Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Optimal Estimation by Using Fuzzy Systems

Oleg S. Amosov*, Lyudmila N. Amosova*

**Amur State University of Humanities and Pedagogy, Komsomolsk-on-Amur, Russia,*
(Tel: +74217-554151; e-mail: aos@kmscom.ru)

Abstract: The paper compares the Bayesian algorithms for estimation of random vectors and the algorithms based on the fuzzy systems. It is shown that the traditional and fuzzy logic algorithms provide the estimates with the similar properties. The comparison results are discussed. The efficiency of applying the Takagi-Sugeno fuzzy systems to the nonlinear estimation problems is investigated by two examples.

1. INTRODUCTION

Since the introduction of the fuzzy set theory by Zadeh (1973), many people have devoted a great deal of time and effort to both the theoretical research and implementation technique for the fuzzy systems (FS). FS have been investigated in the context of adaptive control and system identification (Mamdani and Assilian, 1974; Takagi and Sugeno, 1985; Kreinovich *et al*., 1998; Tanaka and Wang, 2001; Shaaban *et al*., 2006). But only recently they came to be used for filtering problem (Chan *et al*., 1997; Wu and Harris, 1997; Crocetto and Ponte, 2002; Amosov, 2004a, b). Optimal filtering is known to be widely used in estimation of random processes and sequences (Kalman, 1960; Meditch, 1969; Jazwinski, 1970; Yarlykov and Mironov, 1999; Stepanov, 1998). However constructing algorithms requires a comprehensive a priori information about the processes estimated and their measurement errors. Besides, serious difficulties emerge in the constructing nonlinear filtering algorithms (Jazwinski, 1970; Yarlykov and Mironov, 1999; Stepanov, 1998; Dmitriev and Stepanov, 1998). These disadvantages make the researchers look for new approaches to the construction of algorithms.

One of such approaches can be based on fuzzy systems due to their capability to approximate any nonlinear behavior and the possibility to be applied to the solution of difficult (from the calculation standpoint) problems.

However, in our opinion, there is no an unambiguous answer about the advantages or disadvantages of the fuzzy approach in comparison with the traditional one. Most attention in the papers has been concentrated on the methods of applying fuzzy systems to the filtering and estimation. The publications that do compare FS and optimal filtering approaches concern, as a rule, some particular examples and are based on simulation. The authors do not discuss the relation between the traditional and fuzzy systems algorithms. In our opinion, this makes it difficult to use widely FS for the solution of applied problems.

Such relation is investigated for the particular problem of linear estimation in present paper. It is shown that for Takagi-Sugeno (T-S) FS with the linear consequents and the appropriate choice of the criterion used for its off-line generating, the traditional and fuzzy logic algorithms are practically identical and they provide estimates with the similar properties. Besides, the present paper is devoted to a more general nonlinear, non-Gaussian case, for which the problem of linear estimation is a particular case. The main publications (Stepanov and Amosov, 2004; 2005; 2006) devoted to the neural network based estimation have been served as the base of this paper.

2. NONLINEAR ESTIMATION PROBLEM

2.1 Problem Statement

Consider the following problem: to estimate an n-dimensional random vector $\mathbf{x}=[x_1 \; ... \; x_n]^T$ by m-dimensional measurements $\mathbf{y}=[y_1 \; ... \; y_m]^T$ (Stepanov and Amosov, 2006)

$$\mathbf{y}=\mathbf{s}(\mathbf{x})+\mathbf{v}\,, \tag{1}$$

where $\mathbf{s}(\mathbf{x})=[\mathrm{s}_1(\mathbf{x}) \; ... \; s_m(\mathbf{x})]^T$ is the known m-dimensional nonlinear vector-function; $\mathbf{v}=[v_1 \; ... \; v_m]^T$ is a random vector of measurement errors. Suppose that the joint probability density function (p.d.f.) $f(\mathbf{x},\mathbf{v})$ for the vectors $\mathbf{x}$ and $\mathbf{v}$ is known. For simplicity $\mathbf{v}$ and $\mathbf{x}$ are assumed to be zero means random vectors independent of each other, i.e. $f(\mathbf{x},\mathbf{v})=f(\mathbf{x})f(\mathbf{v})$. Thus, taking into account (1), it is possible to get the p.d.f. $f(\mathbf{x},\mathbf{y})$ for $\mathbf{x}$ and $\mathbf{y}$ $f(\mathbf{x},\mathbf{y})=f(\mathbf{x})f_{\mathrm{v}}(\mathbf{y}-\mathbf{s}(\mathbf{x}))$.

The assumptions made allow to state the problem of finding the optimal (minimum variance) estimate $\hat{\mathbf{x}}(\mathbf{y})$ that minimizes the criterion

$$J=E[\|(\mathbf{x}-\hat{\mathbf{x}}(\mathbf{y})\|^2]\,, \tag{2}$$

where $\|a\|^2=a^T a$; E is the mathematical expectation corresponding to $f(\mathbf{x},\mathbf{y})$. Two variants of the solution to this problem are discussed below.

