Takagi-Sugeno Systems with Improved Interpretability and Interpolation Using a Complementary Interpolation Model

Andri Riid, and Ennu Rüstern, *Member, IEEE*

Abstract—**The paper deals with the interpretability problem of 1^{st} order Takagi-Sugeno systems and interpolation issues in particular. Interpolation improvement is carried out by a corrective secondary model (essentially a black box) complementing the primary (interpretable) model. Optimization technique for this two-model configuration is developed. Experimental results suggest that this approach achieves a better accuracy-interpretability tradeoff than the methodologies currently in use.**

Index Terms— **Fuzzy systems, inference mechanism, modeling, interpolation.**

I. Introduction

THE fuzzy inference system proposed by Takagi and Sugeno in [1] (1^{st} order TS system, in short) is a powerful tool for modeling complex nonlinear systems. TS modeling is a multimodel approach in which linear submodels associated with TS rules are combined to describe the global behavior of the system. TS rules have high degrees of freedom to improve their performance that consequently makes it possible to express complicated behaviors with a small number of rules.

This, however, has a downside in overparameterization that results in nonuniqueness in model structure by what different parameter vectors may yield the same input/output behavior. Moreover, large perturbations of consequent parameters may have a very small effect on global approximation. This property - which makes TS systems effective in modeling from purely numerical aspect - implies that local models rarely admit valid interpretation as local linearizations of the modeled nonlinear system (i.e. are non-transparent to interpretation).

Transparency of a local model allows one to gain insight into the global behavior of the system; interpretation in terms of linearizations is also useful in system analysis and local control design, for example in gain-scheduled control [2]. It, however, becomes difficult to obtain transparent as well as accurate TS systems because of the trade-off between these requirements in fuzzy logic systems [3] that can be quite drastic [4].

The system parameters in TS modeling are usually obtained by the use of global learning strategies e.g. ANFIS [5] (a combination gradient descent and least squares estimator (for consequent parameter identification)) that minimize quadratic global cost function. It has been shown that certain approaches (e.g. weighted least squares method [6]) that promote competition between the rules (in contrast to global techniques where learning has cooperative character) can improve interpretability of the model.

Local (or weighted) least squares method estimates the parameters of the local models separately therefore each fuzzy rule is encouraged to produce the whole of the output (rather than a component) in a limited region of input space (where it has nonzero degree of fulfillment) thus each extracted fuzzy rule acts like an independent model related to a subset of training data. In result, weighted parameter estimation gives an optimal estimate of the local models but does not provide an optimal fuzzy model in terms of minimal modeling error because the aggregation of the rules is not taken into account. This problem is handled in [7] where combined global-local least squares method is proposed that aims at striking a good tradeoff between the global approximation and local interpretation.

Another way to promote competition between the rules is to reduce their interpolation (interaction). This can be done either indirectly - by imposing a fixed overlap limit on input membership functions (MFs) contributing to a rule [8] - or directly - by exponenting the rule fulfillment degrees in input space [9]. The basic disadvantage of these techniques is that interpolation reduction generally leads to a model with reduced approximation potential.

All these techniques, however, deal with only a part of the problem (overparameterization) and neglect the second reason behind accuracy-transparency tradeoff - undesirable properties of TS rule interpolation mechanism.

This paper focuses on the latter problem and introduces a two-model system configuration including an additional secondary model complementing the primary (interpretable) one to cancel out the undesired effects of TS inference. To use this approach in practice, an optimization method is developed. The basic advantage of proposed approach when compared to other interpolation improvement schemes [10, 11] is that a replacement of original TS interpolation mechanism is not required and that the technique can be

A. Riid and E. Rüstern are with the Department of Computer Control of Tallinn University of Technology, Ehitajate tee 5, Tallinn 19086 Estonia (phone: +372-6-202-113; fax: +372-6-202-101; e-mail: andri@dcc.ttu.ee).

applied to multidimensional modeling problems without further effort.

