_{f_i} = \pm\rho_l \\ & \text{and } \bar{x}_{f_i} W^l_{f_i} <> 0 \\ 1 & \text{otherwise} \end{cases}. \tag{2.46}$$

In the current notation, the "$\pm$" symbol has a one to one correspondence with the "$<>$" one, meaning that "$+$" case corresponds to "$<$" case and "$-$" case corresponds to "$>$" case.

The above weight adjustment law is the same as (2.26) if $\bar{x}_{f_i} W^l_{f_i}$ belongs to a hypersphere of radius ρ_l. If initially $\bar{x}_{f_i} W^l_{f_i}(0)$ belongs to this hypersphere, one strategy that can be followed is to apply a "hopping" to the weight updating equation whenever a vector is approaching the forbidden outer hyperplane and is directed toward it. The "hopping" could send the weight back to the desired hyperspace allowing thus the algorithm to search the entire space for a better weight solution. Thus, in the case that the weights leave this hypersphere, the weight adjustment law is modified by the addition of a *hopping* term $-\frac{\kappa_l P^l_i \left(\bar{x}_{f_i} W^l_{f_i} \left(\bar{x}_{f_i}\right)^T\right)}{\operatorname{tr}\left\{\left(\bar{x}_{f_i}\right)^T \bar{x}_{f_i}\right\}}$, whose objective is to prevent the weight values from drifting to infinity. This modification appeared first in Boutalis et al. (2009). As it is explained in Chap. 3 (Remark 3.2) the weight hopping does not affect the existence of solutions of the dynamic equations of the model, so that Lyapunov stability arguments can be safely applied.

Now, we are ready to state the following theorem:

Theorem 2.4 *Consider the F-RHONNs model given by* (2.24) *whose weights are adjusted according to* (2.45), (2.46). *Then for* $i = 1, \ldots, n$ and $l = 1, \ldots, k$ *the following properties are guaranteed:*

1. $e_i, W_{f_i}^l, \dot{W}_{f_i}^l \in L_\infty$,
2. there exist constants r, s such that :
   $\int_0^t |e_i(t)|^2 \, \mathrm{d}\tau \leq r + s \int_0^t |\mu_i(t)|^2 \, \mathrm{d}\tau$.

Proof (1) Consider the Lyapunov candidate function:

$$V = \frac{1}{2}\sum_{i=1}^{n}\left(\left|\bar{x}_{f_i}\right|^2 e_i^2\right) + \frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{k}\left[\left(\bar{x}_{f_i}\tilde{W}_{f_i}^l\right)^T\left(P_i^l\right)^{-1}\left(\bar{x}_{f_i}\tilde{W}_{f_i}^l\right)\right]. \quad (2.47)$$

Taking the time derivatives of the Lyapunov function candidate (2.47) and taking into account (2.44), (2.45) we obtain

$$\begin{aligned}
\dot{V} &= \sum_{i=1}^{n}\left(\left|\bar{x}_{f_i}\right|^2 e_i\dot{e}_i\right) + \sum_{i=1}^{n}\sum_{l=1}^{k}\left[\left(\bar{x}_{f_i}\dot{W}_{f_i}^l\right)^T\left(P_i^l\right)^{-1}\left(\bar{x}_{f_i}\tilde{W}_{f_i}^l\right)\right]\\
&= -\sum_{i=1}^{n}\left(a_i\left|\bar{x}_{f_i}\right|^2 e_i^2 + \left|\bar{x}_{f_i}\right|^2 e_i\mu_i\right)\\
&\quad + \left|\bar{x}_{f_i}\right|^2\sum_{i=1}^{n}\left(\sum_{l=1}^{k}\bar{x}_{f_i}e_i\tilde{W}_{f_i}^l s_l - \sum_{l=1}^{k}\bar{x}_{f_i}e_i\tilde{W}_{f_i}^l s_l\right)\\
&\quad + \sum_{i=1}^{n}\frac{\left|\bar{x}_{f_i}\right|^2\sum_{l=1}^{k}\sigma_l\kappa_l\left(\bar{x}_{f_i}W_{f_i}^l\right)\left(\bar{x}_{f_i}\tilde{W}_{f_i}^l\right)}{\left|\bar{x}_{f_i}\right|^2}\\
&= -\left|\bar{x}_{f_i}\right|^2\sum_{i=1}^{n}\left(a_i\left|e_i\right|^2 + e_i\mu_i\right) - \sum_{i=1}^{n}\sum_{l=1}^{k}\left[\sigma_l\kappa_l\left(\bar{x}_{f_i}W_{f_i}^l\right)\left(\bar{x}_{f_i}\tilde{W}_{f_i}^l\right)\right].
\end{aligned} \quad (2.48)$$