2.2 Nonlinear Optimal Estimate

The nonlinear optimal estimate and the covariance matrix of estimation errors are determined as (Jazwinski, 1970)

$$\hat{\mathbf{x}}^{opt}(\mathbf{y}) = \int \mathbf{x} f(\mathbf{x}/\mathbf{y}) d\mathbf{x}\,, \tag{3}$$

$$\mathbf{P}^{opt}(\mathbf{y}) = \int (\mathbf{x} - \hat{\mathbf{x}}^{opt}(\mathbf{y}))(\mathbf{x} - \hat{\mathbf{x}}^{opt}(\mathbf{y}))^T f(\mathbf{x}/\mathbf{y}) d\mathbf{x}\,,$$

where $f(\mathbf{x}/\mathbf{y})$ is the a posteriori (conditional) p.d.f. for the vector $\mathbf{x}$. It should be noted that the symbol of the integral in these expressions and below corresponds to the multiple integrals with the infinite limits. The matrix $\mathbf{P}^{opt}(\mathbf{y})$ (conditional error covariance matrix) characterizes the accuracy of the state-vector estimate for the given set of the measurements $\mathbf{y}$. It is well known that the problem of designing an algorithm for the calculation of $\hat{\mathbf{x}}^{opt}(\mathbf{y})$ and $\mathbf{P}^{opt}(\mathbf{y})$ is easily solved only for the Gaussian $f(\mathbf{x},\mathbf{v})$ and the linear character of the function $\mathbf{s}(\mathbf{x})$, i.e. when $\mathbf{s}(\mathbf{x}) = \mathbf{H}\mathbf{x}$. For this case the dependence of the optimal estimate on measurements has a linear character. In all other cases there arises the problem of designing suboptimal algorithms that do not involve a large size of calculations.

2.3 Linear Optimal Estimate

One of the variants of designing suboptimal algorithms is reduced to finding a linear optimal estimate instead of (3). Then the estimate is calculated as $\hat{\mathbf{x}}^{lin}(\mathbf{y}) = \bar{\mathbf{x}} + \mathbf{K}^{lin}[\mathbf{y} - \mathbf{y}^{lin}]$. The idea of designing a linear optimal algorithm consists in choosing the gain factor matrix $\mathbf{K}^{lin}$ and the vector $\mathbf{y}^{lin}$ in such a way as to minimize the criterion (2) in the class of the linear estimates. It can be shown that the linear optimal estimate is determined as (Medich, 1969):

$$\hat{\mathbf{x}}^{lin}(\mathbf{y}) = \bar{\mathbf{x}} + \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{yy}}^{-1}[\mathbf{y} - \mathbf{y}^{lin}]\,, \tag{4}$$

$$\mathbf{K}^{lin} = \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{yy}}^{-1}\,, \tag{5}$$

$$\mathbf{P}^{lin} = E[(\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}))(\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}))^T] = \mathbf{P}_0 - \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{yy}}^{-1}\mathbf{P}_{\mathbf{yx}}\,, \tag{6}$$

where $\bar{\mathbf{x}}$, $\mathbf{y}^{lin} = \bar{\mathbf{y}}$, $\mathbf{P}_0$, $\mathbf{P}_{\mathbf{yy}}$ are the mathematical expectations and the covariance matrices of the vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{P}_{\mathbf{xy}}$ is the cross covariance matrix for $\mathbf{x}$ and $\mathbf{y}$. Thus the problem of finding the optimal linear estimate is reduced to calculation of the first two moments of the joint vector that includes the vector of the parameters being estimated $\mathbf{x}$ and the measurement vector $\mathbf{y}$. After these moments have been derived, the relations (4)–(6) are used to calculate the estimates and the corresponding covariance matrix.

3. BAYESIAN ESTIMATION IN THE PRESENCE OF A TRAINING SET

For the problem considered it means that there is a set of data

$$\{(\mathbf{y}^{(j)}, \mathbf{x}^{(j)})\}\,,\ j = \overline{1.n_o}\,, \tag{7}$$

in which the pairs $\mathbf{y}^{(j)}$, $\mathbf{x}^{(j)}$, $j = \overline{1.n_o}$ are the independent-of-each-other realizations of the random vector $\mathbf{z} = [\mathbf{x}^T \quad \mathbf{y}^T]^T$, with the p.d.f. $f(\mathbf{x},\mathbf{y})$.