II. SYSTEM DEFINITION

We consider a multi-input/single-input first-order TS fuzzy system consisting of R rules (1), where A_{ir} denote the linguistic labels of the ith input variable ($i = 1...N$), associated with the r^{th} rule, having one-to-one correspondence with normal and convex MFs μ_{ir} in the inference function (2); p_{0r}, p_{ir} denote the consequent parameters of the r^{th} rule and x_i denotes the numerical value of the i^{th} input variable.

$$\text{IF } x_1 \text{ is } A_{1r} \text{ AND...AND } x_i \text{ is } A_{ir} \text{ AND...AND } x_N \text{ is } A_{Nr} \text{ THEN } y_r = p_{0r} + ... + p_{ir}x_1 + ... p_{Nr}x_N, \quad (1)$$

$$y = \frac{\sum_{r=1}^{R} y_r \prod_{i=1}^{N} \mu_{ir}(x_i)}{\sum_{r=1}^{R} \prod_{i=1}^{N} \mu_{ir}(x_i)} = \frac{\sum_{r=1}^{R} \tau_r y_r}{\sum_{r=1}^{R} \tau_r} \quad (2)$$

III. INTERPOLATION PROBLEM

Even though the techniques mentioned in the introduction may help us to extract local models with much stronger local context than global techniques, the very essential interpolation problem inherited from the use of (2) is neglected which needs special consideration. The essence of the problem is illustrated with a simple example that follows. Interpolated global output from two neighboring transparent local models (transparent to admit valid interpretation) is quite different from the one that one would be expecting intuitively (Fig. 1). On the other hand, the only way to produce the desired smooth interpolated output with (2), is to sacrifice interpretability – two interpolating local models give substantially biased local linear estimates of the inferred global function. (Fig 2).

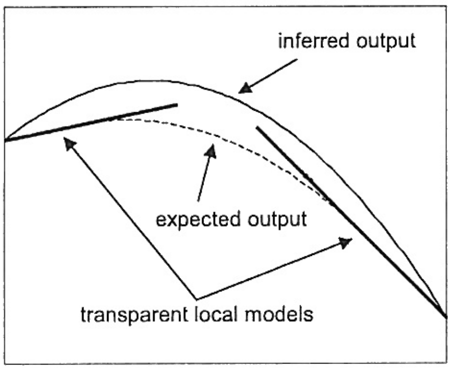

Fig. 1. Biased global output from transparent local models.

Tradeoff between interpretability and accuracy is very evident in this example and there is no straightforward solution to offer other than certain compromise – we need to insert an additional rule that on one hand, would improve interpolation between two existing interpretable rules but on the other hand, interpretability of the inserted rule must be sacrificed. The latter deficiency, however, is acceptable if information about interpretability of any given rule is known (non-transparency can be localized). We accomplish that by organizing interpretable and interpolating rules into separate models.

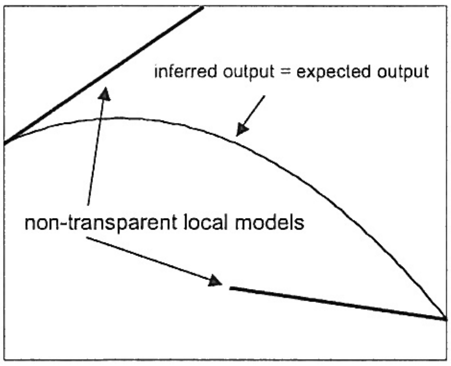

Fig. 2. Biased local models to produce acceptable global output

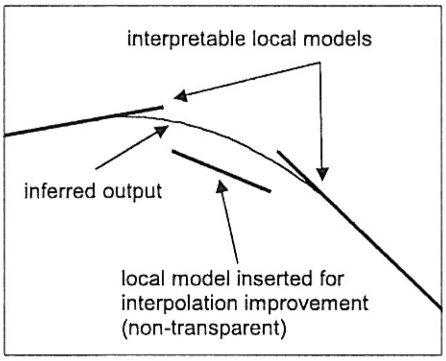

Fig. 3. Interpolation improvement by rule insertion

IV. SYSTEM CONFIGURATION

To distinguish between interpretable and interpolating rules they are divided between two submodels – primary (interpretable) and secondary (interpolative).

