Recursive Estimation of the Takagi-Sugeno Models I: Fuzzy Clustering and the Premise Membership Functions Estimation

Vojislav Filipović[1*], Vladimir Đorđević[1]
[1]Faculty of Mechanical and Civil Engineering, University of Kragujevac, Kraljevo (Serbia)

Fuzzy modelling is an approximation of nonlinear systems by a finite collection of linear systems. On this concept Takagi-Sugeno fuzzy models are based. The procedure for identification of these models include two steps: (a) estimation of membership functions, (b) model parameter estimation. In this paper only the step (a) is considered, where Gustafson-Kessel clustering algorithm is used. The algorithm detects clusters of different shapes. Parameter estimation of the premise membership function is based on the implementation of recursive least squares algorithm. Based on the obtained clusters, recursive least squares algorithm estimates parameters of membership functions. In this paper, it is assumed that the membership functions have triangular shape, performances of the proposed algorithm are demonstrated by simulation.

Keywords: Fuzzy modelling, Fuzzy clustering, Nonlinear systems, Gustafson-Kessel algorithm

1. INTRODUCTION

In general, dynamical model of a system is nonlinear. Identification of this class of systems has been given many attention. Origins of this theory lie in different disciplines: control theory (identification of linear dynamical systems), nonparametric regression and statistics, learning theory, classification theory in pattern recognition, neural networks, fuzzy logic and other disciplines [1]. In this paper is considered the application of fuzzy logic for identification of nonlinear systems. Here will be discussed Takagi-Sugeno models [2]. In these models is used the idea of linearization of nonlinear systems in fuzzy regions of the state space. The structures are obtained, with several linear models. Input space is decomposed into a finite collection of fuzzy regions. The consequent functions describe system behavioural in those regions.

In classic control theory there are approaches that decompose nonlinear model into a finite collection of linear models. Example for that is included angle dividing method [3]. Using this method a finite collection of linear systems is obtained, as a base for further design of the controller. Similar, but more sophisticated methodology is obtained using gap metric concept [4], [5].

Methodologies [3]-[5], as well as the methodology discussed in this paper, are based on fuzzy logic, and they are alternatives to the well-known methodologies for design of controllers for nonlinear systems: feedback linearization [6] and backstepping [7].

The procedure for identification of Takagi-Sugeno models has two steps:

a) Estimation of premise membership functions,
b) Parameter estimation of consequent functions.

In this paper is discussed problem a), while problem b) will be discussed in complete authors’ paper [8].

Problem a) is solved using cluster analysis on Cartesian product space of input and output. For cluster analysis is used Gustafson-Kessel fuzzy clustering algorithm. In order to complete the solution of the problem a), after the clusters are defined, it is necessary to determine the parameters of membership functions. It is assumed that membership functions have triangular shape, and their parameters are estimated using recursive least squares algorithm.

The methodology exposed in this paper is demonstrated, thought simulation, on Hammerstein model.

2. TAKAGI-SUGENO MODELS

A nonlinear model $y=f(x)$ can be expressed in the form of Takagi-Sugeno (TS) model based on input-output measurements $\mathbf{u}_k=[u_{1k},u_{2k},...,u_{nk}]^T$ and y_k where k denotes measurements in the k-th moment, and n is the number of regressors in model.

TS model is a combination of logical and mathematical model. Logical rules are consisted of fuzzy premise, and consequent is a mathematical function. The general form of TS model [2]:

$$\begin{aligned}&\mathrm{R}_i: IF\ \mathbf{u}\ IS\ A_i(\mathbf{u})\ THEN\ y_i=\mathbf{a}_i^T\mathbf{u}+b_i;\\&i=1,2,\dots,c\end{aligned}\tag{1}$$

where, $\mathbf{u}\in\mathbb{R}^n$ and $y_i\in\mathbb{R}^1$ are inputs and outputs of the system, respectively. Values $a_i\in\mathbb{R}^n$ and $b_i\in\mathbb{R}^1$ are parameters of TS model. R_i is the i-th rule, and c is the number of rules in rule base. A_i is multivariable premise membership function of the i-th rule.