Let us consider a possible statement of the estimation problem for the case when, instead of the $f(\mathbf{x},\mathbf{v})$ or $f(\mathbf{x},\mathbf{y})$, the set of data (7) is known. In other words, assume that the a priori information is given in the form of (7) and it is necessary, having this set and the measurement $\mathbf{y}$, to find the estimate $\tilde{\mathbf{x}}(\mathbf{y})$ that minimizes the following criterion:

$$\tilde{J} = \frac{1}{n_o}\sum_{j=1}^{n_o}\left\|\mathbf{x}^{(j)} - \tilde{\mathbf{x}}(\mathbf{y}^{(j)})\right\|^2. \tag{8}$$

As $E\left\|\mathbf{x} - \hat{\mathbf{x}}\ (\mathbf{y})\right\|^2 = \int\int \left\|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})\right\|^2 f(\mathbf{x},\mathbf{y}) d\mathbf{x} d\mathbf{y}$, then, in accordance with the Monte Carlo method, it is possible to write (Zaritsky, *et. al.,* 1975):

$$\lim_{n_o \to \infty} \frac{1}{n_o}\sum_{j=1}^{n_o}\left\|\mathbf{x}^{(j)} - \tilde{\mathbf{x}}(\mathbf{y}^{(j)})\right\|^2 = E\left\|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{y})\right\|^2,$$

i.e. criterion (8) tends to (2) as the n_o increases. It is evident that in these conditions the estimation algorithm, optimal in the sense of criterion (8), will be similar to the traditional Bayesian algorithm (3), optimal in the minimum variance sense. The approximate solution to this problem can be found by introducing a class of parameter-dependent functions used for the calculation of the estimate. Then the criterion (8) can be written as:

$$\tilde{J}^*(\tilde{\mathbf{W}}) = \frac{1}{n_o}\sum_{j=1}^{n_o}\left\|\mathbf{x}^{(j)} - \tilde{\mathbf{x}}(\mathbf{y}^{(j)}, \tilde{\mathbf{W}})\right\|^2, \tag{9}$$

where $\tilde{\mathbf{W}}$ is the vector or matrix determining a set of free parameters that define the function $\tilde{\mathbf{x}}(\mathbf{y}, \tilde{\mathbf{W}})$.

Hence it follows that the problem of deriving the estimation algorithm is reduced to finding the parameters $\tilde{\mathbf{W}}$ determined by the minimization of the criterion formed with the use of the data of the training set (7). The algorithms based on the minimization of the criterion of the types (8), (9) are widely used in the pattern recognition. They are usually called the algorithms of empirical risk minimization (Vapnik, 1982; Haykin, 1994).

4. FUZZY MODEL BASED ESTIMATOR

From the above it follows that the statement of the problem under consideration is in full agreement with the statement for the solution of problems with the use of FS. Thus in order to find the estimate $\tilde{\mathbf{x}}(\mathbf{y},\tilde{\mathbf{W}})$, it is possible to use the fuzzy system, i.e.

$$\hat{\mathbf{x}}^{FS}(\mathbf{y}) = \mathbf{K}^{FS}(\mathbf{y},\tilde{\mathbf{W}}), \qquad (10)$$

where $\mathbf{K}^{FS}(\mathbf{y},\tilde{\mathbf{W}})$ is FS; $\tilde{\mathbf{W}}$ is the matrix that specifies the free parameters (the membership function parameters) and $\mathbf{y}$ is the input of the fuzzy system.

There are four principal parts in a fuzzy system: fuzzifier, fuzzy rule base, fuzzy inference engine, and defuzzifier. The estimation procedure using FS can be divided into two tasks (Shaaban *et al*., 2006): structure identification, which determines the type and number of rules and membership functions, and parameter identification which adjust fuzzy systems parameters such as membership parameters. For both structural and parametric adjustments, a priori knowledge plays an important role.

Thus, using FS, we must realize the unknown mapping $\mathbf{K}^{FS}(\mathbf{y},\tilde{\mathbf{W}})$ (10) for estimating the state vector $\mathbf{x}$ in terms of the input information vector $\mathbf{y}$ (1) if the training set (7) is available.

The corresponding fuzzy rule base consists of a collection of fuzzy *If-Then* rules in the following form:

$$R_k:\ \textit{If}\ \mathbf{y}\ \textit{is}\ \mathbf{A}_k, \textit{Then}\ \hat{\mathbf{x}}^{\mathrm{FS}}\ \textit{is}\ \mathbf{B}_k\ ;\quad k=\overline{1.\eta}, \qquad (11)$$

where $\mathbf{y}$ is the input measurement vector (1), $\mathbf{A}_k = A_{k1}\times...\times A_{km}$ are fuzzy sets defined on the Cartesian product $\mathbf{Y}$ of universal sets of input linguistic variables and having the membership functions $\mu_{A_{ki}}(y_i)$, $k=\overline{1.\eta}$, $i=\overline{1.m}$, and $\mathbf{B}_k = B_{k1}\times...\times B_{kn}$ are fuzzy sets defined on the Cartesian product $\mathbf{X}$ of universal sets of output linguistic variables and having the membership functions $\mu_{B_{ki}}(\hat{x}_i^{FS})$, $k=\overline{1.\eta}$, $i=\overline{1.n}$. η is the total number of fuzzy *If-Then* rules in the rule base.

For designing the fuzzy system it is possible to realize them on the basis of fuzzy neural networks (NN) and, similarly to conventional NN, to use a gradient method for adjusting the parameters of given predicate rules. The matrix $\tilde{\mathbf{W}}$ is determined when FS is generated trained according with the criterion (9), where $\tilde{\mathbf{x}}(\mathbf{y},\tilde{\mathbf{W}}) = \hat{\mathbf{x}}^{FS(j)}(\mathbf{y}^{(j)},\tilde{\mathbf{W}})$ is the estimate generated by FS by the measurements $\mathbf{y}^{(j)}$ corresponding to the realization of $\mathbf{x}^{(j)}$.