$$\mu_i^s(x_i) = \begin{cases} 0, & \text{if } x_i < a_i^s \text{ or } x < d_i^s \\ 2\left(\frac{x_i - a_i^s}{b_i^s - a_i^s}\right)^2, & \text{if } a_i^s \le x_i \le \frac{a_i^s + b_i^s}{2} \\ 1 - 2\left(\frac{b_i^s - x_i}{b_i^s - a_i^s}\right)^2, & \frac{a_i^s + b_i^s}{2} \le x_i \le b_i^s \\ 1, & \text{if } b_i^s < x_i \le c_i^s \\ 1 - 2\left(\frac{x_i - c_i^s}{d_i^s - c_i^s}\right)^2, & c_i^s \le x_i \le \frac{c_i^s + d_i^s}{2} \\ 2\left(\frac{d_i^s - x_i}{d_i^s - c_i^s}\right)^2, & \frac{c_i^s + d_i^s}{2} \le x_i \le d_i^s \end{cases} \quad (3)$$

Each MF μ_i^s (i = 1...N, s = 1 ...S_i (S_i > 1)) of the i^{th} input variable of the primary model is defined by a set of four parameters – a_i^s, b_i^s, c_i^s and d_i^s – and the underlying function (3).

Note that the following constraints apply:

$$\mu_i^s : a_i^s = c_i^{s-1}, d_i^s = b_i^{s+1}, \tag{4}$$

except when s = 1: $a_i^s = b_i^s$ and when $s = S_i$ $d_i^s = c_i^s$ (Fig. 4).

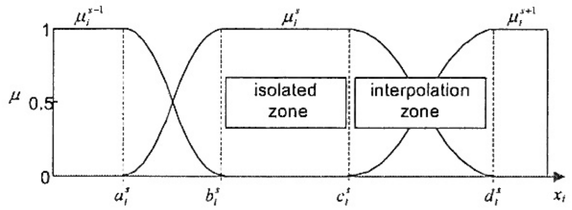

Fig. 4. Input partition of the primary model.

MFs γ_i^t (t = 1, ..., T_i ($T_i = 2S_i$ - 1)) of the secondary model use the same underlying function, (to avoid the confusion, its four parameters are denoted by $\alpha_i^t, \beta_i^t, \chi_i^t$ and δ_i^t) and their parameters are derived directly from the input partition of the primary model (Fig. 5).

$$\gamma_i^t \equiv \mu_i^s \text{, if } t=2s-1 \; (s = 1, \ldots, S_i). \tag{5}$$

$$\gamma_i^t : \alpha_i^t = c_i^s, \delta_i^t = d_i^s, \beta_i^t = \chi_i^t = \frac{c_i^s + d_i^s}{2} \text{, if } t=2s \; (s = 1, \ldots, S_i - 1). \tag{6}$$

Note that trapezoid type of membership function can be used instead of (3), with (4-6) remaining valid.

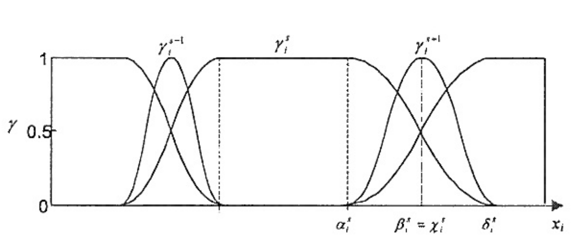

Fig. 5. Input partition of the secondary model

Primary model has fully defined combinatorial rulebase (which means that all possible combinations of input MFs are described by it, bringing the total number of rules to $P = \prod_{i=1}^{N} S_i$). From the rulebase of the secondary model, initially obtained in the similar manner, however, the rules that satisfy (7) are excluded because they are already described by the primary model.

$$\forall i, (\chi_{ir} - \beta_{ir} > 0) \text{, } r = 1, \ldots, R, \left(R = \prod_{i=1}^{N} T_i\right), \tag{7}$$

where χ_{ir} and β_{ir} denote the parameters of the MF of the i^{th} input variable associated with the r^{th} initial rule. After rule filtering we should have $Q = R - P$ rules in the secondary model.

