Structure Identification of Dynamical Takagi-Sugeno Fuzzy Models by Using LPV Techniques

Matthias Kahl and Andreas Kroll

Abstract In this paper the problem of order selection for nonlinear dynamical Takagi-Sugeno (TS) fuzzy models is investigated. The problem is solved by formulating the TS model in its Linear Parameter Varying (LPV) form and applying a recently proposed Regularized Least Squares Support Vector Machine (R-LSSVM) technique for LPV models. In contrast to parametric identification approaches, this non-parametric method enables the selection of the model order without specifying the scheduling dependencies of the model coefficients. Once the correct model order is found, a parametric TS model can be re-estimated by standard methods. Different re-estimation approaches are proposed. The approaches are illustrated in a numerical example.

1 Introduction

Takagi-Sugeno fuzzy models (Takagi and Sugeno, 1985) which permit to approximate nonlinear systems by a weighted superposition of local linear models have been successfully utilized in many industrial applications. Besides

Matthias Kahl · Andreas Kroll
Department of Measurement and Control, University of Kassel, D-34109 Kassel, Germany
✉ matthias.kahl@mrt.uni-kassel.de
✉ andreas.kroll@mrt.uni-kassel.de

Archives of Data Science, Series A
(Online First)
KIT Scientific Publishing
Vol. 5, No. 1, 2018
DOI: 10.5445/KSP/1000087327/19
ISSN 2363-9881

their universal approximation property, their structure with local models permits the transfer of linear controller-design methods to a nonlinear framework, e.g. by gain-scheduling or parallel distributed controllers (Wang et al, 1995). In order to obtain a model which performs well for the considered task an appropriate model structure has to be chosen. In data-driven modeling of dynamical systems (system identification), this involves the choice of relevant physical variables and the individual time lags of each variable as well as the choice of the model terms describing the functional relationship between the system input and output, resulting in a large set of potential model candidates. The structure selection problem of TS models consists of 3 parts:

i) The choice of appropriate scheduling and input variables,

ii) the partitioning of the scheduling space by an appropriate parameterization of the fuzzy basis functions of a predefined type as well as the choice of the number of local models, and

iii) the selection of a suitable local model structure.

While the choice of appropriate system inputs is mostly restricted by the modeling exercise or results from prior knowledge, the selection of the scheduling variables may be more challenging as it mainly determines the nonlinear behavior of the model. As the output of a TS model is nonlinear in the parameters of its basis functions, a nonlinear optimization problem has to be solved in order to partition the scheduling space. Alternatively, heuristic construction strategies were proposed like grid partitioning, data-point-based methods, clustering-based approaches or heuristic tree construction algorithms like LOLIMOT with individual advantages and drawbacks (see, e.g., Nelles, 2001). Once the parameterization of the membership functions is known, the remaining optimization problem regarding the local model parameters $\theta_{i,\mathrm{LM}}$ is linear in the parameters for local linear regression models. In this contribution, autoregressive models with exogenous input (ARX) are used as local model class in order to model nonlinear dynamical systems. Hence, the choice of the local model structure coincides with the choice of the dynamical order of the considered system.

In order to find a suitable local model structure, a higher-level wrapper approach can be used for a given choice of the partitioning strategy (see Kahl et al, 2015 for an overview of different structure selection approaches in the system identification context), which assesses the usefulness of a considered local regressor subset by means of the approximation or generalization properties of a model. This is done by comparing models, which are built of different regressor subsets and can be extended easily to the selection of scheduling variables and hyper-parameters. However, when every possible subset has to be evaluated by exhaustive search, the resulting high-dimensional combinatorial optimization problem may be intractable and greedy strategies like stepwise selection have to be used (see, e.g., Hong and Harris, 2001; Belz et al, 2017). Alternatively, with the aim of sparse local models, the original combinatorial optimization problem can be approximated by lasso like convex relaxation. For the class of TS models also a grouped lasso regularization was used by Luo et al (2014) in order to force sparseness in the number of local models by exploiting the block-structured representation of TS models. With the same aim, Lughofer and Kindermann (2010) introduced a rule weighting, i.e. the inclusion of an additional weighting factor into the fuzzy basis functions, and forced it to zero by incorporating a l_1 penalty into a nonlinear optimization problem. Additionally, they applied a sparse estimator for local parameter estimation.