5. SOLUTION OF THE ESTIMATION PROBLEM WITH THE USE OF A TAKAGI-SUGENO SYSTEMS

For T-S approach the universal approximation property was proven (Kreinovich *et al*., 1998). Let us solve the estimation problem by using a Takagi-Sugeno FS under the assumption that the training set (7) has been specified:

$$R_k:\ \textit{If}\ \mathbf{y}\ \textit{is}\ \mathbf{y}^{(j)}, \textit{Then}\ \hat{\mathbf{x}}^{FS\ (j)}(\mathbf{y}) = \mathbf{K}_k^{FS}(\mathbf{y},\tilde{\mathbf{W}}_k);\quad k=\overline{1.\eta}. \qquad (12)$$

The final output of the T-S fuzzy system can be represented by (Kreinovich *et al*., 1998)

$$\hat{\mathbf{x}}^{\mathrm{FS}\ (j)} = \sum_{k=1}^{\eta}\alpha_k^{(j)}\mathbf{K}_k^{FS}(\mathbf{y}^{(j)},\tilde{\mathbf{W}}_k) / \sum_{k=1}^{\eta}\alpha_k^{(j)}, \qquad (13)$$

where $\alpha_k^{(j)} = f_{\&}(\mu_1(\mathbf{a}_1,y_1^{(j)}),\mu_2(\mathbf{a}_2,y_2^{(j)}),...,\mu_m(\mathbf{a}_m,y_m^{(j)}))$; $f_{\&}(a,b)=\min(a,b)$ *or* $f_{\&}(a,b)=a\cdot b$; $\mu_i(\mathbf{a}_i,y_i^{(j)})$ is the membership function for input $i, i=\overline{1.m}$; $\mathbf{a}_i$ is a vector of parameters for $\mu_i(\mathbf{a}_i,y_i^{(j)})$.

Let's establish the relation between the traditional linear optimal and T-S fuzzy logic algorithms. Consider T-S fuzzy models with the linear consequents described by a set of fuzzy rules as follows

$$R_k:\ \ \textit{If}\ \mathbf{y}\ \textit{is}\ \mathbf{y}^{(j)}, \textit{Then}\ \ \hat{\mathbf{x}}^{FS\ (j)} = \mathbf{w}_0 + \mathbf{W}\mathbf{y}\quad ; k=\overline{1.\eta}. \qquad (14)$$

Taking into consideration the dimensions of the vector to be estimated, the T-S fuzzy system $\hat{\mathbf{x}}^{FS}(\mathbf{y},\tilde{\mathbf{W}})$ (13) with linear consequents (14), and the number of fuzzy rules $\eta=1$ can be simplified as follows:

$$\hat{\mathbf{x}}^{FS(j)}(\mathbf{y}^{(j)},\tilde{\mathbf{W}}) = \mathbf{w}_0 + \mathbf{W}\mathbf{y}^{(j)},\ j=\overline{1.n_o}, \qquad (15)$$

where $\tilde{\mathbf{W}} = [\mathbf{w}_0\quad |\quad \mathbf{W}]$ is an $n\times(m+1)$ -dimensional matrix that includes an n- dimensional biases vector $\mathbf{w}_0 = [w_{10}\ \ ...\ \ w_{n0}]^T$ and an $n\times m$- dimensional matrix of weighing coefficients $\mathbf{W} = [\mathbf{w}_1\ \ |\ \ ...\ \ |\ \ \mathbf{w}_l\ \ |\ \ ...\ \ \mathbf{w}_n]^T$, in which $\mathbf{w}_l = [w_{l1}\ \ ...\ \ w_{lm}]^T$ are m- dimensional vectors $l=\overline{1.n}$. Using (15), the criterion (9) can be represented in the following form

$$\tilde{J}^*(\tilde{\mathbf{W}}) = \frac{1}{n_o}\sum_{j=1}^{n_o}\left\|\mathbf{x}^{(j)} - (\mathbf{w}_0 + \mathbf{W}\mathbf{y}^{(j)})\right\|^2. \qquad (16)$$