Since $\tilde{W}_f = W_f - W_f^*$, we have that

$$\begin{aligned}
\left(\bar{x}_{f_i}W_{f_i}^l\right)\left(\bar{x}_{f_i}\tilde{W}_{f_i}^l\right) &= \bar{x}_{f_i}\left(\tilde{W}_{f_i}^l + W_{f_i}^{l*}\right)\left(\bar{x}_{f_i}\tilde{W}_{f_i}^l\right)\\
&= \left(\bar{x}_{f_i}\tilde{W}_{f_i}^l\right)\left(\bar{x}_{f_i}\tilde{W}_{f_i}^l\right) + \left(\bar{x}_{f_i}W_{f_i}^{l*}\right)\left(\bar{x}_{f_i}\tilde{W}_{f_i}^l\right)\\
&= \left|\bar{x}_{f_i}\tilde{W}_{f_i}^l\right|^2 + \left(\bar{x}_{f_i}W_{f_i}^{l*}\right)\left(\bar{x}_{f_i}\tilde{W}_{f_i}^l\right)\\
&= \frac{1}{2}\left|\bar{x}_{f_i}\tilde{W}_{f_i}^l\right|^2 + \frac{1}{2}\left[\left|\bar{x}_{f_i}\tilde{W}_{f_i}^l\right|^2 + 2\left(\bar{x}_{f_i}W_{f_i}^{l*}\right)\left(\bar{x}_{f_i}\tilde{W}_{f_i}^l\right)\right]\\
&= \frac{1}{2}\left|\bar{x}_{f_i}\tilde{W}_{f_i}^l\right|^2 + \frac{1}{2}\left|\bar{x}_{f_i}W_{f_i}^l\right|^2 - \frac{1}{2}\left|\bar{x}_{f_i}W_{f_i}^{l*}\right|^2.
\end{aligned} \quad (2.49)$$

Since, by definition, $\left|\bar{x}_{f_i} \cdot W_{f_i}^{l*}\right| \leq \rho_l$ and $\left|\bar{x}_{f_i} \cdot W_{f_i}^{l}\right| > \rho_l$ for $\sigma_l = 1$, we have that:

$$\sum_{i=1}^{n}\sum_{l=1}^{k}\left[\frac{\sigma_l \kappa_l}{2}\left(\left|\bar{x}_{f_i} W_{f_i}^{l}\right|^2 - \left|\bar{x}_{f_i} W_{f_i}^{l*}\right|^2\right)\right] \geq 0,$$

therefore (2.48) becomes:

$$\begin{aligned}
\dot{V} &\leq \sum_{i=1}^{n}\left[-\left|\bar{x}_{f_i}\right|^2\left(a_i\left|e_i\right|^2 + e_i\mu_i\right)\right] - \sum_{i=1}^{n}\left[\sum_{l=1}^{k}\frac{\sigma_l\kappa_l}{2}\left|\bar{x}_{f_i}\tilde{W}_{f_i}^{l}\right|^2\right] \qquad (2.50)\\
&\leq \sum_{i=1}^{n}\left[-\frac{\left|\bar{x}_{f_i}\right|^2}{2}a_i\left|e_i\right|^2 - \sum_{l=1}^{k}\frac{\kappa_l}{2}\left|\bar{x}_{f_i}\tilde{W}_{f_i}^{l}\right|^2\right]\\
&\quad + \sum_{i=1}^{n}\left[(1-\sigma_l)\sum_{l=1}^{k}\frac{\kappa_l}{2}\left|\bar{x}_{f_i}\tilde{W}_{f_i}^{l}\right|^2\right]\\
&\quad - \sum_{i=1}^{n}\left[\frac{\left|\bar{x}_{f_i}\right|^2}{2}\left(a_i\left|e_i\right|^2 + 2\mu_i e_i\right)\right]\\
&\leq -\sum_{i=1}^{n}\frac{\left|\bar{x}_{f_i}\right|^2}{2}a_i\left|e_i\right|^2 - \sum_{i=1}^{n}\sum_{l=1}^{k}\frac{\kappa_l}{\lambda_{\max}\left(P_i^{-1}\right)}V\left(e_i,\tilde{W}_{f_i}^{l}\right)\\
&\quad + \sum_{i=1}^{n}\left[(1-\sigma_l)\sum_{l=1}^{k}\frac{\kappa_l}{2}\left|\bar{x}_{f_i}\tilde{W}_{f_i}^{l}\right|^2 + \mu_i^2\right], \qquad (2.51)
\end{aligned}$$

where $\lambda_{\max}\left(P_i^{-1}\right) > 0$ denotes the maximum eigenvalue of P_i^{-1}. Since,

$$\sum_{i=1}^{n}\left[(1-\sigma_l)\sum_{l=1}^{k}\frac{\kappa_l}{2}\left|\bar{x}_{f_i}\tilde{W}_{f_i}^{l}\right|^2\right] = \begin{cases}\sum_{i=1}^{n}\sum_{l=1}^{k}\frac{\kappa_l}{2}\left|\bar{x}_{f_i}\tilde{W}_{f_i}^{l}\right|^2 & \text{if } \bar{x}_{f_i}W_{f_i}^{l} = \pm\rho_l \\ & \text{and } \bar{x}_{f_i}\dot{W}_{f_i}^{l} >< 0 \\ 0 & \text{otherwise}\end{cases},$$

we obtain $\sum_{i=1}^{n}\left[(1-\sigma_l)\sum_{l=1}^{k}\frac{\kappa_l}{2}\left|\bar{x}_{f_i}\tilde{W}_{f_i}^{l}\right|^2\right] \leq \sum_{l=1}^{k}\kappa_l\rho_l^2$. Hence (2.51) can be written in the form