For individual components of vector u, TS model have the following form:

$$\begin{aligned}&\mathrm{R}_i: IF\ u_1\ IS\ A_{i1}(u_1)\ AND...AND\ u_n\ IS\ A_{in}(u_n)\\&THEN\ y_i=\mathbf{a}_i^T\mathbf{u}+b_i;\ i=1,2,...,c\end{aligned}\tag{2}$$

Degree of fulfilment of the rule is equal

$$\beta_i(\mathbf{u})=\prod_{j=1}^{n}\mu_{A_{ij}}(\mathbf{u})\tag{3}$$

where $\mu_{A_{ij}}(u)$ is the membership function of the fuzzy set A_{ij}.

*Corresponding author: Dositejeva 19, 36000 Kraljevo (Serbia) and v.filipovic@open.telekom.rs

The inference is computed using following formula [2]:

$$y=\frac{\sum_{i=1}^{c}\beta_i(\mathbf{u})\left(\mathbf{a}_i^T\mathbf{u}+b_i\right)}{\sum_{i=1}^{c}\beta_i(\mathbf{u})} \quad (4)$$

From relations (2) and (4) is evident that TS model approximates nonlinear system with finite collection of linear systems.

3. FUZZY CLUSTERING

Fundamental property of measurements for defining clusters is similarity. Therefore, it is necessary to determine the appropriate metrics. Consider an n-dimensional vector of measurements $\mathbf{x}_k=\left[x_{1,k},x_{2,k},\ldots,x_{n,k}\right]^T$, $\mathbf{x}_k\in\mathbb{R}^n$. Set of N measurements is denoted with $\mathbf{X}=\{\mathbf{x}_k \mid k=1,2,\ldots,N\}$ and it is represented in form of an $n\times N$ matrix

$$\mathbf{X}=\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,\mathrm{N}} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,\mathrm{N}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,\mathrm{N}} \end{bmatrix} \quad (5)$$

In pattern recognition terminology [10], the columns of matrix $\mathbf{X}$ are called patterns, and the rows are called features or attributes. Matrix $\mathbf{X}$ is called the matrix of patterns or data.

Using exposed, Euclidean distance can be defined as

$$d_2\left(\mathbf{x}_i,\mathbf{x}_j\right)=\left(\sum_{k=1}^{d}\left(x_{i,\mathrm{k}}-x_{j,k}\right)^2\right)^{\frac{1}{2}}=\left\|\mathbf{x}_i-\mathbf{x}_j\right\|_2 \quad (6)$$

The more general form of distance is Minkowski distance

$$d_p\left(\mathbf{x}_i,\mathbf{x}_j\right)=\left(\sum_{k=1}^{d}\left(x_{i,\mathrm{k}}-x_{j,k}\right)^p\right)^{\frac{1}{p}}=\left\|\mathbf{x}_i-\mathbf{x}_j\right\|_p \quad (7)$$

Practice shows that distance (6) is suitable in a case when clusters are isolated. Distance (6) and (7) express weakness if the features are linearly correlated. Methods based on d_2 and d_p distances cannot distinguish between two groups of observations in the Figure [11].

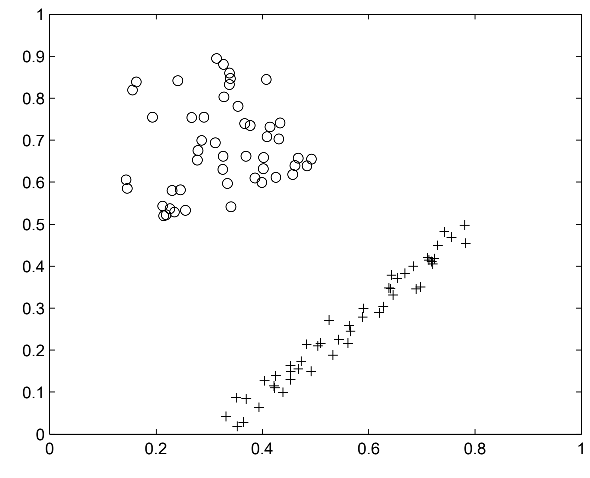

Figure 1: Two different groups of data

In this case Mahalanobis distance is introduced

$$d_M\left(\mathbf{x}_i,\mathbf{x}_j\right)=\left(\mathbf{x}_i-\mathbf{x}_j\right)^T\mathbf{F}^{-1}\left(\mathbf{x}_i-\mathbf{x}_j\right) \quad (8)$$

where $\mathbf{F}$ is a covariance matrix. Using the distance (8) Gustafson-Kessel algorithm is obtained [12].