Global output of the whole system is computed by

$$y = y_1 + y_2 = \sum_{p=1}^{P} y_p \tau_p \Big/ \sum_{q=1}^{Q} \tau_q + \sum_{q=1}^{Q} y_q \tau_q \Big/ \sum_{q=1}^{Q} \tau_q \tag{8}$$

Note that the type of the consequent function y_q of the secondary model does not necessarily need to be a 1$^{\text{st}}$ order function as in the primary model. In some cases higher order consequent functions can be more effective because of their extended interpolative power, or on the other hand, sometimes it may be just sufficient to use a constant consequent (0$^{\text{th}}$ order function). In particular, the following consequent function types (besides the constant ($y_q = p_{0q}$) and the original one given in (2)) have been considered in current paper:

$$y_q = p_{0q} + \sum_{i=1}^{N} p_{iq}^{(1)} x_i + \sum_{i=1}^{N} p_{iq}^{(2)} x_i^2, \tag{9}$$

$$y_q = p_{0q} + \sum_{i=1}^{N} p_{iq}^{(1)} x_i + \sum_{i=1}^{N} p_{iq}^{(2)} x_i^2 + \sum_{i=1}^{N} p_{iq}^{(3)} x_i^3, \tag{10}$$

V. Optimizing the System

The (supervised) optimization method described in this section requires a set of training data consisting of K training samples $[x_k \; y(k)]$, predefined number of interpretable rules (P) and is based on the reasoning that if isolation of the rules promotes interpretability of the system, its approximation capacities depend heavily on the level of rule interpolation that takes place within the system. Therefore, initially we have an interpretable model with high rule isolation level and by gradually increasing interpolation zones in appropriate manner we should ultimately reach a satisfying result. Initialization, consequent and antecedent parameter identification and completion of the optimization algorithm are described in the following sections.

A. Input partition initialization

Initialization of input MF parameters is based on H cluster centers $[x_{1h}, x_{2h}, \ldots, x_{Nh}, y_h]$ extracted from training data by Gustafson-Kessel clustering algorithm [12]. Given a preset interpolation/isolation ratio η = [0, 1], input partition of the primary model is initialized as follows (i = 1, ..., N, h = 1, ..., H – 1)

$$c_i^h = a_i^{h+1} = x_i^h + \frac{(x_i^{h+1} - x_i^h)}{2}(1-\eta) \tag{11}$$

$$d_i^h = b_i^{h+1} = c_i^h + (x_i^{h+1} - x_i^h)\eta \quad (12)$$
$$a_i^1 = b_i^1 = x_i^{\min}, c_i^H = d_i^H = x_i^{\max}$$

Note that $\forall\ i = 1, \ldots, N, (S_i = H)$.

Input partition of the secondary model is constructed from the primary one according to (5) and (6).

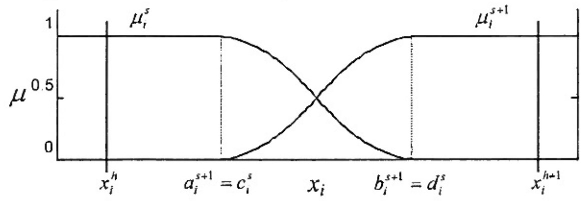

Fig. 6. Extraction of MFs of the primary model from cluster centers.

B. Consequent Parameter Identification

Consequent parameters (on the assumption that the rulebase and input partition of the model are given) of TS systems are generally identified by (global) least squares estimator. Using the notations $X_e = [1\ \mathrm{X}]$ with rows $[1\ x_k]$, $x_k = [x_1(k), x_2(k), \ldots, x_N(k)]$ and $\Gamma = [W_1 X_e, W_2 X_e, \ldots, W_r X_e, \ldots, W_R X_e]$, where

$$W_r = \begin{bmatrix} \beta_r(1) & 0 & \ldots & 0 \\ 0 & \beta_r(2) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \beta_r(K) \end{bmatrix}, \quad (13)$$

where $\beta_r(k) = \tau_r(k) \Big/ \sum_{r=1}^{R} \tau_r(k)$ (normalized rule activation degree), (2) becomes equivalent to a least squares problem $y = \Gamma\theta + \varepsilon$, where ε is the approximation error and has the solution, given by (14)

$$\theta = \left[\Gamma^T \Gamma\right]^{-1} \Gamma^T y, \quad (14)$$

where $\theta = [p_1, \ldots, p_r, \ldots, p_R]^{\mathrm{T}}, p_r = [p_{0r}, p_{1r}, \ldots p_{ir}, \ldots p_{Nr}]$

Note that for computational reasons recursive Kalman filter [1] or Moore-Penrose pseudoinverse computed via singular value decomposition [7] are viable options to compute θ.

