FUZZY INSTRUMENTAL VARIABLE ALGORITHM FOR ONLINE MULTIVARIABLE NEURAL IDENTIFICATION

Ginalber L.O. Serra
*Control and Intelligent Systems Laboratory, State University of Campinas, São Paulo-SP, Brasil
ginalber@dmcsi.fee.unicamp.br

Celso P. Bottura*
bottura@dmcsi.fee.unicamp.br

Abstract. *In this paper an algorithm for neuro-fuzzy identification of multivariable discrete-time nonlinear dynamical systems is proposed based on a decomposed form as a set of coupled multiple input and single output (MISO) Takagi-Sugeno (TS) neuro-fuzzy networks. An* on-line *scheme is formulated for modeling a nonlinear autoregressive with exogenous input (NARX) neuro-fuzzy structure from samples of a multivariable nonlinear dynamical system in a noisy environment. This approach essentially simplifies the original multivariable nonlinear plant to a nonlinear combination of multiple linear MISO subsystems. An adaptive QR factorization weighted instrumental variable (WIV) algorithm based on the numerically robust orthogonal Householder transformation is developed to modify the consequent parameters of the Takagi-Sugeno multivariable neuro-fuzzy network.*

Keywords: *Neuro-fuzzy modeling, Fuzzy systems, Neural networks, Systems identification, Recursive algorithm, Fuzzy instrumental variable.*

1. Introduction

A significant approach for modeling complex, uncertain and highly nonlinear dynamic systems is intelligent identification. The models to be identified are defined on multidimensional spaces and represent input-output mappings $f : \mathbf{x} \rightarrow \mathbf{y}$ which may be finite, infinite, discrete or continuous. These models are also used for simulations, predictions, analysis of the system's behavior, design of controllers (*model based control*), and so forth. For nonlinear dynamical systems identification purposes, Takagi-Sugeno neuro-fuzzy networks are widely investigated, since they provide good interpolation and generalization characteristics (Papadakis and Theocaris, 2002; Takagi and Sugeno, 1985). Identifying TS neuro-fuzzy networks consists of two parts: structure modeling (determining the number of rules and the input variables involved), and parameter optimization (optimizing the consequent parameters). To be applicable to real world problems, the consequent parameter optimization must be highly efficient in the noise treatment. In this paper an algorithm for neuro-fuzzy identification of multivariable discrete-time nonlinear dynamical systems is proposed based on a decomposed form as a set of coupled multiple input and single output (MISO) Takagi-Sugeno (TS) neuro-fuzzy networks. An online scheme is formulated for modeling a nonlinear autoregressive with exogenous input (NARX) neuro-fuzzy structure from samples of a multivariable nonlinear dynamical system in a noisy environment. This approach essencially simplifies the original multivariable nonlinear plant to a nonlinear combination of multiple linear MISO subsystems. An adaptive QR factorization weighted instrumental variable (WIV) algorithm by based on the numerically robust orthogonal Householder transformation is developed to modify the consequent parameters of the Takagi-Sugeno multivariable neuro-fuzzy network. By choosing proper instrument variables it provides a way to obtain consistent estimates with certain optimal properties (Bottura and Serra, 2004; Ljung, 1999).

2. Takagi-Sugeno neuro-fuzzy network

The neuro-fuzzy network is based on Takagi-Sugeno fuzzy model and membership functions of gaussian type, so that the algorithm can utilize all information contained in the training data set to calculate the consequent parameters of the network. The TS neuro-fuzzy network is composed of a fuzzy IF $<$ *antecedent* $>$ THEN $<$ *consequent* $>$ rule base that partitions the input space, so-called *universe of discourse*, into fuzzy regions described by the rule antecedents in which consequent functions are valid (Bottura and Serra, 2004). The consequent of each rule i is usually a simple functional expression $y_i = f_i(x)$. The ith TS rule has the following form:

$$R^i : \text{ IF } x_1 \text{ is } F_1^i \text{ AND } \cdots \text{ AND } x_q \text{ is } F_q^i \text{ THEN } y_i = f_i(\mathbf{x}) \,,\, i = 1, 2, \ldots, l \tag{1}$$

where l is the number of rules. The vector $\mathbf{x} \in \Re^q$ contains the premise variables, which has its own universe of discourse that is partitioned into fuzzy regions by the fuzzy sets describing the linguistic variable F_j^m. The premise variable x_j belongs to a fuzzy set m with a truth value given by a membership function $\mu_{jm} : R \rightarrow [0, 1]$ for $m = 1, 2, \ldots, s_j$ where s_j is the number of fuzzy sets for premise variable j. The truth value h_i for the complete rule i is computed using the aggregation operator, or t-norm, AND denoted by $\otimes : [0,1] \times [0,1] \rightarrow [0,1]$

$$h_i(x) = \mu_1^i(x_1) \otimes \mu_2^i(x_2) \otimes \ldots \mu_q^i(x_q) \tag{2}$$

Among the different t-norms available, in this work the algebraic product will be used, and

$$h_i(\mathbf{x}) = \prod_{j=1}^{q} \mu_j^i(x_j) \tag{3}$$

The degree of activation for rule i is then normalized as

$$\gamma_i(\mathbf{x}) = \frac{h_i(\mathbf{x})}{\sum_{r=1}^{l} h_r(\mathbf{x})} \tag{4}$$