Gustafson-Kessel clustering algorithm is an iterative optimisation algorithm for minimisation the value of objective function

$$J\left(\mathbf{X};\mathbf{V},\mathbf{U},\{\mathbf{M}_i\}\right)=\sum_{i=1}^{c}\sum_{k=1}^{N}\left(\mu_{i,k}\right)^m D_M^2\left(\mathbf{x}_k,\mathbf{v}_i\right) \quad (9)$$

with constraints

$$\mu_{i,k}\in[0,\ 1];\quad 1\le i\le c;\quad 1\le k\le N \quad (10)$$

$$\sum_{i=1}^{c}\mu_{i,k}=1;\qquad k=1,\ldots,N \quad (11)$$

where c is the number of clusters, and m is weighting exponent. The weighting exponent m determines fuzziness of the clusters, and exponent value must be greater than 1. For $m=1$ algorithm performs a hard clustering, a pattern is or is not element of the cluster. With increase of m, the overlap of fuzzy clusters is increased also. Typically, value of m is 2.

In the objective function (9), Mahalanobis distance is replaced with an inner-product norm distance of the form

$$D_{M_i}^2\left(\mathbf{x}_k,\mathbf{v}_i\right)=\left(\mathbf{x}_k-\mathbf{v}_i\right)^T\mathbf{M}_i\left(\mathbf{x}_k-\mathbf{v}_i\right) \quad (12)$$

where $\mathbf{M}_i$ is symmetric and positive-definite norm-inducing matrix.

The arguments of objective function are pattern matrix $\mathbf{X}$, fuzzy partition matrix $\mathbf{U}=\left[\mu_{ik}\right]$, $\mathbf{U}\in\mathbb{R}^{c\times N}$, prototype matrix $\mathbf{V}$ is a set of vector of clusters prototypes (centres) $\mathbf{V}=\left[\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_c\right]$, $\mathbf{v}_i\in\mathbb{R}^n$, and $\{\mathbf{M}_i\}$ is the c-tuple of local norm-inducing matrices.

Using the method of Lagrange multiplier is obtained [12] that the objective function have minimal value in case when

$$\mu_{ik}=\frac{1}{\sum_{j=1}^{c}\left(\frac{D_{\mathbf{M}}^2\left(\mathbf{x}_k,\mathbf{v}_i\right)}{D_{\mathbf{M}}^2\left(\mathbf{x}_k,\mathbf{v}_j\right)}\right)^{\frac{2}{m-1}}},\quad \begin{array}{l}1\le i\le c\\ 1\le k\le N\end{array} \quad (13)$$

and

$$\mathbf{v}_i=\frac{\sum_{k=1}^{N}\left(\mu_{ik}\right)^m\mathbf{x}_k}{\sum_{k=1}^{N}\left(\mu_{ik}\right)^m},\quad 1\le i\le c \quad (14)$$

Matrices $\mathbf{M}_i$ are used as optimisation variables for distance adaptation depending on the layout of data, and as result of optimisation it is obtained the following expression for $\mathbf{M}_i$

$$\mathbf{M}_i=\left[\rho_i\det\left(\mathbf{F}_i\right)\right]^{\frac{1}{n}}\mathbf{F}_i^{-1} \quad (15)$$

where ρ_i are the clusters volumes, and it's value is usually 1.

Fuzzy covariance matrix for the i-th cluster is given by following expression

$$\mathbf{F}_i = \frac{\sum_{k=1}^{N} \left(\mu_{i,k}\right)^m \left(\mathbf{x}_k - \mathbf{v}_i\right)\left(\mathbf{x}_k - \mathbf{v}_i\right)^T}{\sum_{k=1}^{N} \left(\mu_{i,k}\right)^m} \tag{16}$$

Fuzzy covariance matrices contain information of shape and orientation of the cluster. Every cluster can be represented as hyperellipsoid defined by equation

$$\left(\mathbf{x} - \mathbf{v}_i\right)^T \mathbf{F}_i^{-1} \left(\mathbf{x} - \mathbf{v}_i\right) = 1 \tag{17}$$

Figure 2 shows hyperellipsoid defined by equation (17). The semi-axis of the cluster's hyperellipsoid are the eigenvalues of fuzzy covariance matrix $\mathbf{F}_i$, and directions of the axis are corresponding eigenvectors.