All approaches have in common that the partitioning and thereby the fuzzy basis functions have to be determined in advance or in a successive manner and are, therefore, biased by the individually chosen partitioning strategy. Recently, Piga and Tòth (2013) and Mejari et al (2016) developed regularization approaches based on least squares support vector machines allowing to determine the order of LPV-ARX models without the a priori specification of the scheduling dependencies of the model coefficients. The approach from Mejari et al (2016) is used in this contribution to solve the order determination problem iii) for TS fuzzy models formulated in its LPV form to avoid solving of a nonlinear optimization problem to find a suitable partitioning of the TS model.

2 Dynamical TS-fuzzy Model

2.1 Identification Problem

A Takagi-Sugeno fuzzy model consists of $c \in \mathbb{N}_+$ superposed local models $\hat{y}_i(k) = f_i(\theta_{i,\text{LM}}, \varphi(k)) : \mathbb{R}^n \to \mathbb{R}$ weighted by their corresponding fuzzy basis functions $\phi_i(z(k)) : \mathbb{R}^{n_z} \to [0, 1]$, depending on the n_z scheduling variables $z(k) = [z_1(k) \dots z_{n_z}(k)]^\top \in \mathbb{R}^{n_z}$, respectively:

$$\hat{y}(k) = \sum_{i=1}^{c} \phi_i(z(k)) \cdot \hat{y}_i(k), \tag{1}$$

with the discrete time k. As local model type ARX models of the form

$$\hat{y}_i(k) = \sum_{r=1}^{n} \theta_{i,r,\mathrm{LM}} \cdot \varphi_r(k) \tag{2}$$

are considered in this paper. $\varphi_r(k)$ is the r-th element of the vector

$$\varphi(k) = [-y(k-1) \ldots - y(k-n_y), u(k-T_\tau) \ldots u(k-n_u-T_\tau)]^\top, \tag{3}$$

$n = n_y + n_u + 1$, $\theta_{i,r,\mathrm{LM}}$ is the r-th element of the local parameter vector

$$\theta_{i,\mathrm{LM}} = [\theta_{i,y}, \theta_{i,u}]^\top \in \mathbb{R}^n, \tag{4}$$

and T_τ is a potential dead time. $\theta_{i,y} \in \mathbb{R}^{n_y}$ is the parameter vector corresponding to the lagged values of the measured output signal $y(k) \in \mathbb{R}$ of the system and $\theta_{i,u} \in \mathbb{R}^{n_u}$ corresponds to the lagged values of the measured input signal $u(k) \in \mathbb{R}$.

The fuzzy basis functions $\phi_i(z(k))$ define a validity region of the corresponding local models. The basis functions are defined by

$$\phi_i(z(k)) = \frac{\mu_i(z(k))}{\sum_{j=1}^{c} \mu_j(z(k))}, \tag{5}$$

with the membership functions (MF) $\mu_i(z(k))$. Typical types of membership functions are Gaussian, trapezoidal or clustering-based ones (see, e.g., Kroll, 1996; Babuška, 1998). Trapezoidal membership functions have the advantage of easier interpretation and local support. However, they suffer from the curse of dimensionality as they are univariate and can be applied axis aligned only. Multivariate Gaussian or clustering-based membership functions can permit a better adjustment of the partitioning for multivariate problems, such that the identification approach used in this contribution can be easily scaled to higher dimensions of the scheduling space. Furthermore, they can directly be obtained from clustering. But, they are harder to interpret and have no local support. In order to be analogous to the LSSVM approach, in this contribution, Gaussian membership functions are used

$$\mu_i(z(k)) = \exp\left(-\frac{1}{2}\frac{\| z(k) - v_i \|_2^2}{\sigma_i^2}\right), \tag{6}$$

where $v_i \in \mathbb{R}^{n_z}$ represents the partition's prototype and $\sigma_i \in \mathbb{R}_+$ specifies the width of the Gaussian function aggregated in the parameter vector $\theta_{i,\mathrm{MF}}$, so that $\mu_i(z(k)) = \mu_i(\theta_{i,\mathrm{MF}}, z(k))$.