$$\dot{V} \leq -d - bV + c,$$

where $d = \sum_{i=1}^{n}\frac{a_i\left|\bar{x}_{f_i}\right|^2}{2}\left|e_i\right|^2, b = \sum_{i=1}^{n}\sum_{l=1}^{k}\frac{\kappa_l}{\lambda_{\max}\left(P_i^{-1}\right)}$ and $c = \sum_{i=1}^{n}\left(\sum_{l=1}^{k}\kappa_l\rho_l^2 + \bar{\mu}_i^2\right)$ with $\bar{\mu}_i$ an upper bound for μ_i. Therefore, when $V\left(e_i,\tilde{W}_{f_i}^{l}\right) \geq V_0 = \frac{c}{b}$, we have $\dot{V} \leq 0$, which in the sequel implies that V is bounded. Hence, $\tilde{W}_{f_i}^{l} \in L_\infty$ and $\mu_i \in L_\infty$ thus from (2.44) we result to $e_i \in L_\infty$. Furthermore, using (2.45) and the fact that $\bar{x}_{f_i}, e_i, s_l, P_i^l, W_{f_i}^l \in L_\infty$, we obtain $\dot{W}_{f_i}^l \in L_\infty$.

(2) Continuing the analysis, we note that by deleting the second square term in (2.50) we obtain

$$\begin{aligned}\dot{V} &\leq -\sum_{i=1}^{n} \left|\bar{x}_{f_i}\right|^2 \left(a_i \left|e_i\right|^2 + e_i \mu_i\right) \\ &\leq \sum_{i=1}^{n} \left[-\frac{\left|\bar{x}_{f_i}\right|^2}{2} a_i \left|e_i\right|^2 + \mu_i^2\right]. \end{aligned} \tag{2.52}$$

Integrating both sides of (2.52) yields

$$\begin{aligned} V(t) - V(0) &\leq \sum_{i=1}^{n} \left(-\frac{\left|\bar{x}_{f_i}\right|^2 a_i}{2} \int_0^t e_i^2(\tau)\, \mathrm{d}\tau\right) + \sum_{i=1}^{n} \left(\int_0^t \mu_i^2(\tau)\, \mathrm{d}\tau\right) \\ &\leq -\frac{\left|\bar{x}_{f_i}\right|^2 a_i}{2} \int_0^t \left|e_i(\tau)\right|^2 \mathrm{d}\tau + \int_0^t \left|\mu_i(\tau)\right|^2 \mathrm{d}\tau. \end{aligned}$$

Therefore,

$$\begin{aligned} \int_0^t \left|e_i(\tau)\right|^2 \mathrm{d}\tau &\leq \frac{2}{\left|\bar{x}_{f_i}\right|^2 a_i} \left[V(t) - V(0)\right] + \frac{1}{\left|\bar{x}_{f_i}\right|^2 a_i} \int_0^t \left|\mu_i(\tau)\right|^2 \mathrm{d}\tau \\ &\leq r + s \int_0^t \left|\mu_i(\tau)\right|^2 \mathrm{d}\tau \end{aligned}$$

where $r := \left(\frac{2}{\left|\bar{x}_{f_i}\right|^2 a_i}\right) \sup_{t \geq 0} \left[V(t) - V(0)\right]$ and $s := \frac{1}{\left|\bar{x}_{f_i}\right|^2 a_i}$. This proves the second part of Theorem 2.4.

One can observe that if the modeling error is removed, i.e., $\mu_i = 0$, then the *parameter hopping* will not guarantee the ideal properties of the adaptive law since it introduces a disturbance of the order of the design constant κ_l. This is one of the main drawbacks of *parameter hopping* that is removed with the next remark. One of the advantages of *parameter hopping* is that no assumption about bounds or location of the unknown $W_{f_i}^{l*}$ is made.

Remark 2.2 The drawback of parameter hopping is eliminated using a switching term κ_s, which activates the small feedback term around the integrator when the magnitude of $\bar{x}_{f_i} W^l_{f_i}$ exceeds a certain value ρ_0. The assumptions we make in this case are that $\left|\bar{x}_{f_i} W^{l*}_{f_i}\right| \le \rho_0$ and ρ_0 is known. Since ρ_0 is arbitrary, it can be chosen to be high enough to guarantee $\left|\bar{x}_{f_i} W^{l*}_{f_i}\right| \le \rho_0$ in the case where limited or no information is available about the location of $W^{l*}_{f_i}$. The switching parameter constant is given by

$$\kappa_s(t) = \begin{cases} 0 & \text{if } \left|\bar{x}_{f_i} W^l_{f_i}\right| \le \rho_0 \\ \left(\dfrac{\left|\bar{x}_{f_i} W^l_{f_i}\right|}{\rho_0} - 1\right)^{q_0} \kappa_0 & \text{if } \rho_0 < \left|\bar{x}_{f_i} W^l_{f_i}\right| \le 2\rho_0 \\ \kappa_0 & \text{if } \left|\bar{x}_{f_i} W^l_{f_i}\right| > 2\rho_0 \end{cases}, \tag{2.53}$$

where q_0 is any finite integer and ρ_0, κ_0 are design constants satisfying $\rho_0 > \left|\bar{x}_{f_i} W^{l*}_{f_i}\right|$ and $\kappa_0 > 0$. The switching from 0 to κ_0 is continuous to guarantee the existence and uniqueness of solution of the weight updating differential equation.

