Johanyák, Zs. Cs., Kovács, Sz.: Sparse Fuzzy System Generation by Rule Base Extension, 11th IEEE International Conference of Intelligent Engineering Systems (IEEE INES 2007), June 29 – July 1, 2007, Budapest, ISBN 1-4244-1148-3, pp. 99-104, http://www.johanyak.hu

Sparse Fuzzy System Generation by Rule Base Extension

Zsolt Csaba Johanyák[1], Szilveszter Kovács[2]
Institute of Information Technologies
Kecskemét College, GAMF Faculty
Kecskemét, Hungary
[1]johanyak.csaba@gamf.kefo.hu, [2]kovacs.szilveszter@gamf.kefo.hu

***Abstract*—This paper aims the introduction and comparison of two novel fuzzy system generation methods that implement the concept of incremental Rule Base Extension (RBE). Both methods automatically obtain from given input-output data a low complexity fuzzy system with a sparse rule base.**

Keywords-rule base generation; sparse rule base; fuzzy rule interpolation; rule base extension

I. INTRODUCTION

Owing to the powerful hardware development the generation of fuzzy systems from input-output data came into lime light in the last few years. This paper introduces two novel methods aiming the incremental generation of fuzzy systems from data by creating first two starting rules and than extending the rule base in course of the parameter identification process depending on de amelioration velocity of the performance index.

The methods being presented aim the generation of Multiple Input Single Output (MISO) fuzzy systems. In case of Multiple Input Multiple Output (MIMO) phenomenon they can be used by producing an aggregative system that contains a separate subsystem for each output dimension.

The rest of this paper is organized as follows. Section II. presents the parameterization strategy and the condition sets applicable in case of the new linguistic term generation and parameter identification. Section III. introduces a novel concept for fuzzy system generation called RBE and two methods based on it. Section IV. recalls the basic ideas of two fuzzy rule interpolation based reasoning techniques. The applied performance index and tuning algorithm are presented in section V. Section VI. presents the results of several experiments, aiming the testing and comparison of the two methods.

II. PARAMETERIZATION AND CONDITIONS

Due to their tractability and good tuning capability, each partition is built from convex and normal trapezoid shaped (CNF) linguistic terms. The parameterization is the conventional one, the break-points, i.e. end-points of the support and core are considered as parameters. Their ordinate values are fixed (*0* in case of the lower base and *1* in case of the upper base) therefore only the abscissas have to be adjusted in course of the parameter identification process. The parameters are numbered in clockwise direction starting from the lower endpoint of the support (fig. 1).

One can formulate three general conditions a CNF set always has to met, and two special conditions applicable only in case of the first variant of the method.

A. General conditions

The below specified conditions have to be met in course of the tuning in case of trapezoid shaped fuzzy sets in order to ensure the CNF property.

- Starting from the second vertex the value of each parameter has to be greater or equal to its predecessor. It can be expressed by the inequality

$$p_k \geq p_{k-1}, 1 < k \leq 4, \tag{1}$$

where p_k is the current parameter, and its subscript k indicates its position in the sequence of vertices.

- In case of the first three vertices the value of each parameter has to be smaller or equal to its successor. It can be expressed by the inequality

$$p_k \leq p_{k+1}, 1 \leq k < 4. \tag{2}$$

- The reference point, in this case the midpoint of the core, has to be inside the range of the current linguistic

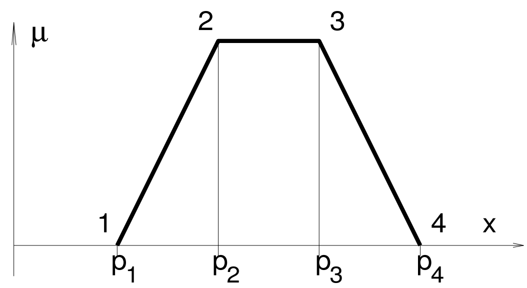

Figure 1. Conventional parameterization of a trapezoid shaped fuzzy set

Research partly supported by the Kecskemét College, GAMF Faculty grant no: 1N076/2006.

variable. It can be expressed by the inequality

$$R_{\min} \le \frac{p_2 + p_3}{2} \le R_{\max}, \tag{3}$$

where $R_{\min}, R_{\max}$ are the lower respective upper endpoints of the range of the actual input/output dimension.