For given $\theta_{i,\mathrm{MF}}$, $i = 1, \dots, c$, the local model parameters $\theta_{i,\mathrm{LM}}$ can be estimated by introducing $y = [y(1) \dots y(N)]^\top \in \mathbb{R}^N$, $\varphi = [\varphi(1) \dots \varphi(N)]^\top \in \mathbb{R}^{N\times n}$, and the extended regression matrix $\Lambda = [\Gamma_1\varphi \dots \Gamma_c\varphi]^\top \in \mathbb{R}^{N\times c\cdot n}$ with $\Gamma_i = \mathrm{diag}(\phi_i(z(1)) \dots \phi_i(z(N))) \in \mathbb{R}^{N\times N}$ for $N \in \mathbb{N}$ observations and applying ordinary least squares:

$$\hat{\theta}_{\mathrm{LM}} = \underset{\theta_{\mathrm{LM}}}{\operatorname{argmin}} \| y - \Lambda\theta_{\mathrm{LM}} \|_2^2, \tag{7}$$

with $\theta_{\mathrm{LM}}^\top = \left[\theta_{\mathrm{LM},1}^\top \dots \theta_{\mathrm{LM},c}^\top\right] \in \mathbb{R}^{n\cdot c}$. However, for an unknown partitioning of the scheduling space, the nonlinear optimization problem

$$\underset{\theta_{\mathrm{MF}},\theta_{\mathrm{LM}}}{\operatorname{argmin}} \sum_{k=1}^{N}\left(y(k) - \sum_{i=1}^{c}\phi_i(z(k), \theta_{\mathrm{MF}}) \cdot \hat{y}_i(k, \theta_{i,\mathrm{LM}})\right)^2, \tag{8}$$

in case of a quadratic cost function, has to be solved. Especially, in combination with a model-based structure selection approach one may have an issue with local minima or this can lead to an intractable problem due to computational complexity. In this contribution, the non-parametric approach described in Section 3.1 is used in order to determine the local structure of a TS model while avoiding such problems.

2.2 Analogies Between TS and LPV Models

According to Mejari et al (2016), an LPV-ARX model can be described by

$$\hat{y}(k) = \sum_{j=1}^{n_g}\vartheta_j(p(k)) \cdot x_j(k), \tag{9}$$

with $n_g = n_a + n_b + 1$, the scheduling variable $p(k) \in \mathbb{R}^{n_p}$, and $\vartheta_j(p(k))$ and $x_j(k)$ being the j-th component of

$$\vartheta(k) = [a_1(p(k)) \dots a_{n_a}(p(k)), b_0(p(k)) \dots b_{n_b}(p(k))]^\top \tag{10}$$

and

$$x(k) = [y(k-1) \dots y(k-n_a), u(k) \dots u(k-n_b)]^\top, \tag{11}$$

respectively. It is assumed that the coefficient functions $\vartheta_j(p(k))$ of the LPV model can be written as

$$\vartheta(p(k)) = \rho_j^\top \cdot \phi_j(p(k)), j = 1, \dots, n_g, \tag{12}$$

with the unknown parameter vector $\rho_j \in \mathbb{R}^{n_\mathrm{H}}$ and the feature maps ϕ_j mapping $p(k)$ to the n_H-dimensional feature space. In this way, and by including (12) in (9), the LPV model can be written in linear regression form:

$$\hat{y}(k) = \sum_{j=1}^{n_g} \rho_j^\top \cdot \phi_j(p(k)) \cdot x_j(k). \tag{13}$$

It is obvious that the scheduling variable of a TS-fuzzy system $z(k)$ and of an LPV system $p(k)$ can be viewed as equal. Furthermore, when assuming an identical structure of all local models for the TS model, also $\varphi(k) = x(k)$, $r = j$, and $n = n_g$ can be stated. By further incorporating (2) in (1):

$$\hat{y}(k) = \sum_{i=1}^{c} \sum_{r=1}^{n} \phi_i(z(k)) \cdot \theta_{i,r,\mathrm{LM}} \cdot \varphi_r(k), \tag{14}$$

and introducing the coefficient functions

$$\tilde{\theta}_r(z(k)) = \sum_{i=1}^{c} \phi_i(z(k)) \cdot \theta_{i,r,\mathrm{LM}}, \tag{15}$$

the TS-fuzzy model can be stated as a special case of the LPV-ARX model (9):

$$\hat{y}(k) = \sum_{r=1}^{n} \tilde{\theta}_r(z(k)) \cdot \varphi_r(k), \tag{16}$$

where the same fuzzy basis function for each regressor is used to parametrize the coefficient functions.