The gradient algorithm with the switching parameter constant κ_s given by (2.53) is described as

$$\dot{W}^l_{f_i} = -\left(\bar{x}_{f_i}\right)^T p_i e_i s_l - \frac{\kappa_s P^l_i \left(\bar{x}_{f_i} W^l_{f_i} \left(\bar{x}_{f_i}\right)^T\right)}{\operatorname{tr}\left\{\left(\bar{x}_{f_i}\right)^T \bar{x}_{f_i}\right\}}. \tag{2.54}$$

As shown in Ioannou and Fidan (2006), the adaptive law (2.53), (2.54) retains all the properties of (2.45), (2.46) and, in addition, guarantees the existence of a unique solution, in the sense of Caratheodory (Hale 1969). The issue of existence and uniqueness of solutions in adaptive systems is treated in detail in Polycarpou and Ioannou (1993) and Ioannou and Fidan (2006).

2.5 Simulation Results

To demonstrate the performance of the proposed identification scheme, we present simulations testing its approximation abilities. First, we present the identification of a system having identical model structure with the NF model, where we investigate the weight convergence to their optimal values. Next, we compare the proposed F-RHONNs scheme against the simple RHONNs in approximating the angular positions of joints 1, 2 of a two link robotic manipulator.

Table 2.1 Parameters of F-RHONNs simulations

Parameters	F-RHONNs values
Recursion constant	$a = 0.5$
Sigmoidal	$\alpha = 4$
	$\beta = 0.3$
	$\gamma = -1$
High-order terms	First order
	$s_f = (s(x), s(u))$
Fuzzy centers	$\bar{x}_f = \begin{bmatrix}1.5 & 3 & 4\end{bmatrix}$
Learning rate	$P = 0.05$
Initial weights	$W_f = [0]$
Optimal weights, scenario 1	$W_f^\star = \begin{bmatrix} -0.0908 & 0.2809 \\ -0.1816 & 0.5618 \\ -0.2421 & 0.7491 \end{bmatrix}$
Optimal weights, scenario 2	$W_f^\star = \begin{bmatrix} 0.818 & 0.298 \\ 2.392 & 0.878 \\ 3.408 & 1.252 \end{bmatrix}$

2.5.1 Parameter Identification in a Known Model Structure

In order to test the ability of the presented modeling and identification approach in regard to the convergence of weights to their optimal values during the identification procedure, we create a known NF structure of the form:

$$\dot{x} = -ax + X_f W_f^\star s_f, \tag{2.55}$$

and an F-RHONNs approximator:

$$\dot{\hat{x}} = -a\hat{x} + X_f W_f s_f, \tag{2.56}$$

with the same parameter values as shown in Table 2.1, except the weights that are adjusted according to the simple gradient descent learning law (2.26).

In the sequel, we consider two different scenarios.

Scenario 1: In this scenario, the train phase consists of 2,500 epochs where each one holds for 3 s, with sampling time 10^{-2} s and control input randomly selected values in the interval $[-1, 1]$. We recall that both x, u participate in the model through s_f. Every epoch trains our algorithm with the same data coming from the known NF structure (2.55). The initial values for the NF structure as well as for the F-RHONNs identifier are $\begin{bmatrix}x(0)\end{bmatrix} = \begin{bmatrix}\hat{x}(0)\end{bmatrix} = \begin{bmatrix}0.2\end{bmatrix}$.

After each epoch, the weight evolution is plotted together with their optimal values, as shown in Fig. 2.4. One can observe the fine behavior of the identifier that finally converges almost perfectly to the known NF structure.

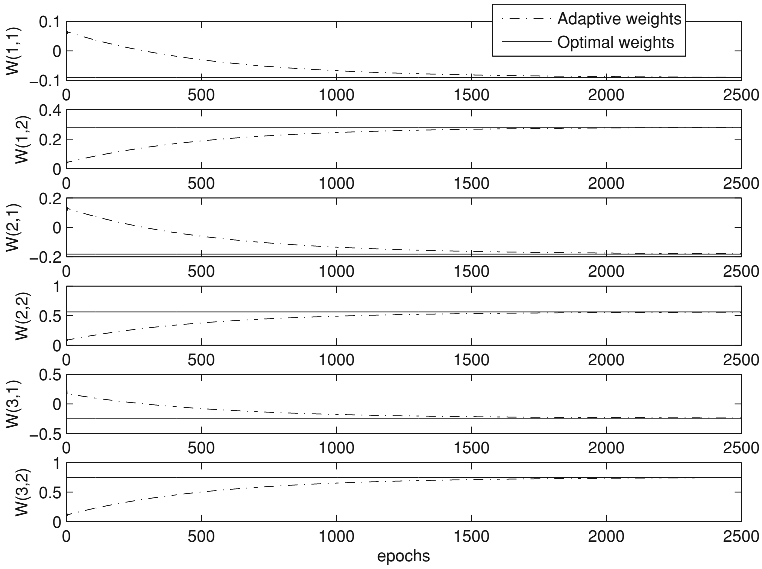

Fig. 2.4 Weight convergence to their optimal values during the training phase by using gradient descent learning law

Scenario 2: In this scenario, we repeat the same procedure with the same parameters as before except the optimal weight matrix target and the learning law, which is now the pure LS with Eqs. (2.30), (2.31). As we can see in Fig. 2.5, the weights converge faster to their optimal values when we use adaptive learning rate.

2.5.2 Two Link Robot Arm

The planar two-link revolute arm shown in Fig. 2.6 is used extensively in the literature for easy simulation of robotic controllers.