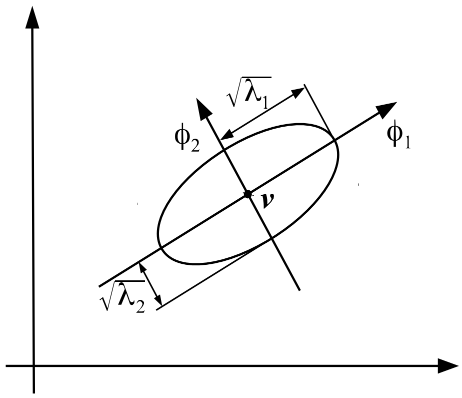

Figure 2: Cluster's hyperellipsoid

Practice shows that Gustafson-Kessel algorithm is suitable method for system identification for several reasons [13]. Since it is based on adaptive distance measure, clusters of different shapes and orientation can be detected. The initialised partition matrix have small influence on results, and also normalisation and standardisation of the data.

On other hand, the large number of clusters and date can result the long execution of algorithm. Also there is singularity problem of covariance matrix in cases when small number of observations is available or when the data are linearly correlated.

The pseudo code for Gustafson-Kessel algorithm is given in following table.

Gustafson-Kessel algorithm

Algorithm inputs are: the pattern matrix $\mathbf{X}$, the number of clusters c, the weighting exponent m, the clusters volumes ρ_i

Initialise random partition matrix $\mathbf{U}^{(0)}$

Do for $l = 1, 2, \ldots$

Compute cluster prototypes:

$$\mathbf{v}_i^{(l)} = \frac{\sum_{k=1}^{N} \left(\mu_{ik}^{(l-1)}\right)^m \mathbf{x}_k}{\sum_{k=1}^{N} \left(\mu_{ik}^{(l-1)}\right)^m}, \quad 1 \le i \le c$$

Compute the cluster covariance matrix:

$$\mathbf{F}_i = \frac{\sum_{k=1}^{N} \left(\mu_{ik}^{(l-1)}\right)^m \left(\mathbf{x}_k - \mathbf{v}_i^{(l)}\right)^T \left(\mathbf{x}_k - \mathbf{v}_i^{(l)}\right)}{\sum_{k=1}^{N} \left(\mu_{ik}^{(l-1)}\right)^m}, \quad 1 \le i \le c$$

Compute distance:

$$\mathbf{M}_i = \rho_i \det\left(\mathbf{F}_i\right)^{\frac{1}{n}} \mathbf{F}_i^{-1}$$

$$D_{M_i}^2\left(\mathbf{x}_k, \mathbf{v}_i\right) = \left(\mathbf{x}_k - \mathbf{v}_i^{(l)}\right)^T \mathbf{M}_i \left(\mathbf{x}_k - \mathbf{v}_i^{(l)}\right),$$

$$1 \le i \le c, \quad 1 \le k \le N$$

Update the partition matrix:

For $1 \le i \le c$

For $1 \le k \le N$

If $D_{M_i}\left(\mathbf{x}_k, \mathbf{v}_i\right) > 0$

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c} \left(\frac{D_{\mathbf{M}_i}^2\left(\mathbf{x}_k, \mathbf{v}_i\right)}{D_{\mathbf{M}_i}^2\left(\mathbf{x}_k, \mathbf{v}_j\right)}\right)^{\frac{1}{m-1}}}$$

Otherwise

$$\mu_{ik}^{(l)} = 0 \text{ and } \sum_{i=1}^{c} \mu_{ik}^{(l)} = 1$$

Until $\max\left(\left|U^{(l)} - U^{(l-1)}\right|\right) < \varepsilon$

After the clusters are obtained, in the following procedure the parameters of membership functions need to be estimated.

For estimation of premise membership functions, the clusters need to be projected on premise variables. The projection of clusters on premise variable is point-wise operation. The rows of partition matrix are projected onto original regression values.

In order to get better results, it is necessary to extract linear part of point-wise set of a premise function using α -cut. After that, the extracted linear part needs to be separated in two groups, one group on each side from the centre. Each group is approximated with straight line.