This normalization implies that

$$\sum_{i=1}^{l} \gamma_i(\mathbf{x}) = 1 \tag{5}$$

The response of the TS neuro-fuzzy network consists in a weighted sum of the consequent functions, i.e, a convex combination of the local functions (models) f_i,

$$y = \sum_{i=1}^{l} \gamma_i(\mathbf{x}) f_i(\mathbf{x}) \tag{6}$$

The general configuration and a multivariable structure of the TS neuro-fuzzy network is shown in Fig. 1, where the layers I and II are the antecedent in (1) and the layers III and IV are (3), (4) and (6) respectively. Such model can be

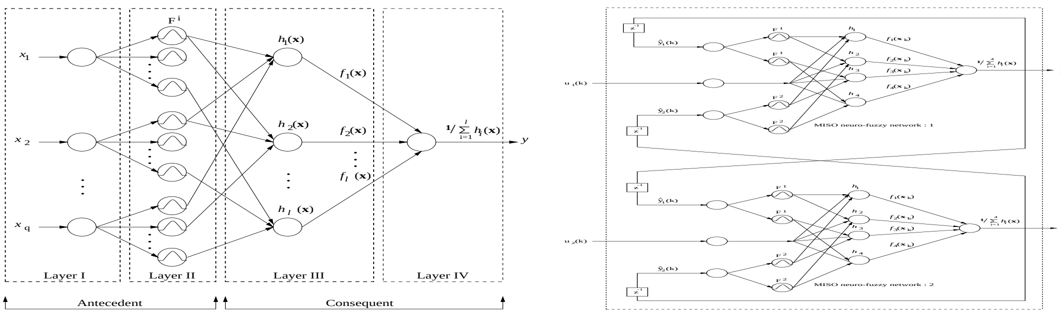

Figure 1. General configuration and multivariable structure of the Takagi-Sugeno neuro-fuzzy network

seen as a linear parameter varying (LPV) system (Bottura and Serra, 2004). In this sense, a TS neuro-fuzzy network can be considered as a mapping from the antecedent (input) space to a convex region (politope) in the space of the local submodels defined by the consequent parameters. This property simplifies the analysis of TS neuro-fuzzy networks in a robust linear system framework for identification, controllers design with desired closed loop characteristics and stability analysis.

2.1 Neuro-Fuzzy Structure Model

The nonlinear autoregressive with exogenous input (NARX) structure model will be used in this paper. This structure is used in most nonlinear identification methods such as neural networks, radial basis functions, cerebellar model articulation controller (CMAC), and also fuzzy models (Bottura and Serra, 2004; Papadakis and Thocaris, 2002; Takagi and Sugeno, 1985). The NARX structure establishes a relation between the collection of past input-output data and the predicted output

$$\hat{y}(k+1) = f(y(k), \ldots, y(k-n_y+1), u(k), \ldots, u(k-n_u+1)) \tag{7}$$

where k denotes discrete time samples, n_y and n_u are integers related to the systems'order. In terms of rules, the structure is given by

$$R^i : \text{IF } y(k) \text{ is } F_1^i \text{ AND } \cdots \text{ AND } y(k-n_y+1) \text{ is } F_{n_y}^i \text{ AND } u(k) \text{ is } G_1^i \text{ AND} \cdots \text{ AND } u(k-n_u+1) \text{ is } G_{n_u}^i$$

$$\text{THEN } \hat{y}_i(k+1) \;=\; \sum_{j=1}^{n_y} a_{i,j} y(k-j+1) + \sum_{j=1}^{n_u} b_{i,j} u(k-j+1) + c_i \tag{8}$$

where $a_{i,j}$, $b_{i,j}$ and c_i are the consequent parameters to be determined. The inference formula of the TS neuro-fuzzy network is a straightforward extension of (6) and is given by

$$\hat{y}(k+1) = \sum_{i=1}^{l} \gamma_i(\mathbf{x}_k)[\sum_{j=1}^{n_y} a_{i,j} y(k-j+1) + \sum_{j=1}^{n_u} b_{i,j} u(k-j+1) + c_i] \quad (9)$$

with

$$\mathbf{x}_k = (y(k), \ldots, y(k-n_y+1), u(k), \ldots, u(k-n_u+1)) \quad (10)$$

and $h_i(\mathbf{x}_k)$ is given as (3). This NARX network represents multiple input and single output (MISO) systems directly and multiple input and multiple output (MIMO) systems in a decomposed form as a set of coupled MISO neuro-fuzzy networks.

3. Consequent Parameters Estimate Problem

The inference formula of the TS neuro-fuzzy network in (9) can be expressed as

$$y(k+1) = \gamma_1(\mathbf{x}_k)(a_{1,1}y(k) + \ldots + a_{1,ny}y(k-n_y+1) + b_{1,1}u(k) + \ldots + b_{1,nu}u(k-n_u+1) + c_1) +$$

$$+ \gamma_2(\mathbf{x}_k)(a_{2,1}y(k) + \ldots + a_{2,ny}y(k-n_y+1) + b_{2,1}u(k) + \ldots + b_{2,nu}u(k-n_u+1) + c_2) + \ldots$$

$$+ \gamma_l(\mathbf{x}_k)(a_{l,1}y(k) + \ldots + a_{l,ny}y(k-n_y+1) + b_{l,1}u(k) + \ldots + b_{l,nu}u(k-n_u+1) + c_l) \quad (11)$$