Its dynamics are given as (Lewis et al. 1993):

$$\begin{aligned}\dot{x}_1 &= x_3\\ \dot{x}_2 &= x_4\\ [\dot{x}_3\ \dot{x}_4]^T &= -M^{-1}\left(N+\tau\right),\end{aligned} \tag{2.57}$$

where $x_1 = \theta_1$ is the angular position of joint 1, $x_2 = \theta_2$ is the angular position of joint 2, $x_3 = \dot{\theta}_1$ is the angular velocity of joint 1, and $x_4 = \dot{\theta}_2$ is the angular velocity of joint 2. Also, the matrices M, N have the following form:

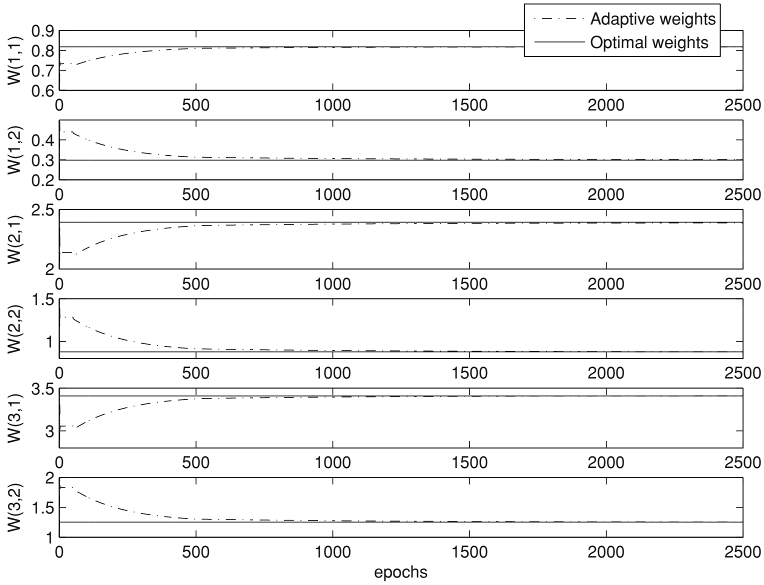

Fig. 2.5 Weight convergence to their optimal values during the training phase using pure LS algorithm

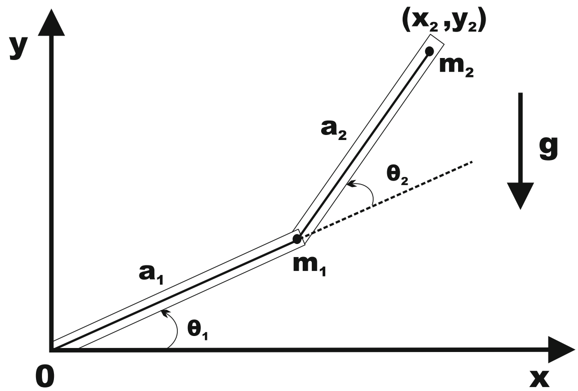

Fig. 2.6 Two-link planar elbow arm

$$M = \begin{bmatrix} M(1,1) & M(1,2) \\ M(2,1) & M(2,2) \end{bmatrix},$$

with

$$M(1,1) = (m_1 + m_2)\, a_1^2 + m_2 a_2^2 + 2 m_2 a_1 a_2 \cos(x_2),$$
$$M(1,2) = m_2 a_2^2 + m_2 a_1 a_2 \cos(x_2),$$

$$M(2,1) = m_2 a_2^2 + m_2 a_1 a_2 \cos(x_2),$$
$$M(2,2) = m_2 a_2^2,$$

and

$$N = \begin{bmatrix} N(1,1) \\ N(2,1) \end{bmatrix},$$

with

$$\begin{aligned} N(1,1) &= -m_2 a_1 a_2 \left(2x_3 x_4 + x_4^2\right) \sin(x_2) \\ &\quad + (m_1 + m_2)\, g a_1 \cos(x_1) + m_2 g a_2 \cos(x_1 + x_2), \\ N(2,1) &= m_2 a_1 a_2 x_3^2 \sin(x_2) + m_2 g a_2 \cos(x_1 + x_2). \end{aligned}$$

We took the arm parameters as $a_1 = a_2 = 1\,\text{m}$, $m_1 = m_2 = 1\,\text{kg}$. In the simulations carried out, the aim is not the control of the system but only to test the identification performance of the proposed scheme. Therefore, we use Eqs. (2.57) as a means for deriving training data. These data help the designer to choose the appropriate fuzzy centers with the help of one of the well-known clustering methods such as fuzzy c-mean clustering. After that, the simulations take place in two different phases, which are presented below.

2.5.2.1 Training Phase

In this phase, our main purpose is to calculate the optimal weight matrix $W_f^\star$ after training our F-RHONNs model with the gradient descent algorithm given by (2.26), in conjunction with the switching parameter hopping condition given by Eqs. (2.53), (2.54). The training phase consists of 1,000 epochs where each one lasts for 2 s, with sampling time 10^{-3} s. Every epoch trains our algorithm with the same data coming from the real system with the same inputs randomly selected in the interval $[-1, 1]$. The initial values for the real system states are:

$$\begin{bmatrix} x_1(0) & x_2(0) & x_3(0) & x_4(0) \end{bmatrix} = \begin{bmatrix} -0.8 & -0.4 & 0 & 0 \end{bmatrix},$$

while for the F-RHONNs algorithm are

$$\begin{bmatrix} \hat{x}_1(0) & \hat{x}_2(0) & \hat{x}_3(0) & \hat{x}_4(0) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}.$$

The weights extracted from every epoch becomes the initial values for the weights during the next run (epoch).