From the previous part it follows that the estimate (13) determined with the use of FS generated in accordance with the criterion (16) will tend to the optimal estimate (4) as the number of realizations n_o increases. To do this, one should, similarly to the way it was done in Reference (Stepanov and Amosov, 2004) calculate partial derivatives with respect to $\mathbf{w}_0$ and $\mathbf{W}$, and put them to zero. After some not complicated but tiresome transformations the derived equations can be resolved with respect to $\mathbf{w}_0$ and $\mathbf{W}$. As the result, the estimate $\hat{\mathbf{x}}^{FS}(\mathbf{y},\tilde{\mathbf{W}})$ derived by the measurements $\mathbf{y}$ with the use of FS (15) trained in accordance with (16) can be given as:

$$\hat{\mathbf{x}}^{FS}(\mathbf{y},\tilde{\mathbf{W}}) = \bar{\mathbf{x}}^* + \mathbf{P}^*_{\mathbf{xy}}\left(\mathbf{P}^*_{\mathbf{yy}}\right)^{-1}[\mathbf{y}-\bar{\mathbf{y}}^*], \qquad (17)$$

where $\bar{\mathbf{x}}^* = \mathbf{m}^*_{\mathbf{x}}$; $\bar{\mathbf{y}}^* = \mathbf{m}^*_{\mathbf{y}}$; $\mathbf{P}^*_{\mathbf{yy}}$, $\mathbf{P}^*_{\mathbf{xy}}$ are the sample values of the mathematical expectations and corresponding covariance matrices:

$$\mathbf{m}^*_{\mathbf{x}} = \frac{1}{n_o}\sum_{j=1}^{n_o}\mathbf{x}^{(j)}; \qquad \mathbf{m}^*_{\mathbf{y}} = \frac{1}{n_o}\sum_{j=1}^{n_o}\mathbf{y}^{(j)}, \qquad (18)$$

$$\mathbf{P}^*_{\mathbf{yy}} = \alpha^*_2[\mathbf{y}] - \mathbf{m}^*_{\mathbf{y}}(\mathbf{m}^*_{\mathbf{y}})^T; \quad \mathbf{P}^*_{\mathbf{xy}} = \alpha^*_{1,1}[\mathbf{x},\mathbf{y}] - \mathbf{m}^*_{\mathbf{x}}(\mathbf{m}^*_{\mathbf{y}})^T. \ (19)$$

$$\alpha^*_{1,1}[\mathbf{x},\mathbf{y}] = \frac{1}{n_o}\sum_{j=1}^{n_o}\mathbf{x}^{(j)}(\mathbf{y}^{(j)})^T; \ \alpha^*_2[\mathbf{y}] = \frac{1}{n_o}\sum_{j=1}^{n_o}\mathbf{y}^{(j)}(\mathbf{y}^{(j)})^T.$$

From Expressions (17) and (4) it is follows that FS, after some adequate generating under the conditions when the specified sample values of mathematical expectations and covariance matrices are close to their true values, provide the determination of the estimate close to optimal in the linear class. Thus, the optimal linear algorithm can be treated as a simplified Takagi-Sugeno FS trained in accordance with (16).

6. EXAMPLES

6.1 Example 1

It is necessary to estimate the random variable x, uniformly distributed on the interval $[0,b]$, from the noisy measurements of the form

$$y_l = x + v_l, \ l = \overline{1.i}, \qquad (20)$$

in which the measurement errors v_l, $l=\overline{1.i}$ are assumed to be zero-mean random variables independent of each other and of x uniformly distributed on the interval $[-a/2, a/2]$. In this example $\mathbf{x} \equiv x$, $\mathbf{y} \equiv [y_1 \ \ldots \ y_i]^T$, $\mathbf{H} = [1 \ \ldots \ 1]^{\mathrm{T}}$, $\mathbf{v} = [v_1 \ \ldots \ v_i]^{\mathrm{T}}$. It should be noted that the a posteriori p.d.f. $f(\mathbf{x}/\mathbf{y})$ is non Gaussian here, as x and v_l, $l=\overline{1.i}$ are the uniformly distributed random variables.

It is possible to find the linear optimal estimate $x^*(\mathbf{y})$ and the corresponding error covariance P^*_e by using (4), (5), i.e.

$$x^*(\mathbf{y}) = \bar{x} + \mathbf{P}_{xy}\mathbf{P}^{-1}_{\mathbf{yy}}[\mathbf{y}-\bar{\mathbf{y}}], \qquad (21)$$

$$P^*_e = P_0 - \mathbf{P}_{xy}\mathbf{P}^{-1}_{\mathbf{yy}}\mathbf{P}_{yx}, \qquad (22)$$

where $\bar{x} = \frac{b}{2}$; $\bar{\mathbf{y}} = \frac{1}{2}[b \ \ldots \ b]^T$;

$$P_0 = \sigma^2_0 = b^2/12; \ \mathbf{P}_{x\mathbf{y}} = \sigma^2_x\mathbf{H}^T; \mathbf{P}_{\mathbf{yy}} = \sigma^2_x\mathbf{I}_i + r^2\mathbf{E}_i. \ (23)$$

Here $\mathbf{I}_i$ is a square matrix composed of 1; $\mathbf{E}_i$ is a unit matrix, $\sigma^2_x = b^2/12$, $r^2 = a^2/12$. It is essential that the optimal nonlinear estimate can be determined exactly for this example (Stepanov and Amosov, 2005). To explain it, let us introduce the domain Ω that represents the crossing of all the intervals $[y_l - a/2, y_l + a/2]$, $l=\overline{1.i}$, i.e.

$$\Omega \equiv [d_1, d_2] = \bigcap_{l=1}^{i}[y_l - a/2, y_l + a/2]. \qquad (24)$$