which is linear in the consequent parameters: **a**, **b** and **c**. For a set of N input-output data pairs $\{(\mathbf{x}_k, y_k)|i = 1, 2, \ldots, N\}$ available, the following vectorial form is obtained

$$\mathbf{Y} = [\psi_1\mathbf{X}, \psi_2\mathbf{X}, \ldots, \psi_l\mathbf{X}]\theta + \mathbf{\Xi} \quad (12)$$

where $\psi_i = diag(\gamma_i(\mathbf{x}_k)) \in \Re^{N\times N}$, $\mathbf{X} = [\mathbf{y}_k, \ldots, \mathbf{y}_{k-ny+1}, \mathbf{u}_k, \ldots, \mathbf{u}_{k-nu+1}, \mathbf{1}] \in \Re^{N\times(n_y+n_u+1)}$, $\mathbf{Y} \in \Re^{N\times 1}$, $\mathbf{\Xi} \in \Re^{N\times 1}$ and $\theta \in \Re^{l(n_y+n_u+1)\times 1}$ are the normalized membership degree matrix of (4), the data matrix, the output vector, the approximation error vector and the parameters vector, respectively. A number of different techniques can be used when the variables associated with the unknown parameters are *exactly known* quantities. In practice, and in the present context, the basic relationship between the parameters is still in the form in (12), but the elements of $\mathbf{X}$ are no longer exactly known quantities and can only be observed with error so that the observed value can be expressed as

$$y_k = \chi_k^T\theta + \eta_k \quad (13)$$

where, at the k-th sampling, $\chi_k^T = [\gamma_k^1(\mathbf{x}_k+\xi_k), \ldots, \gamma_k^l(\mathbf{x}_k+\xi_k)]$ is the vector of the data with noise, $\mathbf{x}_k = [y_{k-1}, \ldots, y_{k-n_y}, u_{k-1}, \ldots, u_{k-n_u}, 1]^T$ is the vector of the data with exactly known quantities, e.g., free noise input-output data, $\gamma_k^i |^{i=1,2,\ldots,l}$ is the normalizad activation degree associated to $(\mathbf{x}_k + \xi_k)$, ξ_k is a vector of noise associated with the observation of $\mathbf{x}_k$, and η_k is a disturbance noise.

The normal equations are formulated as

$$[\sum_{j=1}^{k} \chi_j\chi_j^T]\hat{\theta}_k = \sum_{j=1}^{k} \chi_j y_j \quad (14)$$

and multiplying by $\frac{1}{k}$ gives

$$\{\frac{1}{k}\sum_{j=1}^{k}[\gamma_j^1(\mathbf{x}_j+\xi_j), \ldots, \gamma_j^l(\mathbf{x}_j+\xi_j)][\gamma_j^1(\mathbf{x}_j+\xi_j), \ldots, \gamma_j^l(\mathbf{x}_j+\xi_j)]^T\}\hat{\theta}_k = \frac{1}{k}\sum_{j=1}^{k}[\gamma_j^1(\mathbf{x}_j+\xi_j), \ldots, \gamma_j^l(\mathbf{x}_j+\xi_j)]y_j \quad (15)$$

Noting that $y_j = \chi_j^T\theta + \eta_j$,

$$\{\frac{1}{k}\sum_{j=1}^{j}[\gamma_j^1(\mathbf{x}_j+\xi_j), \ldots, \gamma_j^l(\mathbf{x}_j+\xi_j)][\gamma_j^1(\mathbf{x}_j+\xi_j), \ldots, \gamma_j^l(\mathbf{x}_j+\xi_j)]^T\}\hat{\theta}_k = \frac{1}{k}\sum_{j=1}^{k}[\gamma_j^1(\mathbf{x}_j+\xi_j), \ldots,$$

$$\ldots, \gamma_j^l(\mathbf{x}_j+\xi_j)][\gamma_j^1(\mathbf{x}_j+\xi_j), \ldots, \gamma_j^l(\mathbf{x}_j+\xi_j)]^T\theta + \frac{1}{k}\sum_{j=1}^{k}[\gamma_j^1(\mathbf{x}_j+\xi_j), \ldots, \gamma_j^l(\mathbf{x}_j+\xi_j)]\eta_j \quad (16)$$

and

$$\tilde{\theta}_k = [\frac{1}{k}\sum_{j=1}^{k}[\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)][\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]^T]^{-1}\frac{1}{k}\sum_{j=1}^{k}[\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]\eta_j \quad (17)$$

where $\tilde{\theta}_k = \hat{\theta}_k - \theta$ is the parameters error, $\hat{\theta}_k$ is the estimated parameters vector. Taking the probability in the limit as $k \to \infty$,

$$\text{p.lim }\tilde{\theta}_k = \text{p.lim }\{\frac{1}{k}\mathbf{C}_k^{-1}\frac{1}{k}\mathbf{b}_k\} \quad (18)$$

with

$$\mathbf{C}_k = \sum_{j=1}^{k}[\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)][\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]^T \quad (19)$$

$$\mathbf{b}_k = \sum_{j=1}^{k}[\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]\eta_j \quad (20)$$