To obtain consequent parameters for submodels, the following two-step procedure is used: in first pass consequent parameters of both models are identified together by using (14), with $\Gamma = [\Gamma_1, \Gamma_2]$, $\Gamma_1 = [W_1 X_e, \ldots, W_p X_e, \ldots, W_P X_e]$, $\Gamma_2 = [W_1 X_e, \ldots, W_q X_e, \ldots, W_Q X_e]$, $\theta = [\theta_1, \theta_2]$, $\theta_1 = [p_1, \ldots, p_p, \ldots, p_Q]^{\mathrm{T}}$, $\theta_2 = [p_1, \ldots, p_q, \ldots, p_Q]^{\mathrm{T}}$,

Regardless of the actual configuration of the secondary model, we are making the assumption that it is a 1st order TS system just as the primary one.

In the second pass, however, consequent parameters of the secondary model will be properly re-identified, using

$$\theta_2 = \left[\Gamma_2^T \Gamma_2\right]^{-1} \Gamma_2^T (y - \Gamma_1 \theta_1) \quad (15)$$

Note that the formation of Γ_2 and what is contained in θ_2 depends on the order of consequent function and in case of 0th order function

$$x_k = [\,], p_q = p_{0q}. \quad (16)$$

In case of 2nd order function (9)

$$x_k = [x_1(k), (x_1(k))^2, x_2(k), (x_2(k))^2, \ldots, x_N(k), (x_N(k))^2] \quad (17)$$
$$p_q = \left[p_{0q}, p_{1q}^{(1)}, p_{1q}^{(2)}, \ldots, p_{iq}^{(1)}, p_{iq}^{(2)}, \ldots, p_{Nq}^{(1)}, p_{Nq}^{(2)}\right]$$

and in case of 3rd order function (10)

$$x_k = [x_1(k), (x_1(k))^2, (x_1(k))^3, \ldots, x_N(k), (x_N(k))^2, (x_N(k))^3] \quad (18)$$
$$p_q = \left[p_{0q}, p_{1q}^{(1)}, p_{1q}^{(2)}, p_{1q}^{(3)} \ldots, p_{iq}^{(1)}, p_{iq}^{(2)}, p_{iq}^{(3)}, \ldots, p_{Nq}^{(1)}, p_{Nq}^{(2)}, p_{Nq}^{(3)}\right]$$

C. Input partition optimization

After the model is initialized and consequent parameters are identified, we proceed with iterative input partition optimization. In each step of the cycle, k^{th} training sample (of all K samples) responsible for the maximum error $\varepsilon(k)$ is identified. Each i^{th} component of this sample will then be projected onto respective axis of x_i and fired MFs ($\mu_i^s(x_i(k)) > 0, \gamma_i^t(x_i(k)) > 0$) of both models will be updated according to the following rules. There are three possibilities (Figs. 7-8).

a) $x_i(k)$ falls into the isolated zone of the primary model (i.e. only μ_i^s has nonzero firing degree);

b) $x_i(k)$ falls into the left side of interpolation zone (i.e. μ_i^s and μ_i^{s+1} have both nonzero firing degrees and $\mu_i^s(x_i(k)) > \mu_i^{s+1}(x_i(k))$);

c) $x_i(k)$ falls into the right side of interpolation zone (i.e. μ_i^{s-1} and μ_i^s have both nonzero firing degrees and $\mu_i^s(x_i(k)) > \mu_i^{s-1}(x_i(k))$).