The last constraint is applicable only in case of inference methods that allow a fuzzy set situated at the margin of a range to lap over the touched boundary. The inference methods based on the concept of linguistic term shifting (LTS) [2] belong to this category. Both of the techniques LESFRI [3] and FRIPOC [2], which were used in course of the preparation of this paper, apply LTS.

B. *Conditions specific to the first variant of the method*

One of the advantageous features of the fuzzy systems is the ability to extract linguistic rules that are easily interpretable by humans. In order to ensure this characteristic, two conditions have to be met. The first one states, that the support of the fuzzy set being adjusted cannot overlap the core of another set. Considering the case of the i^{th} set from fig. 2 it can be expressed by

$$p_1(A_i) \ge p_3(A_{i-1}), \tag{4}$$

$$p_4(A_i) \le p_2(A_{i+1}), \tag{5}$$

where p_j is the j^{th} parameter of a set.

The second condition specifies that the core of the linguistic term being adjusted cannot overlap the support of another linguistic term. It can be expressed by

$$p_2(A_i) \ge p_4(A_{i-1}), \tag{6}$$

$$p_3(A_i) \le p_1(A_{i+1}). \tag{7}$$

Conditions (4)..(7) define weaker constraints than the conditions necessary for ensuring a Ruspini character of the partition described e.g. in [4]. We used here less strict conditions in order to allow creation of sparse partitions as well.

III. Rule Base Generation

This section introduces the concept of fuzzy system generation based on *Rule Base Extension* (RBE) and two methods (RBE-DSS and RBE-SI) as its implementations. The methods differ only in the way they generate the shape of the antecedent respective consequent linguistic terms of the new rules, and in the conditions they apply in course of the tuning, therefore they are presented together.

RBE starts with an empty rule base and a set of training data points given in form of coherent input and output values. At the beginning of the process one defines the default core and support width values that will be applied in case of the newly generated linguistic terms. These values are expressed in percentage of the range of the input variables that can be given as a specification or can be determined from the available data. Thus the characteristic size values of the default shape are not identical in each dimension unless the widths of the domains are equal. In case of the RBE-SI method the default set shapes will be used only for the generation of the first two rules.

Next, the starting rule base is defined by determining the first two rules. They aim the description of the minimum and maximum output. First one seeks the two extreme output values and a representative data point for each of them. If several data points correspond to an extreme value, one should select the one that is closer to an endpoint of the input domain. For example, in the case of the function (11) the maximum value appears twice, at $\pi/2$ and at $5\pi/2$ (see fig. 4). We selected the first one, because it is closer to the lower bound of the studied x interval.

The reference points of the antecedent sets of the first rule will be identical with the corresponding input values of the minimum point. The reference point of the consequent set will be identical with the output value of the minimum point. The shape of the linguistic terms is determined by the default set shape, which is a characteristic feature of the dimension. The antecedent and consequent linguistic terms of the second rule are determined in a similar way taking into consideration the maximum point.

At this point the system contains two linguistic terms in each dimension. In order to fulfill the conditions specified in the previous section, the widths of the cores and supports are shrunk the same amount if it is necessary. However, their reference points are kept always unmodified. In case of a coincidence in any of the dimensions the sets are unified and only the first specified set is kept.

Having the first two rules determined, next a parameter identification process is started, which interactively adjusts the values of the inference parameters and the parameters of the linguistic terms. The details of the applied algorithm are presented in section *V.B*. If the decreasing velocity of the root mean square error (*RMSE*) (8) applied as performance indicator of the system falls below a specified threshold after a system evaluation a new rule is generated. It is because the system tuning reached a local or global minimum of *RMSE* and the performance cannot ameliorate further by the applied parameter identification algorithm. The new rule introduces

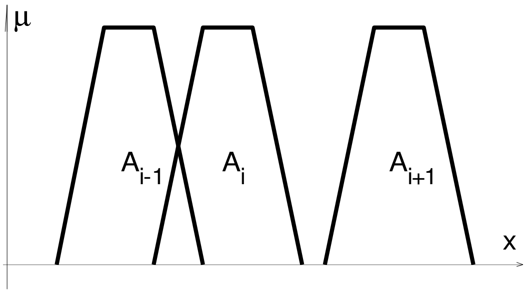

Figure 2. Neighboring sets fulfilling the special conditions (4)..(7)

additional tuning possibilities. However, in some cases *RMSE* will increase temporarily after the insertion of the new rule into the rule base.