Applying Slutsky's theorem and assuming that the elements of $\frac{1}{k}\mathbf{C}_k$ and $\frac{1}{k}\mathbf{b}_k$ converge in probability, we have

$$\text{p.lim }\tilde{\theta}_k = \text{p.lim }\frac{1}{k}\mathbf{C}_k^{-1}\text{ p.lim }\frac{1}{k}\mathbf{b}_k \quad (21)$$

Thus,

$$\text{p.lim }\frac{1}{k}\mathbf{C}_k = \text{p.lim }\frac{1}{k}\sum_{j=1}^{k}[\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)][\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]^T \quad (22)$$

Assuming $\mathbf{x}_j$ and ξ_j statistically independent,

$$\text{p.lim }\frac{1}{k}\mathbf{C}_k = \text{p.lim }\frac{1}{k}\sum_{j=1}^{k}\mathbf{x}_j\mathbf{x}_j^T[(\gamma_j^1)^2+\ldots+(\gamma_j^l)^2] + \text{p.lim }\frac{1}{k}\sum_{j=1}^{k}\xi_j\xi_j^T[(\gamma_j^1)^2+\ldots+(\gamma_j^l)^2] \quad (23)$$

with $\sum_{i=1}^{l}\gamma_j^i = 1$. Hence, the asymptotic analysis of the consequent parameters estimation of TS neuro-fuzzy network is based in a weighted sum of the covariance matrices of $\mathbf{x}$ and ξ. Similarly,

$$\text{p.lim }\frac{1}{k}\mathbf{b}_k = \text{p.lim }\frac{1}{k}\sum_{j=1}^{k}[\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]\eta_j\text{p.lim }\frac{1}{k}\mathbf{b}_k = \text{p.lim }\frac{1}{k}\sum_{j=1}^{k}[\gamma_j^1\xi_j\eta_j,\ldots,\gamma_j^l\xi_j\eta_j] \quad (24)$$

Substituting from (24) and (23) in (21), results

$$\text{p.lim }\tilde{\theta}_k = \{\text{p.lim }\frac{1}{k}\sum_{j=1}^{k}\mathbf{x}_j\mathbf{x}_j^T[(\gamma_j^1)^2+\ldots+(\gamma_j^l)^2] + \text{p.lim }\frac{1}{k}\sum_{j=1}^{k}\xi_j\xi_j^T[(\gamma_j^1)^2+\ldots+(\gamma_j^l)^2]\}^{-1}\text{p.lim }\frac{1}{k}\sum_{j=1}^{k}[\gamma_j^1\xi_j\eta_j,$$

$$\ldots,\gamma_j^l\xi_j\eta_j] \quad (25)$$

with $\sum_{i=1}^{l}\gamma_j^i = 1$. Provided that the input u_k continues to excite the process and, at the same time, the coefficients in the submodels from the consequent are not all zero, then the output y_k will exist for all k observation intervals. As a result, the covariance matrix $\mathbf{x}_j\mathbf{x}_j^T$ will also be non-singular and its inverse will exist. Thus, the only way in which the asymptotic error can be zero is for $\xi_j\eta_j$ identically zero at the limit. But, in general, ξ_j and η_j are correlated and the asymptotic error will not be zero and the least squares estimates will be asymptotically biased to an extent determined by the relative ratio of noise to signal variances (Bottura and Serra, 2004; Ljung, 1999).

4. Fuzzy Instrumental variable QR identification algorithm

To be applicable to real world problems, the consequent parameter optimization must be highly efficient in the noise treatment. To obtain consistent parameter estimates without modeling the noise, the instrumental variable (IV) method can be used. Thus, generating a vector of variables, the instrumental variables vector, which is independent of the noise inputs and correlated with data vetor $\mathbf{x}_j$ from the system is required. If this is possible, then it will be easy to see that the

choice of this vector becomes effective to remove the asymptotic bias from the consequent parameters estimates.
The least squares estimation results according to

$$\{\frac{1}{k}\sum_{j=1}^{k}[\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)][\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]^T\}\hat{\theta}_k=\frac{1}{k}\sum_{j=1}^{k}[\gamma_j^1(\mathbf{x}_j+\xi_j),$$

$$\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]\{[\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]\theta+\eta_j\} \tag{26}$$

Using the new vector of variables of the form $[\beta_j^1\mathbf{z}_j,\ldots,\beta_j^l\mathbf{z}_j]$, the *fuzzy instrumental variable* vector, the last equation can be placed as

$$\{\frac{1}{k}\sum_{j=1}^{k}[\beta_j^1\mathbf{z}_j,\ldots,\beta_j^l\mathbf{z}_j][\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]^T\}\hat{\theta}_k=\frac{1}{k}\sum_{j=1}^{k}[\beta_j^1\mathbf{z}_j,\ldots,\beta_j^l\mathbf{z}_j]\{[\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]\theta+\eta_j\} \tag{27}$$

where $\mathbf{z}_j$ is a vector with the order of $\mathbf{x}_j$, associated to the dynamic behavior of the system, and $\beta_j^i\mid^{i=1,\ldots,l}$ is the normalized degree of activation, as in (4), associated to $\mathbf{z}_j$. As a consequence, three conditions must be satisfyied:

condition 1

$$\lim_{n\to\infty}\frac{1}{k}\sum_{j=1}^{k}[\beta_j^1\mathbf{z}_j,\ldots,\beta_j^l\mathbf{z}_j]\xi_j^T=\mathbf{0} \tag{28}$$

condition 2

$$\lim_{n\to\infty}\frac{1}{k}\sum_{j=1}^{k}[\beta_j^1\mathbf{z}_j,\ldots,\beta_j^l\mathbf{z}_j]\eta_j=\mathbf{0} \tag{29}$$

condition 3

$$\lim_{n\to\infty}\frac{1}{k}\sum_{j=1}^{k}[\beta_j^1\mathbf{z}_j,\ldots,\beta_j^l\mathbf{z}_j][\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]^T=\mathbf{C}_{\mathbf{zx}}\neq 0 \tag{30}$$