In Tables 2.2 and 2.3 we present the parameter values that have been used for the simulations of F-RHONNs and RHONN approaches (as described in Sect. 1.2 and Rovithakis and Christodoulou (2000)), respectively.

Table 2.2 Parameters of F-RHONNs algorithm and updating law for robot two link simulations

Parameters	F-RHONNs values
Recursion constant	$a_1 = 7.7$
	$a_2 = 4.04$
Sigmoidal	$\alpha = 6.08$
	$\beta = 0.3$
	$\gamma = -7.696$
High-order terms	First order
	$s_f = (s(x_1), s(x_2), s(x_3), s(x_4), s(u_1), s(u_2))$
Fuzzy centers	$\bar{x}_{f_1} = \begin{bmatrix} -2.1 & -1.7 & -1.2 & -0.5 \end{bmatrix}$
	$\bar{x}_{f_2} = \begin{bmatrix} -1.8 & -1.4 & 1.5 & 1.8 \end{bmatrix}$
Learning rate	$P_1 = 0.00428$
	$P_2 = 0.00384$
Initial weights	$W_{f_1} = [0]$
	$W_{f_2} = [0]$
Hopping constants	$\kappa_l = 0.675$
	$\rho_l = 181$

Table 2.3 Parameters of RHONN algorithm and updating law for robot two-link simulations

Parameters	RHONN values
Recursion constant	$a_1 = 5.06$
	$a_2 = 5.19$
Sigmoidal	$\alpha = 7.93$
	$\beta = 0.25$
	$\gamma = -7.74$
High-order terms	First order
	$s_f = (s(x_1), s(x_2), s(x_3), s(x_4), s(u_1), s(u_2))$
Learning rate	$P_1 = 0.00692$
	$P_2 = 0.00766$
Initial weights	$W_{f_1} = [0]$
	$W_{f_2} = [0]$

2.5.2.2 Testing Phase

When the training stops, we proceed to test the abilities of the trained NF model using as input signal, values constraint in the interval $[-1, 1]$, which have the following form:

$$u_1(k) = \sum_{j=1}^{16} \frac{j}{196} \cdot \sin\left(\frac{\pi}{100} \cdot j \cdot k\right), \tag{2.58}$$

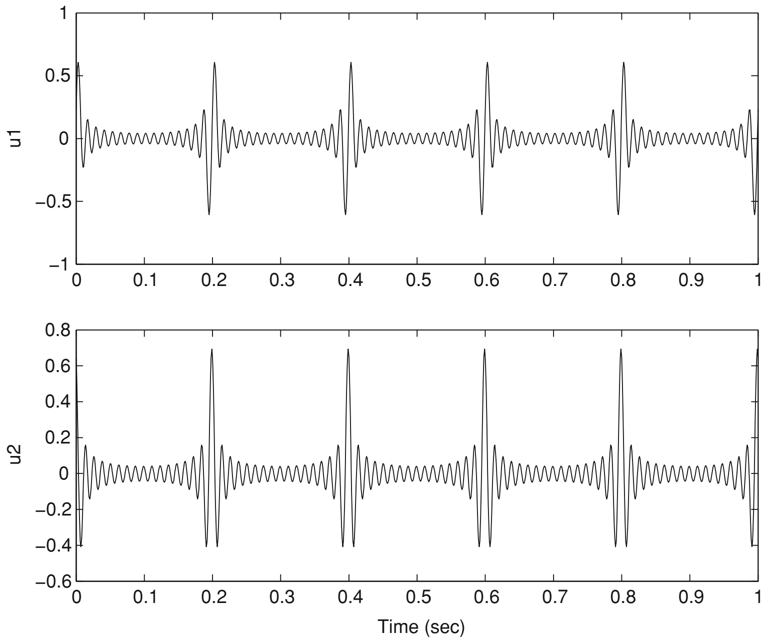

Fig. 2.7 Evolution of input signals u_1 and u_2

$$u_2(k) = \sum_{j=1}^{16} \frac{j}{196} \cdot \cos\left(\frac{\pi}{100} \cdot j \cdot k\right). \tag{2.59}$$

Figure 2.7 shows the evolution of the input signals (2.58) and (2.59), respectively.