3 Identification Approach

3.1 LPV Model Estimation Using R-LSSVM

In order to select the order and dead time of the local models of a TS model characterized by n_y, n_u, and T_τ the Regularized Least Squares Support Vector Machine approach (R-LSSVM) introduced in Mejari et al (2016) is used. The approach is based on the method developed in Tòth et al (2011) and incorporates an additional regularization step in order to select the dynamical order of an LPV model.

The approach consists of 3 steps. In a first step, the approach proposed by Tòth et al (2011) is used to estimate the coefficient functions of an over-parametrized LPV model in a non-parametric manner. Starting from the LSSVM formulation for the estimation of the LPV model (13):

$$\begin{aligned} \underset{\rho,e}{\operatorname{argmin}}\ \mathcal{I}(\rho,e) &= \frac{1}{2}\sum_{j=1}^{n_g}\rho_j^\top\rho_j + \frac{\lambda}{2}\sum_{k=1}^{N}e^2(k) \\ \text{s.t. } e(k) &= y(k) - \sum_{j=1}^{n_g}\rho_j^\top\phi_j(p(k))x_j(k), \end{aligned} \tag{17}$$

with $\lambda \in \mathbb{R}_+$ being the regularization parameter of the primal problem, the Lagrangian dual problem associated with (17) is constructed:

$$\mathcal{L}(\rho,e,\alpha) = \mathcal{I}(\rho,e) - \sum_{k=1}^{N}\alpha_k\left[e(k) - y(k) + \sum_{j=1}^{n_g}\rho_j^\top\phi_j(p(k))x_j(k)\right], \tag{18}$$

with $\alpha_k \in \mathbb{R}$ being the Lagrangian multipliers. In the LSSVM setting, the kernel trick can be applied which enables the non-parametric description of the scheduling dependencies of the model coefficient functions. For a detailed description of the solution of (18), see Tòth et al (2011). The coefficient functions to be estimated are given as

$$\hat{\vartheta}_j(\cdot) = \rho_j^\top \phi_j(\cdot) = \sum_{k=1}^{N} \alpha_k K_j(p(k), \cdot) x_j(k), \tag{19}$$

where K_j is a positive definite kernel function. As stated in Tòth et al (2011) or Mejari et al (2016), a common choice of the kernel function is the Radial Basis Function (RBF):

$$K_j(p(k), p(m)) = \exp\left(-\frac{\| p(k) - p(m) \|_2^2}{\sigma_j^2}\right), \tag{20}$$

with the hyper-parameter σ_j specifying its width. The choice of the kernel defines the class of dependencies that can be represented, and thus any kernel function can be chosen if it matches the dependencies of the coefficient functions of the system under consideration on the scheduling variables. But, as it is assumed that no prior knowledge of these dependencies is available in the considered setup, a general purpose kernel like the Gaussian kernel is used which is capable of reproducing a wide range of smooth nonlinear functions.

In order to shrink the previously estimated coefficient functions $\hat{\vartheta}_j$ corresponding to insignificant lagged values of the input and the output, that is the elements of x, towards zero, in the second step, the following regularized convex optimization problem is solved

$$\underset{\{\mathrm{w}_j\}_{j=1}^{n_g}}{\operatorname{argmin}} \sum_{k=1}^{N} \left(y(k) - \sum_{j=1}^{n_g} \mathrm{w}_j^\top \zeta(p(k)) \hat{\vartheta}_j(p(k)) x_j(k) \right)^2 + \gamma \sum_{j=1}^{n_g} \| \mathrm{w}_j \|_\infty, \tag{21}$$

where $\zeta(p(k))$ is a vector of monomials in $p(k)$ which has to be specified a priori. $\mathrm{w}_j \in \mathbb{R}^{n_\mathrm{w}}$ is a vector of unknown parameters, and $\gamma \in \mathbb{R}_+$ is a regularization parameter. The term

$$\mathrm{w}_j^\top \zeta(p(k)) \hat{\vartheta}_j(p(k)) = \bar{\vartheta}_j(p(k)) \tag{22}$$

represents the scaled versions of the original coefficient functions introduced for the regularization. The regularization term $\gamma \sum_{i=j}^{n_g} \| \mathrm{w}_j \|_\infty$, i.e. the sum of the infinity norms ($l_{1,\infty}$), forces the vector w_i either to be equal to zero or full.