Based on relation (1), the j-th set can be represented as

$$y_{k_j}^j = a^j u_{kj} + b^j = \begin{bmatrix} u_{k_j} & 1 \end{bmatrix} \begin{bmatrix} a^j \\ b^j \end{bmatrix} = \left(\boldsymbol{\varphi}_{k_j}^j\right)^T \boldsymbol{\theta}_{k_j}^j \tag{18}$$

For estimation of parameters of membership functions, a^j and b^j, is used recursive least squares algorithm

$$\hat{\boldsymbol{\theta}}_{k_j}^j = \hat{\boldsymbol{\theta}}_{(k-1)_j}^j + \mathbf{P}_{k_j} \boldsymbol{\varphi}_{k_j} \left(y_{k_j} - \boldsymbol{\varphi}_{k_j}^T \hat{\boldsymbol{\theta}}_{(k-1)_j}^j\right) \tag{19}$$

$$\mathbf{P}_{k_j} = \mathbf{P}_{(k-1)_j} - \frac{\mathbf{P}_{(k-1)_j} \boldsymbol{\varphi}_{k_j}^j \left(\boldsymbol{\varphi}_{k_j}^j\right)^T \mathbf{P}_{(k-1)_j}}{1+\left(\boldsymbol{\varphi}_{k_j}^j\right)^T \mathbf{P}_{(k-1)_j} \boldsymbol{\varphi}_{k_j}^j} \qquad (20)$$

The initial values are $\hat{\theta}_0^j = 0$ and $P_{0_j} = 10^4 I$.

Once the parameters of both lines are estimated, it is needed to calculate value of premise variable, denoted with β, where lines intersect each other, and also x-intercept of each line, denoted with α and γ.

Now, when all three parameters of fuzzy set are known, the membership function can be defined as normalised point-wise triangular membership function:

$$\mu_{A_{ij}}\left(x,\alpha,\beta,\gamma\right) = \max\left(\min\left(\frac{x-\alpha}{\beta-\alpha},\frac{\gamma-x}{\gamma-\beta}\right),0\right) \quad (21)$$

4. SIMULATIONS

The methodology is demonstrated on Hammerstein model, Figure 3.

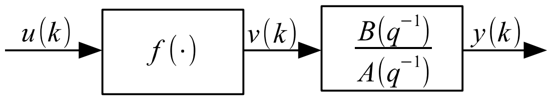

Figure 3: Hammerstein model

Nonlinear function of Hammerstein model is

$$v(k) = u(k) + 0.5u^2(k) + 0.25u^3(k)$$

and polynomials $A\left(q^{-1}\right)$ and $B\left(q^{-1}\right)$ are

$$A\left(q^{-1}\right) = 1 - 1.6q^{-1} + 0.8q^{-2}$$

$$B\left(q^{-1}\right) = 0.85q^{-1} + 0.65q^{-2}$$

As input signal is used following multi-sinusoidal function

$$u_k = 10\sin\left(0.01\cdot t\right) + 5\sin\left(0.1\cdot t\right) + 2.5\cdot\sin\left(0.25\cdot t\right) + \\ +0.75\cdot\sin\left(t\right)$$

where $t = 1,2,\ldots,N$. Figure 4 shows input and output of Hammerstein model.

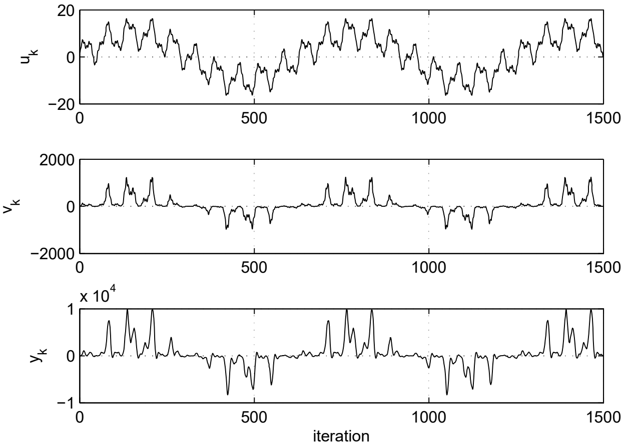

Figure 4: Input and output of Hammerstein model

The Hammerstein model is represented as first-order NARX model,

$$y(k+1) = F\left(y(k), u(k)\right)$$

with following pattern matrix:

$$\mathbf{X} = \begin{bmatrix} y(1) & y(2) & \ldots & y(N-1) \\ u(1) & u(2) & \ldots & u(N-1) \\ y(2) & y(3) & \ldots & y(N) \end{bmatrix}$$

The optimal number of cluster can be find using performance measures [14]. The first performance measure is fuzzy hypervolume and it is defined by:

$$F_{HV} = \sum_{i=1}^{c} \left[\det\left(F_i\right)\right]^{\frac{1}{2}} \qquad (22)$$

where F_i are obtained from Gustafson-Kessel algorithm.