It can be shown that the a posteriori density in the example considered is uniform on the interval $[c_1, c_2]$, which represents the crossing of the a priori domain $[0,b]$ and the domain Ω so that $c_1 = \max\{0, d_1\}, c_2 = \min\{b, d_2\}$. Then it follows that

$$\hat{x}(\mathbf{y}) = (c_2 + c_1)/2. \qquad (25)$$

Assume that the a priori information is represented by a set of pairs $x^{(j)}$, $\mathbf{y}^{(j)}$, $j=\overline{1.n_o}$. Then the estimation problem can be solved by using FS. Let us use both T-S FS1 (13)-(14) with i inputs and with number of the rules 2^i (number of membership functions per input is equal 2) and a simplified T-S FS2 (15) with i inputs and with number of the rules equal 1. Below there are the simulation results corresponding to the linear and nonlinear optimal estimates and T-S FS estimates derived for different number of measurements i. The simulation was performed under the assumption that $b=1$, $a=1$, $i=\overline{1.k}$, $k=1,2,\ldots,10$. The calculation of the optimal nonlinear estimate was carried out in accordance with (25).

Generating of T-S FS with Gaussian membership functions was performed in accordance with the criterion (16). To provide generating, the realizations $x^{(j)}$, $\mathbf{y}^{(j)}$, $j=\overline{1.n_o}$, $n_o = 3000$ were simulated in accordance with (20). Generating was followed by testing. For this purpose $n_\omega = 1000$ pairs of the realizations $x^{(j)}$, $\mathbf{y}^{(j)}$, $j=\overline{1.1000}$ were additionally simulated for various $i=\overline{1.k}$, $k=10$.

The values $\tilde{\sigma}_i$, $\tilde{\sigma}^\mu_i$, $\mu = FS1, FS2$ were calculated as

$$\tilde{\sigma}_i \approx \sqrt{\frac{1}{n_\omega}\sum_{j=1}^{n_\omega}(e_i^{(j)})^2}, \ e_i^{(j)} = x^{(j)} - \hat{x}^{(j)}(\mathbf{y}^{(j)}),$$

$$\tilde{\sigma}^\mu_i \approx \sqrt{\frac{1}{n_\omega}\sum_{j=1}^{n_\omega}(e^{\mu\,(j)}_i)^2}, \ e^{\mu\,(j)}_i = x^{(j)} - \hat{x}^{\mu\,(j)}(\mathbf{y}^{(j)},\tilde{\mathbf{W}}).$$

Figure 1 shows the sample root-mean- square (r.m.s.) errors: $\tilde{\sigma}^*_i \approx \sqrt{P^*_e}$ – for the linear optimal estimates; $\tilde{\sigma}_i$ – for the nonlinear optimal estimates; $\tilde{\sigma}^{FS1}_i$ – for T-S FS estimates; $\tilde{\sigma}^{FS2}_i$ – for the simplified T-S FS estimates.

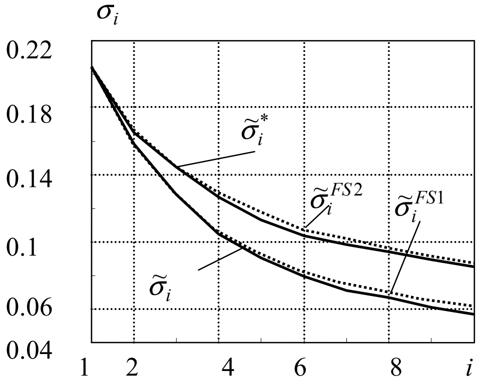

Fig. 1. The r.m.s. estimation errors x_i

As may be seen from Fig. 1 the estimate of the simplified Takagi-Sugeno FS2 and the optimal linear estimate are identical, but they differ very much from the optimal nonlinear estimate. At the same time the estimate of T-S FS1 is close to the optimal nonlinear estimate. It is of importance to note that the derivation of the optimal estimate involved a priori information presented in the form of the analytic dependence (20) and the known joint probability distribution density function $f(x,\mathbf{y}) = f(x) f(\mathbf{y}/x)$. At the same time the derivation of estimates with the use of FS only a set of realizations $x^{(j)}$, $\mathbf{y}^{(j)}$, $j=\overline{1.n_o}$ involved.

6.2 Example 2: APPLICATION TO NAVIGATION PROBLEM USING REFERENCE BEACONS

One of the navigation problems, whose nonlinear character has to be taken into account, is the so-called navigation with the use of measurements of distances $\mathbf{y} = (y_1, ..., y_m)^T$ to m reference beacons whose coordinates are assumed to be known.

As an illustration let us consider the problem of determining the unknown vector $\mathbf{x} = (x_1, x_2)^T$ with the use of measurements of distances $\mathbf{y} = (y_1, y_2)^T$ to two reference beacons whose coordinates are assumed to be known (Stepanov and Amosov, 2006). Assume that there is one, two and more pairs (l) of measurements which have the form

$$y_i^l = s_i(x) + v_i^l = \sqrt{\left(x_1 - x_1^i\right)^2 + \left(x_2 - x_2^i\right)^2} + v_i^l ,$$
$$i = \overline{1.2}, \quad l = 1,2,...,$$

where x_1, x_2; x_1^i, x_2^i, $i=\overline{1.2}$ are the unknown vector and the coordinates of the beacons, correspondingly.