The proof of these conditions are given in (Bottura and Serra, 2005). Now, (27) can be expressed in the form

$$\{\frac{1}{k}\sum_{j=1}^{k}[\beta_j^1\mathbf{z}_j,\ldots,\beta_j^l\mathbf{z}_j][\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]^T\}(\hat{\theta}_k-\theta)=\frac{1}{k}\sum_{j=1}^{k}[\beta_j^1\mathbf{z}_j,\ldots,\beta_j^l\mathbf{z}_j]\eta_j \tag{31}$$

Therefore, provided $\mathbf{C}_{\mathbf{zx}}$ is non-singular, the limit value of the parameter error, in probability, is

$$\text{p.lim}\,\tilde{\theta}=0 \tag{32}$$

and the estimates are asymptotically unbiased, as required.
An effective choice of the vector $\mathbf{z}_j$ would be the one based on the

$$\mathbf{z}_j=[y_{k-1-dl},\cdots,y_{k-n_y-dl},u_{k-1-dl},\cdots,u_{k-n_u-dl}]^T$$

where dl is the applied delay. The choice of this instrumental variable provides consistent estimates with certain optimal properties (Ljung, 1999). In a structural model situation, the IV normal equations must be formulated as

$$\sum_{j=1}^{k}[\beta_j^1\mathbf{z}_j,\ldots,\beta_j^l\mathbf{z}_j][\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]^T\hat{\theta}_k-\sum_{j=1}^{k}[\beta_j^1\mathbf{z}_j,\ldots,\beta_j^l\mathbf{z}_j]y_j=0 \tag{33}$$

so that the IV estimate is obtained as

$$\hat{\theta}_k=\{\sum_{j=1}^{k}[\beta_j^1\mathbf{z}_j,\ldots,\beta_j^l\mathbf{z}_j][\gamma_j^1(\mathbf{x}_j+\xi_j),\ldots,\gamma_j^l(\mathbf{x}_j+\xi_j)]^T\}^{-1}\sum_{j=1}^{k}[\beta_j^1\mathbf{z}_j,\ldots,\beta_j^l\mathbf{z}_j]y_j \tag{34}$$

and, in vectorial form, the problem of interest may be placed as

$$\hat{\theta}=(\mathbf{\Gamma}^T\mathbf{\Sigma})^{-1}\mathbf{\Gamma}^T\mathbf{Y} \tag{35}$$

where $\mathbf{\Gamma}^T \in \Re^{l(n_y+n_u+1)\times N}$ is the extended instrumental variable matrix with rows given by ζ_j, $\mathbf{\Sigma} \in \Re^{N\times l(n_y+n_u+1)}$ is the extended data matrix with rows given by χ_j and $\mathbf{Y} \in \Re^{N\times 1}$ is the output vector. The motivation for this work was the development of an algorithm that provides frequent estimates of the parameters by properly processing the I/O data on-line and adapts itself to variations on the parameters with time. The interest problem may be placed as

$$\mathbf{\Gamma}^T\mathbf{\Sigma}\theta = \mathbf{\Gamma}^T\mathbf{Y} \tag{36}$$

Equation (36) can be written as

$$\mathbf{\Gamma}^T\mathbf{W}\mathbf{\Sigma}\theta = \mathbf{\Gamma}^T\mathbf{W}\mathbf{Y} \tag{37}$$

where $\mathbf{W} = diag\left(\lambda^{n-1}, \lambda^{n-2}, \ldots, 1\right) \in \Re^{N\times N}$, with $0 < \lambda < 1$. The scalar λ, *forgetting factor*, is used to place less weight on past data. Developing both sides in (37), results

$$\mathbf{S}\theta = \mathbf{b} \tag{38}$$

for $\mathbf{S} = \mathbf{\Gamma}^T\mathbf{W}\mathbf{\Sigma} \in \Re^{l(n_y+n_u+1)\times l(n_y+n_u+1)}$ and $\mathbf{b} = \mathbf{\Gamma}^T\mathbf{W}\mathbf{Y} \in \Re^{l(n_y+n_u+1)\times 1}$. It is worth emphasizing that the resulting order of the $\mathbf{S}$ matrix and of the $\mathbf{b}$ vector are lower than the $\mathbf{\Sigma}$ matrix and $\mathbf{Y}$ vector orders, respectively, because $l(n_y + n_u + 1)$ is equal to the number of parameters to be estimated, implying less computational effort and, in consequence, greater speed to solve for θ. Generically, for $\mathbf{S}$ and $\mathbf{b}$ results:

$$\mathbf{S} = \begin{bmatrix} \mathbf{Z}^T\mathbf{\Psi}_1^2\mathbf{W}\mathbf{X} & \cdots & \mathbf{Z}^T\mathbf{\Psi}_1\mathbf{\Psi}_l\mathbf{W}\mathbf{X} \\ \vdots & \vdots & \vdots \\ \mathbf{Z}^T\mathbf{\Psi}_l\mathbf{\Psi}_1\mathbf{W}\mathbf{X} & \cdots & \mathbf{Z}^T\mathbf{\Psi}_l^2\mathbf{W}\mathbf{X} \end{bmatrix} \tag{39}$$

$$\mathbf{b} = \begin{bmatrix} \mathbf{Z}^T\mathbf{\Psi}_1\mathbf{W}\mathbf{Y} \\ \mathbf{Z}^T\mathbf{\Psi}_2\mathbf{W}\mathbf{Y} \\ \vdots \\ \mathbf{Z}^T\mathbf{\Psi}_l\mathbf{W}\mathbf{Y} \end{bmatrix} \tag{40}$$

Letting $\mathbf{S}_{ij} = \mathbf{Z}^T\mathbf{\Psi}_i\mathbf{W}\mathbf{\Psi}_j\mathbf{x}$ and $\mathbf{b}_i = \mathbf{Z}^T\mathbf{\Psi}_i\mathbf{W}\mathbf{Y}$, results

$$\mathbf{S}_{ij} = \begin{bmatrix} \sum_{t=1}^{n}\gamma_{ij}(\mathbf{x}_{k-1})u_{k-1}y_{k-1}\lambda^{n-k} & \cdots \\ \vdots & \vdots \\ \sum_{k=1}^{n}\gamma_{ij}(\mathbf{x}_{k-1})u_{k-p}y_{k-1}\lambda^{n-k} & \cdots \\ \sum_{k=1}^{n}\gamma_{ij}(\mathbf{x}_{k-1})y_{k-1}\lambda^{n-k} & \cdots \end{bmatrix} \tag{41}$$

$$\mathbf{b}_i = \begin{bmatrix} \sum_{k=1}^{n}\gamma_{ij}(\mathbf{x}_{k-1})u_{k-1}y_k\lambda^{n-k} \\ \vdots \\ \sum_{k=1}^{n}\gamma_{ij}(\mathbf{x}_{k-1})u_{k-p}y_k\lambda^{n-k} \\ \sum_{k=1}^{n}\gamma_{ij}(\mathbf{x}_{k-1})y_k\lambda^{n-k} \end{bmatrix} \tag{42}$$

where $\gamma_{ij}(\mathbf{x}_{k-1}) = \gamma_i(\mathbf{x}_{k-1})\gamma_j(\mathbf{x}_{k-1})|^{(i=j=1,2,\ldots,l)}$. From (41) and (42), it can be seen that the $\mathbf{S}$ matrix and the $\mathbf{b}$ vector are summations that depend of the actual and immediately former values, based on the dimension of the problem, the input and output measurements. These structures imply in generating, directly, i.e., in each sample, $\mathbf{S}$ and $\mathbf{b}$, without need of a priori batch of matricial operations, as in (37), with the advantage that the order is lower for QR factorization. An orthogonal Householder matrix has the following form

$$\mathbf{H} = \mathbf{I} - 2\frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|_2^2} \tag{43}$$

where $\mathbf{H} = \mathbf{H}^T$ and $\mathbf{H} = \mathbf{H}^{-1}$. Householder transformations are often used to annul a block of elements in matrices or vectors by appropriately selecting the Householder vector $\mathbf{v}$ in (43). If $\mathbf{a}$ is a nonzero vector and $\mathbf{e}_i$ is an unit vector with 1 in the *i*-th position, when

$$\mathbf{v} = \mathbf{a} \pm \parallel \mathbf{a} \parallel \mathbf{e}_i \tag{44}$$

then

$$\mathbf{H}\mathbf{a} = \mp \parallel \mathbf{a} \parallel \mathbf{e}_i \tag{45}$$

The vectors $\mathbf{v}$ and $\mathbf{a}$ are identical except for the *i*-th element. Thus, the problem may be placed as that of finding the solution:

$$\hat{\theta} = \text{ arg min } \parallel \mathbf{S}\theta - \mathbf{b} \parallel_2^2 \tag{46}$$

Applying the QR factorization with the Householder orthogonal transformation results

$$\hat{\theta} = \text{ arg min } \parallel \mathbf{Q}^T\mathbf{S}\theta - \mathbf{Q}^T\mathbf{b} \parallel_2^2 \tag{47}$$

and

$$\hat{\theta} = \text{ arg min } \parallel \mathbf{R}\theta - \mathbf{d} \parallel_2^2 \tag{48}$$

where $\mathbf{Q} \in \Re^{l(n_y+n_u+1)\times l(n_y+n_u+1)}$ is an orthogonal matrix, the upper triangular matrix is given by $\mathbf{R} \in \Re^{l(n_y+n_u+1)\times l(n_y+n_u+1)}$ and $\mathbf{d} \in \Re^{l(n_y+n_u+1)}$ is a resulting vector. Hence, the minimizer of (46) may be found by solving $\mathbf{R}\hat{\theta} = \mathbf{d}$ by back substitution. For automatic initial estimation, the algorithm receives an initial batch data and updating is obtained by simple acquisition of input and output data and insertion of it into summations of the matrix $\mathbf{S}$ and of the vector $\mathbf{b}$. Thus, for the *n*-th sample