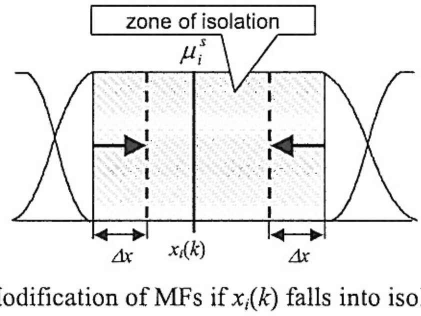

Fig. 7. Modification of MFs if $x_i(k)$ falls into isolated zone.

In case (a), the core ($c_i^s - b_i^s$) of μ_i^s (Fig. 7) is reduced from both sides by a preset value of Δx. Note that in order to satisfy conditions (4), some parameters of neighboring MFs also need to be updated.

In case (b), the core of μ_i^s is reduced from the right side and a_i^{s+1} has to be updated as well.

In case (c), the core of μ_i^s is reduced from the left side and d_i^{s-1} has to be updated as well.

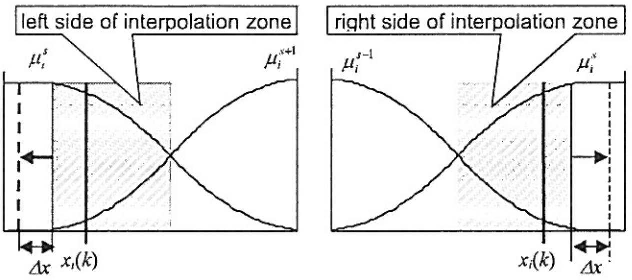

Fig. 8. Modification of MFs if $x_i(k)$ falls into interpolation zone.

Obviously, respective input MFs of the secondary model also need to be updated on the basis of (5-6). To complete the training step, it is followed by consecutive application of (14) and (15). Optimization is finished if the stopping criterion (which may be a preset number of training epochs, preset error value or preset error change rate) becomes satisfied.

VI. RESULTS

This section presents three examples of function approximation to demonstrate how the proposed model configuration and the optimization algorithm deal with accuracy-transparency tradeoff. The first example is a function

$$y = 0.6\sin(\pi x) + 0.3\sin(3\pi x) + 0.1\sin(5\pi x)\,, \tag{19}$$

approximated from 201 data points placed at equal intervals in [-1, 1] of input space.

We model this function using models with 3, 5 and 9 rules and using different types (constant, 1st order, (9) and (10)) of consequent functions. The results, evaluated with modeling root mean square error (RMSE) and final interpolation/isolation rate

$$\eta_{final} = 1 - \frac{1}{x_i^{\max} - x_i^{\min}} \sum_{s=1}^{S_i} b_i^s - c_i^s\,, \tag{20}$$

are given in Table 1, where L denotes the number of training steps necessary to obtain minimum value of RMSE. Note that $\eta = 0.5$ and $\Delta x = 0.02$ in all experiments.

The question here, as it turns out, is not so much how to obtain small RMSE (as it appears, the error falls into the same range, independent of P, except for some experiments with 0th order consequent functions in the secondary model) but how much interpretability we need to sacrifice (expressed by η_{final}). In present case, 2nd order consequent function seems to be the optimal choice. One must take into account, however, that computation of higher order function parameters requires more computational power and increases model complexity. Increase of P similarly pays back with more interpretability and faster convergence but must be weighted against system complexity.

TABLE I
MODELING RESULTS OF (19)

P (Q)	rank of y_q	RMSE	η_{final}	L
3 (5)	0 / 1	0.0071/0.0071	0.9679/0.9679	66/62
3 (5)	2 / 3	0.0071/0.0085	0.9679/0.9579	61/60
5 (9)	0 / 1	0.0143/0.0087	0.6000/0.6900	16/21
5 (9)	2 / 3	0.0064/0.0063	0.6900/0.6900	21/21
9 (17)	0 / 1	0.1180/0.0059	0.5100/0.5900	7/15
9 (17)	2 / 3	0.0045/0.0044	0.5900/0.5900	13/13