It is assumed that $\mathbf{x} = (x_1, x_2)^T$ is the zero mean Gaussian vector with the diagonal covariance matrix and similar variances σ_0^2; v_i are the zero mean, independent-of-each-other and of $\mathbf{x}$, Gaussian random values with similar variances equal to r^2. Under the assumptions made $\mathbf{P}_0 = \sigma_0^2 \mathbf{E}_2$, $\mathbf{P}_\mathrm{v} = r^2 \mathbf{E}_{2i}$, where $\mathbf{E}_2$ and $\mathbf{E}_{2i}$ are 2×2 and $2i \times 2i$ unit matrices. It is also supposed that $\sigma_0 = 500\, m$; $r = 30\, m$; $\mathbf{x}^1 = (3000\, m, 0\, m)^T$; $\mathbf{x}^2 = (0\, m, 3000\, m)^T$.

The Cramer-Rao inequality was used to evaluate the potential accuracy. The latter makes it possible to find the lower Cramer-Rao boundary (CRB) for the unconditional covariance matrix of optimal estimate errors (Stepanov, 1998, Bergman 1999). It can be shown that for the problems considered the inequality can be written as

$$\mathbf{G}^{opt} \geq \mathbf{J}^{-1} , \tag{26}$$

where $\mathbf{J}^{-1} = \left(\frac{1}{\sigma_0^2} \mathbf{E}_2 + \frac{1}{r^2} \int \left[\frac{d\mathbf{s}^T(\mathbf{x})}{d\mathbf{x}} \frac{d\mathbf{s}(\mathbf{x})}{d\mathbf{x}^T} \right] f(\mathbf{x}) d\mathbf{x} \right)^{-1}$,

$$\mathbf{G}^{opt} = \int \int (\mathbf{x} - \hat{\mathbf{x}}^{opt}(\mathbf{y}))(\mathbf{x} - \hat{\mathbf{x}}^{opt}(\mathbf{y}))^T f(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} , \tag{27}$$

is the unconditional covariance matrix for the optimal estimation errors. It is easy to show that the square roots of the diagonal elements (26), determining CRB, can be calculated as $\sigma_k^{CRB} \approx \left(1/\sigma_0^2 + r^2/k\right)^{-1/2}$, where k is number of measurements. The square roots of the diagonal elements (27) are the r.m.s. errors for the corresponding algorithms. They characterize the accuracy of algorithms needed for comparison. The diagonal elements (27) were calculated as

$$\mathbf{G}_{ss}^{\mu} \approx \frac{1}{L} \sum_{j=1}^{L} ((\mathbf{x}_s^j - \hat{\mathbf{x}}_s^{\mu}(\mathbf{y}^j))^2 , s = 1, 2 ; \mu = lin, FS1, FS2.$$

By analogy with the example 1 let us use both T-S FS1 (13)-(14) with number of the rules 2^{2l} and a simplified T-S FS2 (15) with number of the rules equal 1. The Gaussian membership functions for FS are used. Figure 2 presents σ_l^{CRB} and $\tilde{\sigma}_l^*$, σ_l^{FS1}, σ_l^{FS2} which are CRB and r.m.s. errors at a different number of pairs measurements $l = 1,2,...$ corresponding to the linear optimal and FS methods for T-S FS1 and the simplified T-S FS2 estimates.

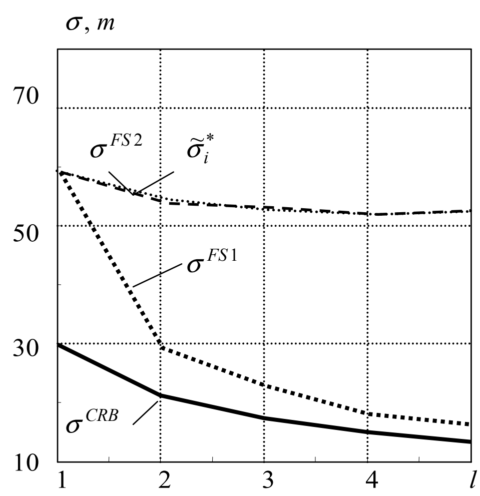

Fig. 2. The r.m.s. for x_l, $l = \overline{1.2}$ ($\sigma_0 = 500\, m$)

The number of the samples for generating of the fuzzy system based algorithms is equal to 3000. The number of the samples for testing is $L = 300$. The simulation results are shown for one of the component coordinates. They look similar for the other component as well. For the sake of simplicity the indices l are not shown in the plots.

The simulation results allow the following conclusions. The r.m.s. errors for the linear optimal and simplified T-S FS2 estimates are different from CRB because of the errors caused by the linear character of the algorithms.