The weight matrix derived from the training phase, which corresponds to the angular position of joint 1, 2, is given for the F-RHONNs as

$$W^{\star}_{f_{1,2}} = \begin{bmatrix} -2.0785 & -0.0284 & -0.2853 & 0.0154 & 1.2161 & 1.1566 \\ -4.9884 & -0.0680 & -0.6847 & 0.0370 & 2.9187 & 2.7759 \\ -7.0669 & -0.0964 & -0.9699 & 0.0524 & 4.1348 & 3.9325 \\ -8.7297 & -0.1190 & -1.1982 & 0.0647 & 5.1076 & 4.8578 \\ -0.1399 & -2.5484 & 0.0234 & -0.7536 & 1.5823 & 1.8583 \\ -0.1089 & -1.9821 & 0.0182 & -0.5862 & 1.2307 & 1.4453 \\ 0.1166 & 2.1236 & -0.0195 & 0.6279 & -1.3186 & -1.5485 \\ 0.1399 & 2.5484 & -0.0234 & 0.7536 & -1.5823 & -1.8583 \end{bmatrix}$$

while for the RHONN as:

$$W^{\star}_{f_{1,2}} = \begin{bmatrix} 8.0216 & -0.1592 & 1.8826 & -0.0931 & -4.7609 & -5.0891 \\ -0.7083 & 9.1152 & -0.0492 & 1.9712 & -4.9517 & -5.4681 \end{bmatrix}.$$

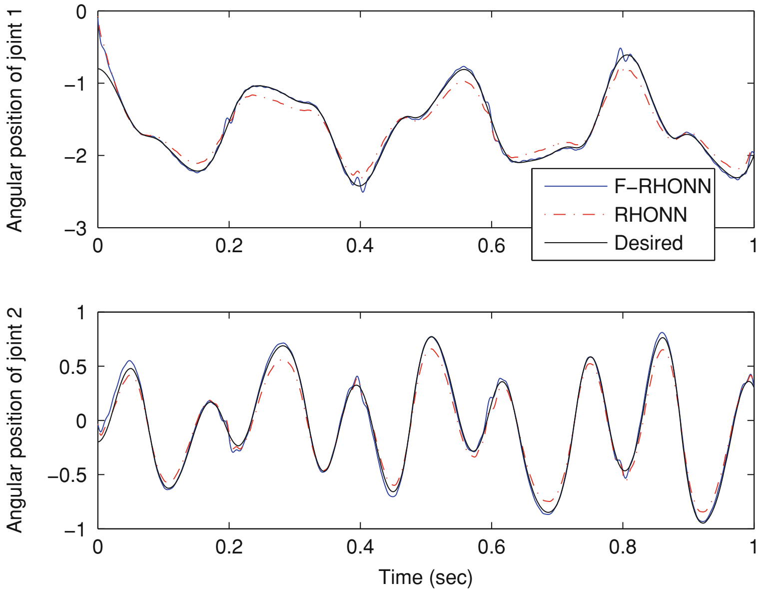

Fig. 2.8 Approximation of robot angular position 1 and 2

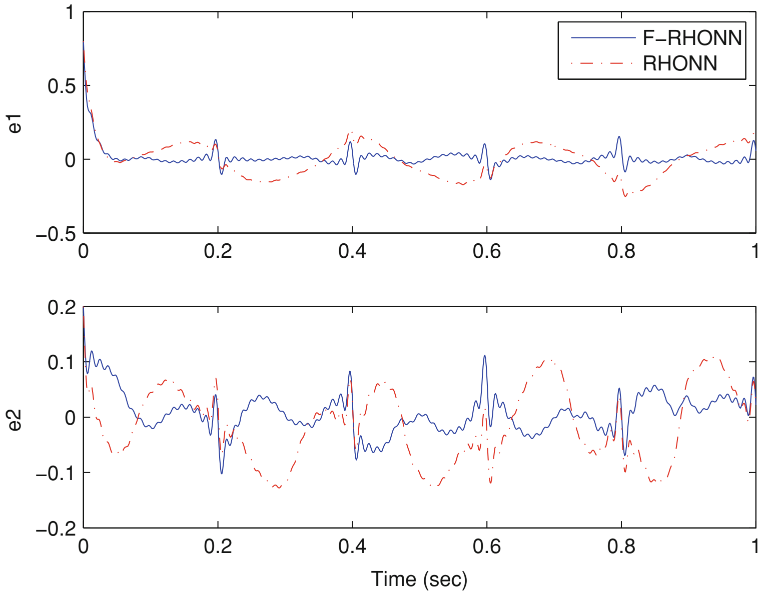

Fig. 2.9 Identification errors of robot angular positions 1 and 2 when we use different initial conditions

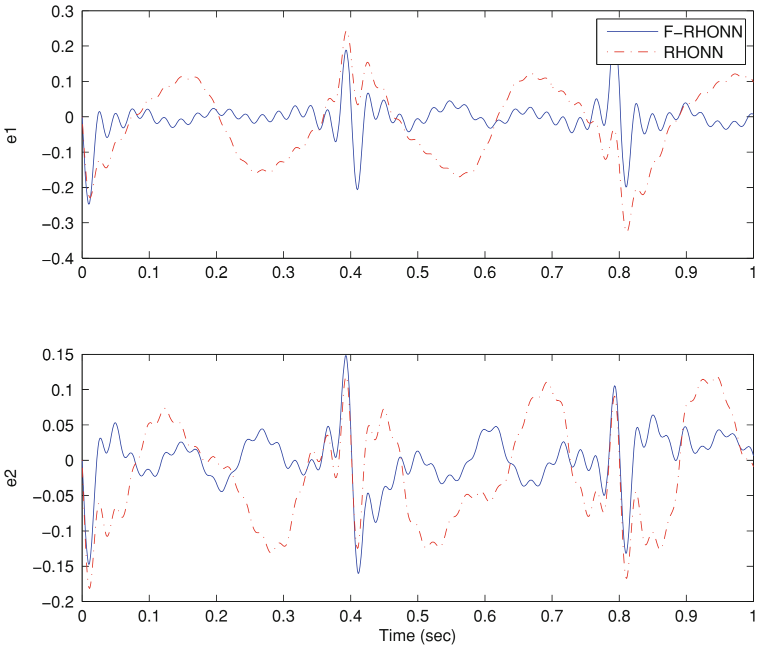

Fig. 2.10 Identification errors of robot's angular positions 1 and 2 when we use the same initial condition

Finally, we run our algorithm and RHONN approach for the above optimal weight matrices and the same control sequence with sampling time 10^{-3} s for 1 s. It is our intention to compare the approximation abilities of the proposed dynamic NF network (2.24) with RHONNs, (Rovithakis 2000) in approximating Eq. (2.57).