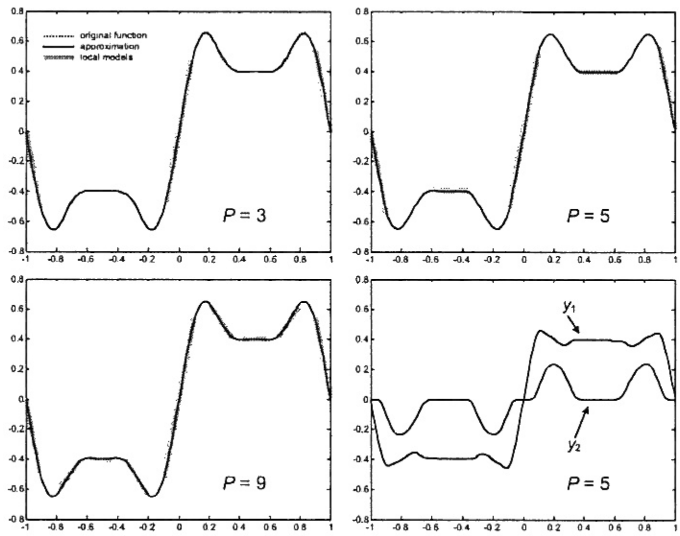

Fig. 9. Approximation of (19). Lower right figure depicts the outputs of both models, respectively.

In the second example the algorithm has to deal with noisy/corrupt motorcycle crash data taken from [13]. The iterative approach is not appropriate here because a single strongly biased outlier can outbalance it. Therefore the procedure is skipped and the input partition obtained by initialization serves as the final partition. Secondly, to ensure smooth interpolation (because of noise secondary model may obtain non-smooth interpolating local models), linear coefficients p_{iq} of the 1st order secondary model are computed as the average of respective coefficients from the relevant (active in the same region) rules of the primary model so that

$$p_{iq} = \underset{\tau_p(c_q)>0}{avg}\,(p_{ip})\,, \tag{21}$$

where $c_q = [(\beta_{1r} + \chi_{1r})/2, \ldots, (\beta_{Nr} + \chi_{Nr})/2]$ is the center of qth rule in input space. The remaining unknown parameters of the secondary model (p_{0q}) are identified using least squares estimator, so that

$$\left[p_{01},\ldots,p_{0Q}\right]^{T} = \left[\Gamma_2^{\,T}\Gamma_2\right]^{-1}\Gamma_2^{\,T}(y-\Gamma_1\theta_1-\Gamma_2\varphi)\,, \tag{22}$$

where $\varphi = [p_{11}, \ldots, p_{N1}, \ldots, p_{1q}, \ldots, p_{iq}, \ldots, p_{Nq}, \ldots, p_{1Q}, \ldots, p_{NQ}]^{T}$ contains the coefficients computed by (21).

Using the settings $P = 4$, $Q = 3$, $\eta = 0.5$ and applying (21-22) we see from Fig. 10 that four extracted interpretable local models capture the essence of the process under consideration. Moreover, in comparison with the results obtained in [7] with different combinations of global and local least squares, our result (MSE = 467.68) is outperformed only by a pure global least-squares approach (MSE = 460.62) with very poor interpretability, even then, in our experiment overall 7 rules are used, compared to 8 in [7].

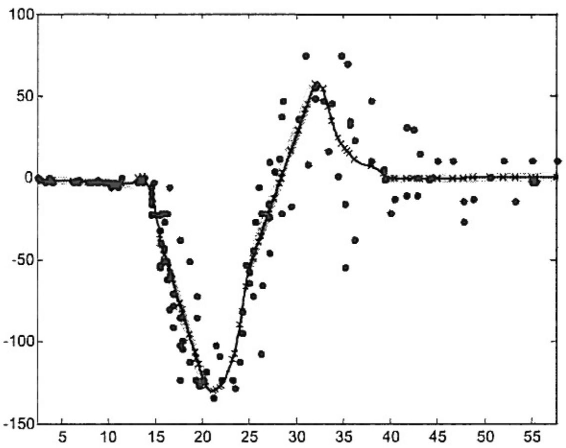

Fig. 10. Approximation of motorcycle data. Legitimate local linear models are depicted with gray lines.