As the $l_{1,\infty}$-norm induces a bias in the estimated coefficient functions, in a third step, the non-zero coefficient functions are re-estimated with the approach proposed in Tòth et al (2011), that is minimizing (18), in order to obtain unbiased estimates.

3.2 Re-Estimation of a TS Model

As the parameter vectors ρ_j of the parametric LPV model (13), which we think of as TS model (14), are not accessible in the LSSVM framework, a re-estimation is necessary in order to obtain a parametric system description. Note, that although similar membership functions for TS models and kernel functions for the LSSVM approach are considered in this contribution, only an approximate reconstruction of the TS model is possible as the TS model uses far less parameters and normalized basis functions. Hence, the R-LSSVM approach is viewed as a pre-processing step for order selection of a TS model to avoid solving a non-convex optimization problem potentially multiple times. Subsequently, the extracted information of the dependency structure of the underlying process, that is the dynamical order, can be further exploited for TS modeling. For this purpose, 3 approaches are investigated in the following.

Approach 1 (Standard Methods)

Once the model order is found by applying the R-LSSVM approach, a TS model can be identified by standard methods. In this contribution, the following approaches are applied. In a first step, a fuzzy c-means clustering with multiple initializations is used to determine an appropriate partitioning of the scheduling space where the prototypes are used as centers of the membership functions (6). Afterwards, the local model parameters θ_{LM} are estimated using (7). In a third step, the obtained model is used as initialization for a nonlinear optimization of (8) where the prototypes and local model parameters are optimized simultaneously regarding the simulation performance of the model. In order to solve (8) the Matlab function lsqnonlin is used which by default uses a Trust-Region Reflective algorithm. In this contribution, the scheduling variable $z(k)$, the number of local models c, and σ_i are supposed to be known in order to keep the optimization problem simple. For the clustering, the fuzziness parameter $\nu \in \mathbb{R}^{>1}$ is chosen to be $\nu = 1.2$ following the recommendations in Kroll (2011).

Approach 2 (Pre-Filtering with LPV Model)

In order to further exploit the results from the R-LSSVM approach, the performance of a TS model which is identified by the aforementioned techniques is examined. But instead of optimizing (7) the following optimization problem is solved:

$$\underset{\theta_{\mathrm{LM}}}{\operatorname{argmin}} \parallel \hat{y}_{\mathrm{LPV}} - \Lambda\theta_{\mathrm{LM}} \parallel_2^2, \tag{23}$$

where $\hat{y}_{\mathrm{LPV}} = [\hat{y}_{\mathrm{LPV}}(1) \ldots \hat{y}_{\mathrm{LPV}}(N)]^\top \in \mathbb{R}^N$ is the vector of the simulated output of the LPV model identified by the R-LSSVM approach. Also, the system output $y(k)$ in (8) is replaced by $\hat{y}_{\mathrm{LPV}}$ for the subsequent nonlinear optimization. In this way, a pre-filtering of the estimation data with the non-parametric LPV model is performed yielding a pre-conditioned training-data set with reduced noise in order to obtain better estimates.

Approach 3 (Coefficient-function-based Cost Function)

In a third approach, a TS model is determined by minimizing the following nonlinear cost function:

$$\underset{\theta_{\mathrm{MF}},\theta_{\mathrm{LM}}}{\operatorname{argmin}} \left\|\hat{\vartheta}_r - \tilde{\theta}_r\left(\theta_{\mathrm{MF}},\theta_{\mathrm{LM}}\right)\right\|_2^2 . \tag{24}$$

That is, instead of minimizing the squared distance between prediction and measured output, the parametric TS model is determined such that the squared distance between the non-parametric estimation of the coefficient functions obtained by the LSSVM approach and the coefficient functions of a parametric TS model is minimized.

4 Simulation Example

In order to evaluate the performance of the R-LSSVM approach in the TS-fuzzy framework and appraise the proposed re-estimation procedures to obtain a TS model, a slightly modified version of the case study of Gringard and Kroll (2017) is considered. The test system is a TS-fuzzy system consisting of $c = 5$ superposed second-order lag elements with input-dependent attenuation and amplification. Gaussian membership functions like (6) are used for partitioning.