In order to create the new rule, one seeks for the calculated data point, which is the most differing one from its corresponding training point. The input and output values of this training point will be the reference points of the antecedent and consequent sets of the new rule. The shape of the new linguistic terms is determined in two different ways by the two methods introduced in the followings.

The method *Rule Base Extension using Default Set Shapes* (RBE-DSS) applies a similar technique as seen in the case of the second rule. One starts with the default set shapes and than adjusts the core and support width values of the new sets and the neighboring linguistic terms as well conform to the conditions (1)..(7).

The method *Rule Base Extension using Set Interpolation* (RBE-SI) aims the minimization of the possible drawbacks of the introduction of the new rule. In the interest of better system performance it creates the antecedents and the consequent of the new rule by applying a set interpolation. Here the used technique depends on the applied fuzzy reasoning method. While in case of FRIPOC [2] the method FEAT-p [2] serves for the determination of the shape of the new linguistic terms, in case of LESFRI [3] the technique FEAT-LS [3] is used for the same purposes.

In order to alleviate the excessive increase of linguistic terms as a result of new rule generation two meta-rules are applied. The first one states that if the distance between the reference point of the new linguistic term and the reference point of an old fuzzy set is smaller or equal than a specified threshold value, instead of creating a new linguistic term, the old set is used as antecedent/consequent of the new rule in the current input/output dimension. The threshold value is expressed in percentage of the range of the actual partition. Its default value is *0.01 %*.

The second meta-rule allows the merging of the new linguistic set and an old set of the partition, when the average distance between their corresponding parameters is smaller or equal than a specified threshold. Also in this case the threshold is expressed in percentage of the range of the actual partition. Its default value is *0.1 %*.

Further on, the last two steps (parameter adjustment and new rule creation) are repeated until the specified iteration number has been reached, or the value of *RMSE* falls under a prescribed threshold.

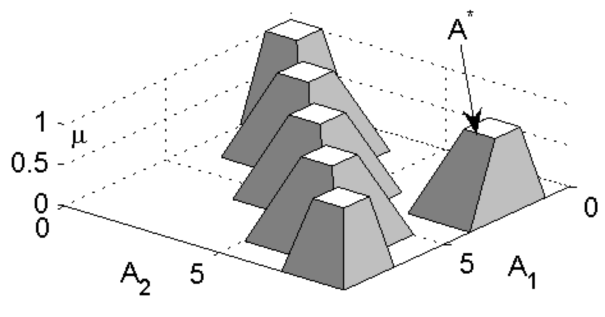

Figure 3. Antecedent space of a sparse rule base

IV. Inference Techniques

Both of the fuzzy system generation methods RBE-DSS and RBE-SI produce a sparse rule base in most of the cases. The “sparse” attribute means that there are no rules for all the possible input values. For example fig. 3 presents the input space of a system applying a sparse rule base. There are two input dimensions and the rule base consists of five rules. Each rule antecedent is represented by a pyramid. In case of the observation A* there is no rule whose antecedent part would overlap the observation at least partially. Therefore the classical compositional reasoning methods cannot afford an acceptable output and special approximate inference techniques are needed. In [7] Perfilieva studies the solvability of a fuzzy system approximating a given set of fuzzy data (fuzzy points), by compositional rule of inference, where the fuzzy rules are in the form of fuzzy implication.

Section IV.A and IV.B present shortly the basic concepts of two methods, which are applicable in cases when the full coverage of the input space is not ensured.

A. Fuzzy Rule Interpolation based on Polar Cuts (FRIPOC)

The first method we used in course of the experiments as fuzzy inference technique is FRIPOC. It was introduced in [2] and also belongs to the group of two-step fuzzy rule interpolation techniques. In the first step it determines a new rule whose antecedent is situated in the same position as the observation, i.e. their reference points are identical in each antecedent dimension. The shape of the antecedent and consequent linguistic terms is calculated by a set interpolation technique that is based on the concept of linguistic term shifting and polar cuts. The position of the consequent sets is determined by an adapted version of the Shepard interpolation [8].

The fuzzy sets representing the final conclusion are determined in the second step of FRIPOC applying a special single rule reasoning technique, which is also based on polar cuts and determines the shape of the new linguistic term going out from the differences between the antecedent sets of the interpolated rule and the sets that represent the observation.