The second performance measure is partition density, and given by formula:

$$P_D = \frac{S}{F_{HV}} \qquad (23)$$

where

$$S = \sum_{i=1}^{c}\sum_{j=1}^{N} \mu_{ij}, \\ \forall \mathbf{x}_j \in \left\{\mathbf{x}_j : \left(\mathbf{x}_j - \mathbf{v}_i\right)^T \mathbf{F}_i^{-1}\left(\mathbf{x}_j - \mathbf{v}_i\right) < 1\right\} \qquad (24)$$

Both performance measures are calculated for different numbers of cluster, and the results are shown on Figure 5.

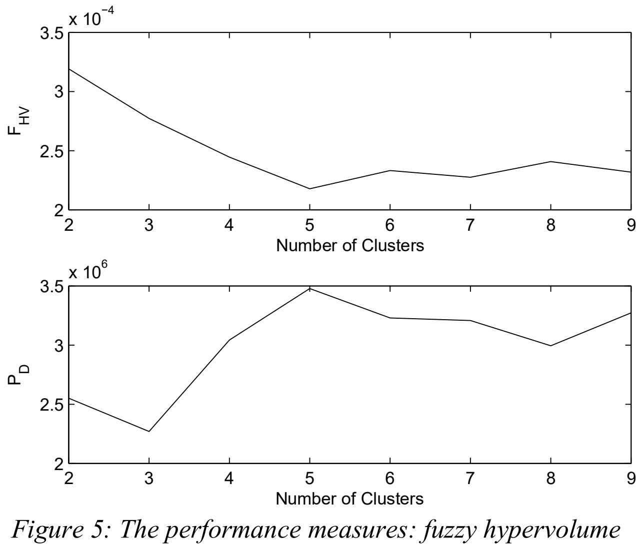

Figure 5: The performance measures: fuzzy hypervolume and partitiom density

The minimum of fuzzy hypervolumes and the maximum of partition densities is when the number of clusters is 5. This is the optimal number of clusters for this first-order NARX model and this matrix of patterns.

Figure 6 shows a premise membership function, obtained by projecting cluster on premise variable. The membership degrees of premise variable are presented with dots. The centre of the membership function is presented with dotted line, and linear approximations of the membership function are presented with dashed lines.

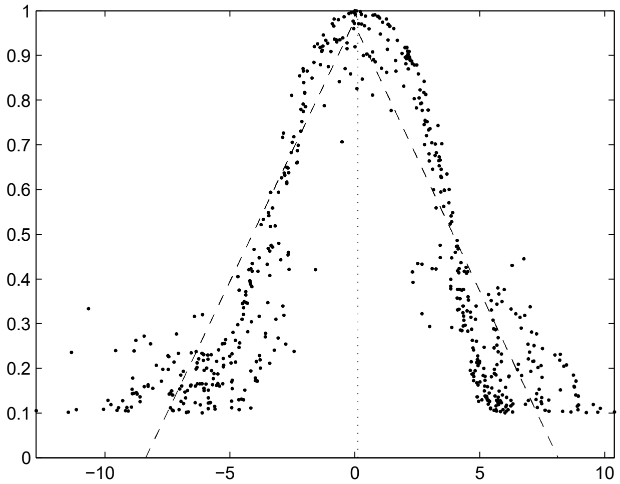

Figure 6: Estimation of membership function

Figures 7 and 8 show normalised estimated membership functions for regressors $u(k)$ and $y(k)$.