Figure 2.8 gives the approximation of robot states x_1, x_2 while Fig. 2.9 presents the evolution of identification errors, for RHONN and F-RHONNs models. The MSE measured as 0.0042, 0.0056 for F-RHONNs and 0.0148, 0.0194 for RHONN concerning states x_1, x_2, respectively. One can see that the dynamic NF networks are more powerful than the simple neural networks.

Once again, we test our approximator capabilities against RHONNs changing the frequency of input signals (divided by 2) and the initial values of both methods that are equal to the initial values of real system. Figure 2.10 shows that when the initial conditions of the robotic system and F-RHONNs identifier become equal the behavior of our approximator remains or becomes even better.

2.6 Summary

In this chapter, we presented a new recurrent neurofuzzy model, termed F-RHONN, for the identification of unknown nonlinear dynamical systems. It is based on multiple HONN approximators and the fuzzy partitioning of a fuzzy system output. Every HONN is specialized to work in a small region of the whole unknown system around the fuzzy output center.

In the sequel, we investigated the approximation capabilities of the proposed F-RHONNs when they are trained by different algorithms such as gradient descent and pure LS, which are proved to be Lyapunov stable. Furthermore, we examine the robustness of the training algorithms by employing a novel approach of switching parameter hopping instead of the classical σ-modification and prove once again that it is Lyapunov stable.

In the simulation results section, we demonstrated the approximation capabilities of our approach by presenting the convergence of identifier weights to their optimal values when we use a known model structure. In the sequel, the F-RHONN model was compared with simple RHONN showing the superiority of the NF model in approximating two link robot arm state variables.

References

Bishop, E. (1961). A generalization of the Stone–Weierstrass theorem. *Pacific Journal of Mathematics*, *11*, 777–783.

Boutalis, Y. S., Theodoridis, D. C., & Christodoulou, M. A. (2009). A new neuro fds definition for indirect adaptive control of unknown nonlinear systems using a method of parameter hopping. *IEEE Transactions on Neural Networks*, *20*, 609–625.

Christodoulou, M. A., Theodoridis, D. C., & Boutalis, Y. S. (2007). Building optimal fuzzy dynamical systems description based on recurrent neural network approximation. *Conference of Networked Distributed Systems for Intelligent Sensing and Control* (pp. 82–93). Greece: Kalamata.

Hale, J. (1969). *Ordinary differential equations*. New York: Wiley-Interscience.

Ioannou, P., & Fidan, B. (2006). Advances in design and control. In *Adaptive control tutorial*. Philadelphia: SIAM.

Kosmatopoulos, E. B., & Christodoulou, M. A. (1996). Recurrent neural networks for approximation of fuzzy dynamical systems. *International Journal of Intelligent Control and Systems*, *1*, 223–233.

Kosmatopoulos, E. B., Polycarpou, M. M., Christodoulou, M. A., & Ioannou, P. A. (1995). High-order neural network structures for identification of dynamical systems. *IEEE Transactions Neural Networks*, *6*, 422–431.

Lewis, F. L., Abdallah, C. T., & Dawson, D. M. (1993). *Control of robot manipulators*. New York: Macmillan.

Polycarpou, M., & Ioannou, P. (1993). On the existence and uniqness of solutions in adaptive control systems. *IEEE Transactions on Automatic Control*, *38*, 474–479.

Rovithakis, G. & Christodoulou, M. A. (2000). Adaptive control with recurrent high order neural networks (theory and industrial applications). In *Advances in industrial control*. London: Springer.

Rovithakis, G. A. (2000). Performance of a neural adaptive tracking controller for multi-input nonlinear dynamical systems in the presence of additive and multiplicative external disturbances. *IEEE Transaction SMC—Part A*, *30*, 720–730.

Stone, M. H. (1948). The generalized weierstrass approximation theorem. *Mathematics Magazine*, *21*, 167–184.

Theodoridis, D., Boutalis, Y., & Christodoulou, M. (2009). A new neuro-fuzzy dynamical system definition based on high order neural network function approximators. In *European Control Conference ECC-09*. Budapest, Hungary.

Theodoridis, D., Boutalis, Y., & Christodoulou, M. (2012). Dynamical recurrent neuro-fuzzy identification schemes employing switching parameter hopping. *International Journal of Neural Systems, 22*, 16 pages.

http://www.springer.com/978-3-319-06363-8

System Identification and Adaptive Control
Theory and Applications of the Neurofuzzy and Fuzzy Cognitive Network Models
Boutalis, Y.; Theodoridis, D.; Kottas, T.; Christodoulou, M.A.
2014, XII, 313 p. 120 illus., 56 illus. in color., Hardcover
ISBN: 978-3-319-06363-8