B. Fuzzy Rule Interpolation by the Least Squares Method (LESFRI)

The second method applied as fuzzy rule interpolation based reasoning technique was LESFRI [3]. It was developed for the case when all linguistic terms can be characterized by the same shape type and the break-points in case of piece-wise linear membership functions are situated at the same α-levels.

LESFRI also belongs to the group of two-step methods. In the first step it determines the antecedent and consequent linguistic terms of the new rule applying the concepts of LTS and weighted least squares (WLS). Thus the resulting set shapes always belong to the characteristic shape type of the partition. Similar to FRIPOC the position of the consequent sets is determined by an adapted version of the Shepard interpolation [8]. The shape of the conclusion is calculated by preserving the weighted average difference measured between the observation and the antecedent of the interpolated rule on the consequent side and applying WLS.

V. Parameter Identification

A. Performance Index

In course of the parameter identification process after each parameter adjustment the resulting parameter set is evaluated by calculating the system output for a collection of predefined input data, for which the expected output values are known. In order to compare the results obtained with different parameter sets a performance index is calculated after each system evaluation.

Our algorithm uses the root mean square (quadratic mean) of the error (RMSE) as performance index for the evaluation of the fuzzy system. We chose it owing to its good comprehensibility and comparability to the range of the output linguistic variable. Its value is calculated by

$$RMSE = \sqrt{\frac{\sum_{j=1}^{M}\left(y_j - \hat{y}_j\right)^2}{M}}, \tag{8}$$

where M is the number of training data points, y_j is the output of the j^{th} data point and $\hat{y}_j$ is the output calculated by the system.

B. Parameter Identification Algorithm

The parameter identification method used in course of the calculations is a heuristic algorithm, a variant of the gradient descent method like the technique presented in [10]. Our algorithm starts the tuning with the parameters of the applied fuzzy inference method and continues with the adjustment of the parameters of the input and output linguistic terms. The final system considered as optimal (corresponding to a local or global minimum of RMSE) is iteratively approximated.

In course of iteration each parameter is modified one by one in both of the possible upper and lower (increasing and decreasing) directions. After determining a new value for the current parameter the system is evaluated calculating the actual value of the performance index. If this is better than the previous minimum the new parameter value is stored ("hill climbing").

The amount of modification of the set parameters is dependent on the range of the current input/output dimension, i.e. the step is calculated by multiplying the range by a coefficient. We use an adaptive approach, which determines the actual value of the coefficient depending on the change of RMSE during an iteration stage, the history of previous iteration stages, and a prescribed minimum value for the coefficient (C_{min}). For the calculation of C_{min} we determine first the range of each linguistic variable

$$DR_j = R_{j\max} - R_{j\min}, \quad 1 \le j \le n_{in} + n_{out}, \tag{9}$$

where n_{in} is the number of the input dimensions, $R_{j\min}$ and $R_{j\max}$ denote the lower respective upper endpoints of the current (j^{th}) dimension. The threshold for the coefficient is

$$C_{\min} = \frac{10^{-dn}}{\min\limits_{j\in[1,n_{in}+n_{out}]}\left(DR_j\right)}, \tag{10}$$

where dn is the maximal number of decimals used by the description of the parameters.

The iteration starts with a prescribed value of the coefficient (default: *0.2*). If the improvement of the system slows down or even stops, i.e. the value of RMSE does not reduce more than a prescribed threshold (default: *0.01*) during one iteration, the coefficient is divided by two unless its vale is already equal to C_{min}. In that case we generate a new rule.

On the other hand the coefficient is increased multiplying it by two when the improvement of the system speeds up, i.e. the value of RMSE increases more than a specified threshold (default: *10*) during one iteration. The algorithm also stops when the prescribed number of iteration is done.

VI. Experimental Results

In course of the experiments we used two data sets, one having one input and one output dimension and one having two input dimensions and one output dimension. In case of the first data set the data were generated by the function (see fig. 4)

$$y = \sin x, \quad x \in [0, 10], \tag{11}$$

where the value of x is interpreted in (radian) in course of the calculations.

The second data set were generated by the function

$$y = \left(1 + x_1^{-2} + x_2^{-1.5}\right)^2 \quad x \in [0.8, 1.8]. \tag{12}$$

It is the same function as used in [1], [9] and [10]. Fig. 5 shows the surface described by the second function. The systems were trained to the fuzzy rule interpolation based inference techniques FRIPOC [2] and LESFRI [3] separately. The time demand of one iteration could be different in case of the studied methods.