$$\mathbf{S}_{ij}^{new} = \mathbf{S}_{ij} + \lambda \begin{bmatrix} \gamma_{ij}(\mathbf{x}_{n-1})u_{n-1}y_{n-1} & \cdots \\ \vdots & \vdots \\ \gamma_{ij}(\mathbf{x}_{n-1})u_{n-p}y_{n-1} & \cdots \\ \gamma_{ij}(\mathbf{x}_{n-1})y_{n-1} & \cdots \end{bmatrix} ; \mathbf{b}_i^{new} = \mathbf{b}_i + \lambda \begin{bmatrix} \gamma_{ij}(\mathbf{x}_{n-1})u_{n-1}y_n \\ \vdots \\ \gamma_{ij}(\mathbf{x}_{n-1})u_{n-p}y_n \\ \gamma_{ij}(\mathbf{x}_{n-1})y_n \end{bmatrix} \tag{49}$$

5. Neuro-fuzzy MIMO identification

The proposed algorithm is applied to the following MIMO nonlinear plant, which presents a highly nonlinear behaviour and has outputs estatistically correlated:

$$\begin{bmatrix} y_1(k+1) \\ y_2(k+1) \end{bmatrix} = \begin{bmatrix} \frac{y_1(k)}{1+y_2^2(k)} \\ \frac{y_1(k)y_2(k)}{1+y_2^2(k)} \end{bmatrix} + \begin{bmatrix} u_1(k) \\ u_2(k) \end{bmatrix} \tag{50}$$

The TS multivariable neuro-fuzzy network is composed by two coupled TS MISO neuro-fuzzy networks, in lattice, as shown in Fig. 1, of three inputs $y_1(k)$, $y_2(k)$ and $u_{ne}(k) \mid^{ne=1,2}$ and a single output $\hat{y}_{ns}(k+1) \mid^{ns=1,2}$, respectively. In this point, is important to highlight that the presented neuro-fuzzy network is able to represent a general nonlinear MIMO plant where its outputs are estatistically correlated as in (50), for instance. In particular, the advantage of this lattice structure is that it simplifies the original multivariable nonlinear plant to a nonlinear combination of multiple linear MISO subsystems as well as captures the correlation between the outputs of the multivariable nonlinear plant. The rule base, for each TS MISO neuro-fuzzy network, is of the form:

$$R^i: \text{ IF } y_1(k) \text{ is } F_{1,2}^i \text{ AND } y_2(k) \text{ is } G_{1,2}^i \text{ THEN } \hat{y}_{ns,i}(k+1) = a_{i,1}y_1(k) + a_{i,2}y_2(k) + b_{i,1}u_{ne}(k) + c_i \tag{51}$$

where $F_{1,2}^i|^{i=1,2,\ldots,l}$ are gaussian fuzzy sets. Thus, the multivariable neuro-fuzzy mapping is of the form:

$$\begin{bmatrix} \hat{y}_1(k+1) \\ \hat{y}_2(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{F}^1\left(y_1(k), y_2(k), u_1(k)\right) \\ \mathbf{F}^2\left(y_1(k), y_2(k), u_2(k)\right) \end{bmatrix} \tag{52}$$

where $\mathbf{F}^i$ represents the i-th MISO neuro-fuzzy network. An experimental data set of 300 points is created from (50), with random inputs $u_1(k)$ and $u_2(k)$ uniformly distributed in the interval $[-1, 1]$, with $\sigma^2 = 0.02$, meaning that the noise level applied to outputs takes values between 0 and $\pm15\%$ from its nominal values, which is an acceptable practical percentage of noise. This data set is presented to the proposed algorithm based on the local approach: The consequent parameters are estimated for each rule i, independently of each other, minimizing a set of weighted local criteria $(i = 1, 2, \ldots, l)$:

$$\hat{\theta}_i = \text{ arg min } \parallel \mathbf{S}_i\theta_i - \mathbf{b}_i \parallel_2^2 \tag{53}$$

where $\mathbf{S}_i = \mathbf{Z}^T\mathbf{\Psi}_i\mathbf{X}$ and $\mathbf{b}_i = \mathbf{Z}^T\mathbf{\Psi}_i\mathbf{Y}$.

In this application, the number of rules is 4 for each MISO neuro-fuzzy network, the antecedent parameters are obtained by the ECM (*Evolving Clustering Method*) method proposed in (Kasabov and Song, 2002), 30 points were used for the initial estimate, $\lambda = 0.99$ and $dl = 1$. The linguistic variables partitions were obtained by the ECM method. These partitions correspond to the membership functions of the layer II in the neuro-fuzzy network. For (50), the obtained TS multivariable neuro-fuzzy network rules base is:

$R^1:$ IF $y_1(k)$ is F_1 AND $y_2(k)$ is G_1 THEN

$$\begin{bmatrix} \hat{y}_1(k+1) \\ \hat{y}_2(k+1) \end{bmatrix} = \begin{bmatrix} 0.6177 & 0.0038 \\ 0.0316 & 0.0454 \end{bmatrix} \begin{bmatrix} \hat{y}_1(k) \\ \hat{y}_2(k) \end{bmatrix} + \begin{bmatrix} 1.0133 & 0.0000 \\ 0.0000 & 1.0006 \end{bmatrix} \begin{bmatrix} u_1(k) \\ u_2(k) \end{bmatrix} + \begin{bmatrix} 0.0056 \\ -0.0014 \end{bmatrix}$$