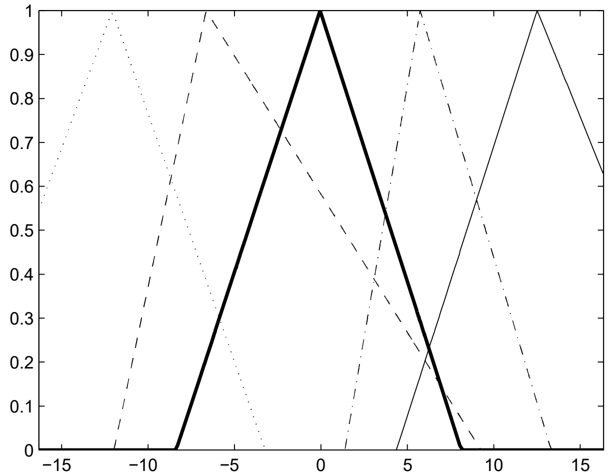

Figure 7: Estimated membership function for $u(k)$

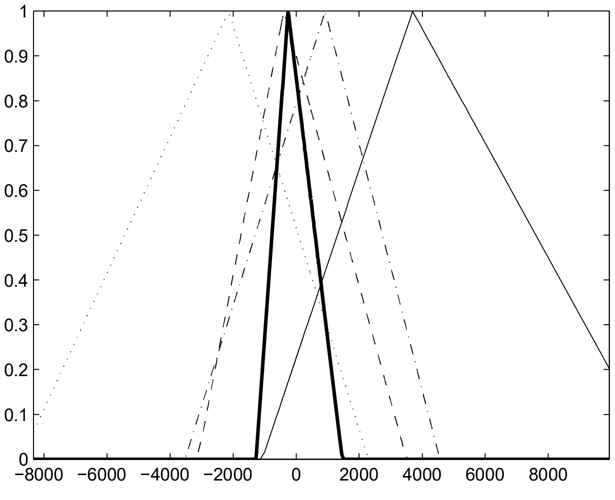

Figure 8: Estimated membership functions for $y(k)$

5. MISCELLANEOUS

In this paper, a methodology for determination of premise membership functions is presented, and its performances are demonstrated through simulation on Hammerstein model. The methodology employs Gustafson-Kessel algorithm to find the clusters, and recursive least squares algorithm to estimates parameters of premise membership functions. The results of this methodology are point-wise triangular membership functions, which are very suitable for use, especially in online applications.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to the Ministry of Education, Science and Technological Development of the Republic of Serbia for supporting this paper through project TR33026 and project TR33027.

REFERENCES

[1] L. Ljung, "Some aspects of nonlinear system identification," Proceedings of the 14th IFAC Symposium on System Identification, Newcastle (Australia), p. 761-766, (2006)

[2] T. Takagi and M. Sugeno, "Fuzzy identification of Systems and its application to modelling and control," IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-15, p. 116-132, (1985)

[3] J. Du, X. Zhang and C. Song, "Multi-PID Control of Hammerstein Systems with Input Multyplicity", Proceedings of 30th Chinese Control Conference, Yantai (China), 22-24 July 2011, pp. 303-306, (2011)

[4] O. Galán, J. A. Romagnoli, A. Palazoğlu and Y. Arkun, "Gap Metric Concept and Implications for Multilinear Model-Based Controller Design", Ind. Eng. Chem. Res. Vol. 42, pp. 2189-2197, (2003)

[5] G. Vinnicombe, "Uncertainty and Feedback. H_∞ Loop-Shaping and the ν-Gap metric", Imperial College Press, London (UK), (2000)

[6] A. Isidori, "Nonlinear Control Systems", Springer, Berlin (Germany), (1995)

[7] M. Krstić, I. Kanellakopoulos and P. Kokotović, "Nonlinear and Adaptive Control Design," Wiley, New York, (1995)

[8] V. Filipović and V. Đorđević, "Recursive Estimation of the Takagi-Sugeno Models II: Recursive estimation of fuzzy Hammerstein models," (to be published)

[9] J. Abonyi and B. Feil, "Cluster Analysis for Data Mining and System Identification", Birkhauser, Basel (Switzerland), (2007)

[10] R. O: Duda, P. E. Hart and D. G. Stork, "Pattern Classification", Wiley, New York (USA), (2000)

[11] S. Miyamoto, H. Ichihashi and K. Honda, “Algorithms for Fuzzy Clustering: Methods in c-Mean Clustering with Applications”, Springer, Berlin (Germany), (2008)

[12] E. E. Gustafson and W.C. Kessel, “Fuzzy Clustering with a Fuzzy Covariance Matrix,” IEEE Conference of Decision and Control including the 17th Symposium on Adaptive Processes, Vol. 17, p. 761-766, (1978)

[13] R. Babuška, “Fuzzy Modeling for Control”, Springer Science + Business Media, New York (USA), (1998)

[14] I. Gath and A. B. Geva, “Unsupervised Optimal Fuzzy Clustering”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, pp. 773-781, (1989)