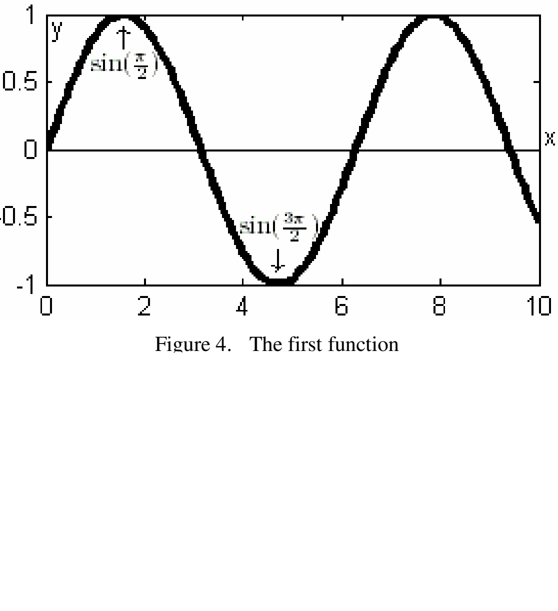

Figure 4. The first function

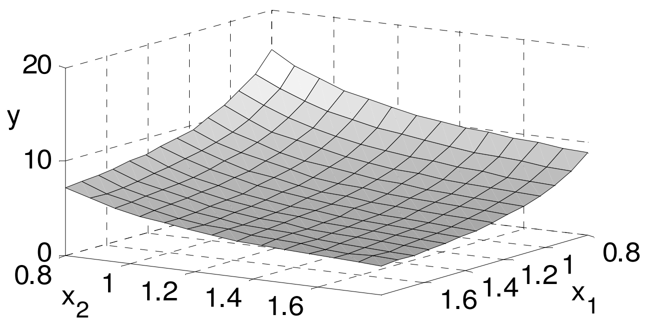

Figure 5. The second function

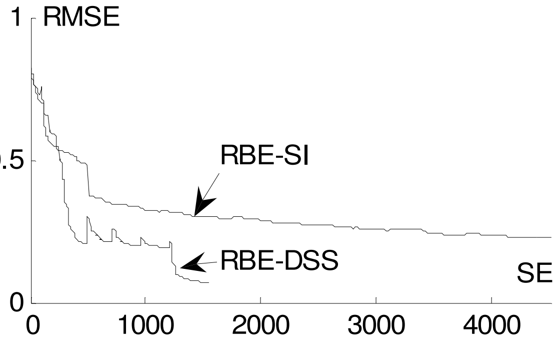

Figure 6. RMSE in course of the tuning of the first system for LESFRI with both methods

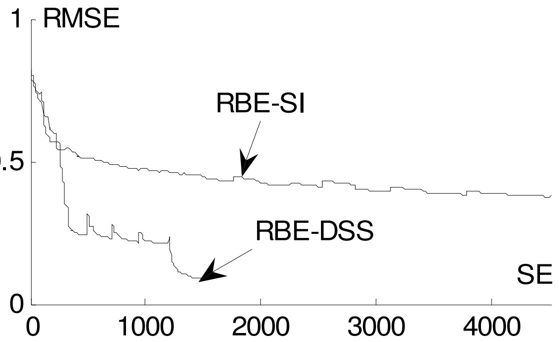

Figure 7. RMSE in course of the tuning of the first system for FRIPOC with both methods

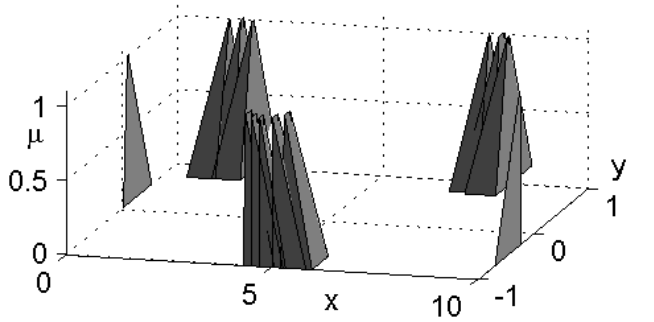

Figure 8. Rule base of the best performing system tuned for the first function with the RBE-DSS method and the LESFRI inference technique