R^2 : IF $y_1(k)$ is F_1 AND $y_2(k)$ is G_2 THEN

$$\begin{bmatrix} \hat{y}_1(k+1) \\ \hat{y}_2(k+1) \end{bmatrix} = \begin{bmatrix} 0.6580 & 0.0201 \\ -0.1021 & 0.0749 \end{bmatrix} \begin{bmatrix} \hat{y}_1(k) \\ \hat{y}_2(k) \end{bmatrix} + \begin{bmatrix} 1.0033 & 0.0000 \\ 0.0000 & 1.0217 \end{bmatrix} \begin{bmatrix} u_1(k) \\ u_2(k) \end{bmatrix} + \begin{bmatrix} 0.0062 \\ 0.0310 \end{bmatrix}$$

R^3 : IF $y_1(k)$ is F_2 AND $y_2(k)$ is G_1 THEN

$$\begin{bmatrix} \hat{y}_1(k+1) \\ \hat{y}_2(k+1) \end{bmatrix} = \begin{bmatrix} 0.5508 & 0.0456 \\ 0.0395 & -0.3209 \end{bmatrix} \begin{bmatrix} \hat{y}_1(k) \\ \hat{y}_2(k) \end{bmatrix} + \begin{bmatrix} 1.0701 & 0.0000 \\ 0.0000 & 1.1199 \end{bmatrix} \begin{bmatrix} u_1(k) \\ u_2(k) \end{bmatrix} + \begin{bmatrix} -0.0722 \\ 0.0639 \end{bmatrix}$$

R^4 : IF $y_1(k)$ is F_2 AND $y_2(k)$ is G_2 THEN

$$\begin{bmatrix} \hat{y}_1(k+1) \\ \hat{y}_2(k+1) \end{bmatrix} = \begin{bmatrix} 0.5847 & -0.1090 \\ -0.1058 & -0.3805 \end{bmatrix} \begin{bmatrix} \hat{y}_1(k) \\ \hat{y}_2(k) \end{bmatrix} + \begin{bmatrix} 1.0733 & 0.0000 \\ 0.0000 & 1.1119 \end{bmatrix} \begin{bmatrix} u_1(k) \\ u_2(k) \end{bmatrix} + \begin{bmatrix} -0.1177 \\ -0.1139 \end{bmatrix}$$

The recursive estimate results are shown in Fig. 2.

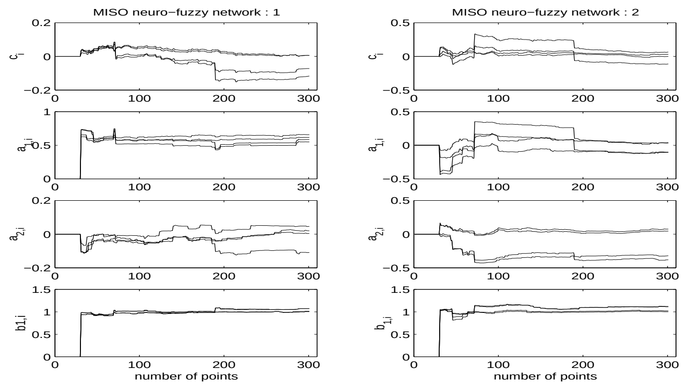

Figure 2. Parameters estimates of the multivariable neuro-fuzzy network

6. Conclusions

An approach for neuro-fuzzy identification of multivariable nonlinear discrete-time dynamical systems was proposed. The structure of the TS neuro-fuzzy network as well as an interpretation of the LPV systems view point were presented. Convergence conditions for identification in a noisy environment were introduced. Based on the TS neuro-fuzzy network, an *on-line* scheme was developed to form a NARX neuro-fuzzy structure from samples of a multivariable nonlinear dynamical system where the consequent parameters were modified by an adaptive WIV algorithm based on the numerically robust orthogonal Householder transformation. Simulation results have shown that with the proposed methodology a multivariable TS neuro-fuzzy network rules base for a multivariable nonlinear plant was efficiently obtained in a noisy environment.

7. References

Bottura, C.P. and Serra, G.L.O., 2004, “ An Algorithm for Fuzzy Identification of Nonlinear Discrete-Time Systems”, 43^{rd} IEEE Conference on Decision and Control CDC’04, pp. 5421-5426.

Bottura, C.P. and Serra, G.L.O., 2005, “ Fuzzy Instrumental Variable Concept and Identification Algorithm”, Accepted in IEEE International Conference on Fuzzy Systems FUZZY-IEEE’05, May 20-22, Reno-Nevada (EUA).

Kasabov, N.K. and Song, Q., 2002, “DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and Its Application for Time-Series Prediction”, IEEE Transactions on Fuzzy Systems, Vol. 10, No. 2, pp. 144-154.

Ljung, L., 1999, “System Identification: Theory for the user”, 2nd ed., Prentice Hall.

Papadakis, S.E. and Theocaris, J.B., 2002, “A GA-based Fuzzy Modeling Approach for Generating TSK models”, Fuzzy Sets and Systems, 131, 2002, pp. 121-152.

Takagi, T. and Sugeno, M., 1985, “Fuzzy Identification of Systems and its Aplication to Modeling and Control”, IEEE Transations on Systems, Man, and Cybernetics, Vol. 15, no. 1, pp. 116-132.