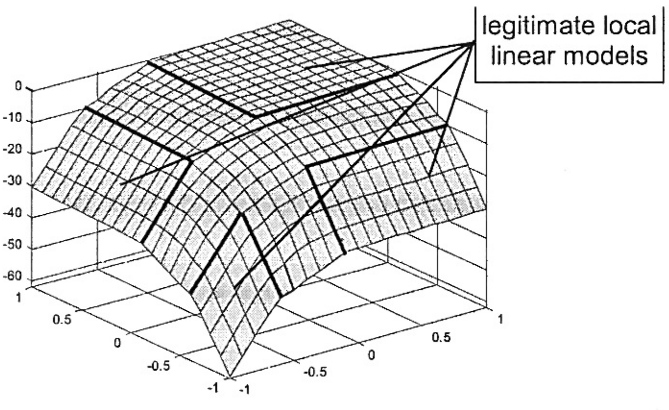

Fig. 11. Approximation of 3D data.

Finally, an approximation of 3D function

$$y = (x_1 - 1)^5 + (x_2 - 1)^5 \quad (21)$$

is presented, using 441 data points spaced equidistantly in [-1, 1]×[-1, 1] of input space.

After 30 steps of training (*P = 4, Q = 5,* $\eta = 0.02$*,* Δx = 0.02) using (9) in the secondary model, we obtain RMSE = 0.5596, which is about 2.5 times less than the number (1.3927) obtained in [10] (important note: the result in [10] is obtained with just four rules). The current approach, however, is more universal, as it does not assume anything about the type of rule interpolation unlike the method described in [10].

VII. Conclusions

We have introduced a two-model system configuration to improve both interpretability and interpolation in TS modeling and developed the optimization method to benefit from the new features. The experiments show that the proposed approach is able to extract legitimate TS local models from data. The results evaluated in terms of RMSE and interpolation/isolation ratio depend on the number and initialization of interpretable rules. Special measures to deal with noisy data are suggested.

References

[1] T. Takagi, and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," IEEE Trans. Syst., Man, Cybernetics, vol. SMC-15, no. 1, pp. 116-132, 1985.

[2] P. Viljamaa, "Fuzzy Gain Scheduling and Tuning of Multivariable Fuzzy Control - Methods of Fuzzy Computing in Control Systems," Ph.D. dissertation, Automation and Control Institute, Tampere Univ. of Technology, 2002.

[3] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena, Ed. *Interpretability Issues in Fuzzy Modeling*. Berlin: Springer, 2003.

[4] J. Abonyi, *Fuzzy Model Identification for Control*. Boston: Birkhäuser, 2003.

[5] J.-S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. System, Man, Cybernetics*, vol. 23, no.3, pp. 665-685, 1993.

[6] W.S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *J. Amer. Statistical Assoc.*, vol. 74, pp. 829–836, 1979.

[7] J. Yen, L. Wang, and C.W. Gillespie, "Improving the Interpretability of TSK Fuzzy Models by Combining Global Learning and Local Learning," *IEEE Trans. Fuzzy Systems*, vol. 6, no. 4, pp. 530-537, 1998.

[8] A. Riid, R. Isotamm, and E. Rüstern, "Transparency Enhancement of 1st order TS systems: Promoting the Competition Between the Rules by Controlling the Overlap of Input Fuzzy Sets," in *Proc. 8th Biennal Baltic Electronic Conference*, Tallinn, 2002, pp. 137-140.

[9] A. Riid, "Transparent Fuzzy Systems: Modeling and Control," Ph.D. dissertation, Dept. of Comp. Control, Tallinn Technical Univ., Tallinn, 2002.

[10] R. Babuska, C. Fantuzzi, U. Kaymak, and H.B. Verbruggen, "Improved inference for Takagi-Sugeno models," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, New Orleans, pp. 701-706, 1996.

[11] B. Butkiewicz, "Simple modification of Takagi-Sugeno Model," in *Neural Networks and Soft Computing*, L. Rutkowski, and J. Kacprzyk, Eds. Heidelberg: Physica-Verlag, 2003, pp. 504-509.

[12] D.E. Gustafson, and W.C. Kessel, Fuzzy clustering with a fuzzy covariance matrix." in *Proc. IEEE Conf. Decision and Control*, San Diego, pp. 761-766, 1979.

[13] W.Hardle, *Applied Nonparametric Regression*. Cambridge, MA: Cambridge Univ. Press, 1990.