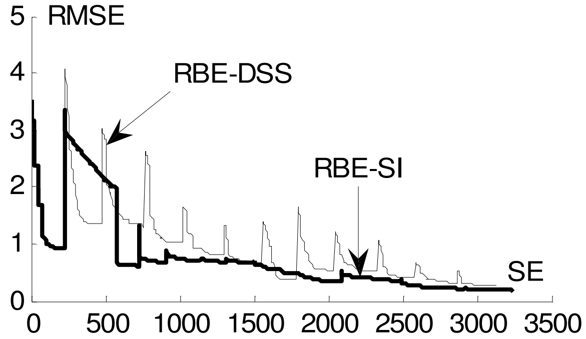

Figure 9. RMSE in course of the tuning of the second system for LESFRI with both methods

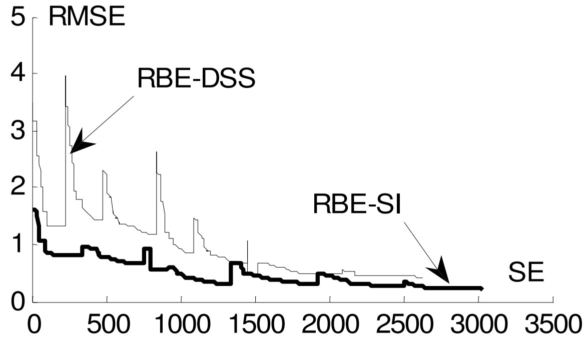

Figure 10. RMSE in course of the tuning of the second system for FRIPOC with both methods

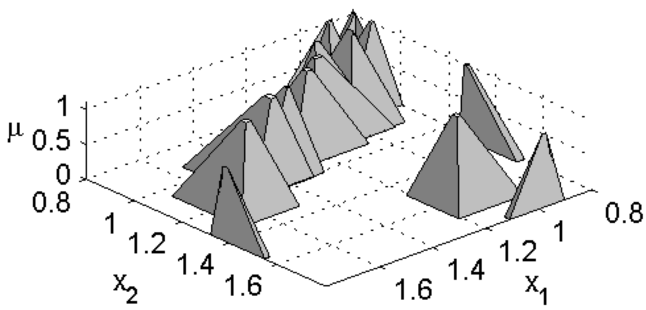

Figure 11. Antecedent space of the system tuned for the second function with the RBE-SI method and LESFRI inference technique

It is because they do not apply the same constraints for the newly calculated value of the parameter being adjusted. Thus it easily can happen that after applying the specified conditions one gets back the original value of the parameter. Under such circumstances no system evaluation is made and therefore the time need is negligible.

In order to ensure the comparability of the methods the horizontal axis indicates the number of system evaluations (SE) on figures showing the variation of the performance in course of the process (see e.g. fig. 6). During a system evaluation one calculates the output for all values belonging to the training data set, and calculates the performance index.

A. The SISO system

In case of the SISO system, which approximates function (11), RBE-DSS ensured the best results by both reasoning techniques (see fig. 6 and fig. 7). The performance curve of RBE-DSS shows some repetitive salient portions, which correspond to the introduction of new rules. In case of RBE-SI some protuberances also can be recognized, but these are much milder. Hence it can be stated that our expectations regarding the positive effects of the application of set interpolation were only partly fulfilled. The introduction of a new rule was not followed by a temporarily increase of the value of RMSE as it has been seen by RBE-DSS, but the error measure also did not decrease sufficiently in course of the examined number of system evaluations.

This case the better performance index values were obtained by the inference technique LESFRI. The rule base of the tuned system is represented on fig. 8. If one compares it with fig. 4 it can observe easily that the rules are grouped at the turning points of the function, like by the so-called optimal fuzzy rules discussed by Kosko in [6]. Probably the same results could be obtained by a smaller number of rules as well by an improved tuning algorithm. This topic is subject to further research work. Table I. presents the RMSE values of the final systems.

TABLE I. RMSE VALUES AT THE END OF THE PROCESS OF GENERATING THE FIRST SYSTEM

	LESFRI	FRIPOC
RBE-DSS	0.0709	0.0915
RBE-SI	0.2252	0.3742

B. *The MISO system*

In case of the MISO system, which approximates function (12), there are also observable peek points corresponding to the creation of new rules (see fig. 9 and 10). Surprisingly they appear significantly in case of the set interpolation based rule base extension as well. In case of both figures some portions of the performance curves are overlapped and therefore the one corresponding to RBE-SI is indicated by a bold line.