Under the assumptions made even T-S FS1 (13)-(14) with the linear consequents provide accuracies close to the potential accuracy of the nonlinear optimal algorithm.

7. CONCLUSIONS

The research has proved that after some adequate generating the suggested simplified T-S FS provides the determination of the estimate close to the optimal in the linear class.

It is shown that the Bayesian and fuzzy logic algorithms provide the estimates with the similar properties in nonlinear, non-Gaussian case.

REFERENCES

Amosov, O.S. (2004a). Markov sequence filtering on the basis of Bayesian and neural network approaches and fuzzy logic systems in navigation data processing. *Journal of Computer and System Sciences International*, 43(4), pp. 551–560.

Amosov, O.S. (2004b). Fuzzy Logic Systems for Filtering of the Markovian Sequences. *Information technologies,* 11, pp. 16–24. In Russian.

Chan, K.C.C., V. Lee, and H. Leung (1997). Radar tracking for air surveillance in a stressful environment using a fuzzy-gain filter. *IEEE Trans. On Fuzzy Systems,* Vol. 5 (1), pp. 80–89.

Crocetto, N. and S. Ponte (2002). Blunder detection and estimation with fuzzy logic: applications to GPS code- and carrier-phase measurements. In: *Proc. of the 9th International Conference on Integrated Navigation Systems,* pp. 95–104. State Research Center of Russia "Elektropribor", Saint Petersburg.

Dmitriev, S.P. and O.A. Stepanov (1998). Nonlinear filtering and navigation. In: *Proc. of 5th International Conference on Integrated Navigation Systems.* Russia, CSRI Elektropribor, Saint Petersburg, pp. 138–149.

Haykin, S. (1994). *Neural networks: A comprehensive foundation.* MacMillan College, New York.

Jazwinski, A.H. (1970). *Stochastic processes and filtering theory*. Academic Press, New York.

Kalman, R.E. (1960). New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng. March*, 35-46.

Kreinovich, V., G.C. Mouzouris, and H.T. Nguyen (1998). Fuzzy rule based modeling as a universal approx-imation tool. In: H.T. Nguyen, M. Sugeno (Eds.) *Fuzzy Systems: Modeling and Control*. Kluwer, Boston, MA, pp.135-195.

Mamdani, E.H. and Assilian S. (1974). Applications of Fuzzy Algorithms for Control of Simple Dynamic Plant. *IEEE Proc. Part-D,* Vol. 121, pp. 1585-1588.

Meditch, J.S. (1969). Stochastic optimal linear estimation and control. Mc. Graw Hill, New York.

Shaaban, A. Salman, R. Puttige Vishwas, and G. Anavatti Sreenatha (2006). Real-time Validation and Comparison of Fuzzy Identification and State-space Identification for a UAV Platform. In: *Proc. of the CAC/CACD/ICC*, Munich, Germany, October 4-6, pp. 2138-2143.

Stepanov, O.A. (1998). *Nonlinear filtering and its application in navigation.* Russia, CSRI Elektropribor, Saint Petersburg. In Russian.

Stepanov, O.A. and O.S. Amosov (2004). Nonrecurrent linear estimation and neural networks. *IFAC Workshop on Adaptation and Learning in Control and Signal Processing, and IFAC Workshop on Periodic Control Systems.* Yokohama, Japan, August 30 – September 1, pp. 213–218.

Stepanov, O. A. and O. S. Amosov (2005). Optimal estimation by using neural networks. In: *Proc. of the 16th IFAC World Congress*, Prague, Czech Republic, July 3–8, 6 p.

Stepanov, O.A. and O.S. Amosov (2006). Optimal Estimation Algorithms Based on the Monte Carlo Method and Neural Networks for Nonlinear Navigational Problems. In: *Proc. of the CAC/CACD/ICC,* Munich, Germany, October 4-6, pp. 1432-1437.

Takagi, T. and Sugeno M. (1985). Fuzzy Identification of Systems and its Applications to Modeling and Control. *IEEE Trans. Syst., Man, Cybern.,* 15, pp. 116-132.

Tanaka, K. and Wang H.O. (2001). *Fuzzy Control System Design and Analysis - A Linear Matrix Inequality Approach*. John Wiley & Sons Inc.

Yarlykov, M.S. and M.A. Mironov (1999). The Markov theory of estimating random processes. In: *Telecommunications and Radioengineering.* 50(2–12). Published by Begell House, Inc., New York.

Zadeh, L.A. (1973). Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. *IEEE Trans. Systems Man Cybernetics*, 3, pp. 28-44.

Zaritsky, V.S., V.B. Svetnik, and L.I. Shimelevich (1975). The Monte-Carlo techniques in problems of optimal information processing. *Automaion and Remote Control*, 36, 2015–2022.

Vapnik, V.N. (1982). *Estimation of dependences based on empirical data.* New York: Springer-Verlag.

Wu, Z.Q. and C.J. Harris (1997). A neurofuzzy network structure for modeling and state estimation of unknown nonlinear systems. *International Journal of Systems Science,* Vol. 28(4), pp. 335–345.