This case the final RMSE values obtained by the two methods are more similar. However the best performance is attained by RBE-SI in course of the tuning of the system for inference technique LESFRI. The input space (rule antecedents) of the best tuned system is represented on fig. 11. Table II. presents the RMSE values of the final systems.

TABLE II. RMSE VALUES AT THE END OF THE PROCESS OF GENERATING THE SECOND SYSTEM

	LESFRI	FRIPOC
RBE-DSS	0.2701	0.4143
RBE-SI	0.1809	0.2143

VII. CONCLUSIONS

This paper introduced a new approach called the concept of Rule Base Extension (RBE) and two automatic system generation methods (RBE-DSS and RBE-SI) based on it. One of their important features is that in most of the cases the resulting rule base is sparse, which ensures a reduced system complexity. However, this character implies the application of Fuzzy Rule Interpolation (FRI) based reasoning techniques. Both methods can be used for the improvement of the design of fuzzy logic controllers for non-linear systems introduced in [11].

The experimental results with two FRI based inference techniques (LESFRI and FRIPOC) showed that good system performance can be obtained by both of the RBE-DSS and RBE-SI methods. The grouping of the rules (see fig. 8) in case of the SISO system proved the expectation that the coverage by rules of the functions' turning points plays a significant role in system performance. The improvement of the tuning algorithm, the influence of the selection of its parameters, and the applicability of the methods in case of other FRI techniques are subject to further research work.

The Matlab implementation of the presented system generation methods can be freely downloaded from [12]. This website is dedicated to a fuzzy rule interpolation Matlab toolbox development project (introduced in [5]) aiming the implementation of various FRI techniques.

REFERENCES

[1] A. Chong, "Constructing Sparse and Hierarchical Fuzzy Rulebases," PhD Thesis, Murdoch University, Perth, W.A., 2004.

[2] Z. C. Johanyák and S. Kovács, "Fuzzy Rule Interpolation Based on Polar Cuts," Computational Intelligence, Theory and Applications, Springer Berlin Heidelberg, 2006, pp. 499-511.

[3] Z. C. Johanyák and S. Kovács, "Fuzzy Rule Interpolation by the Least Squares Method," 7th International Symposium of Hungarian Researchers on Computational Intelligence (HUCI 2006), November 24-25, 2006 Budapest, Hungary, pp. 495-506.

[4] Z. C. Johanyák and S. Kovács, "The effect of different fuzzy partition parameterization strategies in gradient descent parameter identification," 4th International Symposium on Applied Computational Intelligence and Informatics (SACI 2007), May 17-18, 2007 Timisoara, Romania, pp. 141-146.

.[5] Z. C. Johanyák, D. Tikk, S. Kovács, K. W. Wong "Fuzzy Rule Interpolation Matlab Toolbox – FRI Toolbox," Proc. of the IEEE World Congress on Computational Intelligence (WCCI'06), 15th International Conf. on Fuzzy Systems (FUZZ-IEEE'06), July 16–21, 2006, Vancouver, BC, Canada, pp. 1427–1433, Omnipress.

[6] B. Kosko, "Optimal fuzzy rules cover extrema," International journal of interlligent systems, vol. 10, 249 - 255, 1995.

[7] I. Perfilieva, "Fuzzy function as an approximate solution to a system of fuzzy relation equations," Fuzzy Sets and Systems, 147(3):363-383, 2004.

[8] D. Shepard, "A two dimensional interpolation function for irregularly spaced data," Proc. Of the 23rd ACM International Conf., 1968, pp. 517–524.

[9] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," IEEE Transactions on Fuzzy Systems, vol. 1, pp. 7-31, 1993.

[10] D. Tikk, T. D. Gedeon, L. T. Kóczy and G. Bíró, "Implementation details of problems in Sugeno and Yasukawa's qualitative modeling," . Research Working Paper RWP-IT-02-2001, School of Information Technology, Murdoch University, Perth, W.A., 2001.

[11] M. L. Tomescu, "Fuzzy Logic Controller for the Liénard System," 4th International Symposium on Applied Computational Intelligence and Informatics (SACI 2007), May 17-18, 2007 Timisoara, Romania, pp. 129-133.

[12] http://fri.gamf.hu