The prototypes are chosen to be $\{v_i\} = \{-2; -1; 0; 1; 2\}$ and the parameter specifying the width of each Gaussian is $\sigma_i = 0.3$. The i-th local model is defined by the following difference equation:

$$\begin{aligned} y_i^0(k) = &\underbrace{(2 - 2D_i\omega_0 T_s)}_{\theta_{i,1,\mathrm{LM}}} y^0(k-1) - \underbrace{(2D_i\omega_0 T_s - \omega_0^2 T_s^2 - 1)}_{\theta_{i,2,\mathrm{LM}}} y^0(k-2) \\ &+ \underbrace{K_i\omega_0^2 T_s^2}_{\theta_{i,3,\mathrm{LM}}} u(k-2), \end{aligned} \tag{25}$$

with the sample time $T_s = 10\,\mathrm{ms}$, $\omega_0 = 50\,\mathrm{rad/s}$, $\{K_i\} = \{6; 1.5; 3; 7.5; 4.5\}$, and $\{D_i\} \approx \{0.45; 0.71; 0.2; 0.58; 0.32\}$. The global system is given by the following NARX process:

$$y^0(k) = f(\varphi(k), z(k)) + e(k), \tag{26}$$

where $\varphi(k)^\top = [y^0(k-1), y^0(k-2), u(k-2)]$, $z(k) = u(k-2)$, and the Gaussian distributed additive zero-mean white noise $e(k)$. The resultant test system shows nonlinear behavior in the static as well as the dynamic part. Note, that the scheduling space is chosen to be one-dimensional for the sake of simplicity. But, the approaches are also applicable to higher dimensions of z.

Five different models are estimated from a training-data set of length $N = 1000$ and tested on a separate noise-free validation data set of length $N_\mathrm{v} = 1000$ in 50 Monte-Carlo runs with different realizations of the noise and the input. The input is chosen to be a uniformly distributed white noise process $u(k) \sim \mathcal{U}(-5, 5)$. Two Monte-Carlo studies are performed. In the first one, the average of Signal-to-Noise Ratio (SNR) over the 50 Monte-Carlo runs is equal to 18 dB and in the second one equal to 12 dB, corresponding to a standard deviation of the noise of 0.5 and 1, respectively. The SNR is defined as

$$\mathrm{SNR} = 10\ \mathrm{dB} \cdot \log_{10}\left(\frac{\sum_{k=1}^{N}(y_0(k))^2}{\sum_{k=1}^{N}(y(k) - y_0(k)^2}\right), \tag{27}$$

with $y_0(k)$ being the noise-free system output. The models are evaluated in a simulation which means that the output is only based on current inputs and the previous predictions of the output. To assess the generated models, the Best Fit Rate (BFR) is used:

$$\mathrm{BFR} = 100\,\% \cdot \max\left\{1 - \sqrt{\frac{\sum_{k=1}^{N_\mathrm{v}}(y(k) - \hat{y}(k))^2}{\sum_{k=1}^{N_\mathrm{v}}(y(k) - \bar{y}(k))^2}}, 0\right\}. \tag{28}$$

The five estimated models to be compared are:

M_{LPV} : LPV model estimated with the approach described in Section 3.1.

M_{TS1} : Over-parametrized TS model with $n_y = n_u = 5$ estimated with approach 1 described in Section 3.2.

M_{TS2} : TS model with the correct dynamical order as it should result from M_{LPV} estimated with approach 1.

M_{TS3} : TS model with the correct dynamical order estimated with approach 2.

M_{TS4} : TS model with the correct dynamical order estimated with approach 3.

4.1 Order Selection Results

For the identification of the LPV model, also an over-parametrized model with $n_a = n_b = 5$ is considered. σ_i of all RBF kernels (20) are kept equal. The values of the hyper-parameters are determined via a combination of trial and error and grid search optimizing the BFR on an independent calibration data set for the two noise levels and are fixed in the Monte-Carlo studies. For 18 dB, the obtained values are $\sigma = 1.0$, $\lambda = 1001$, and $\gamma = 1.1 \cdot 10^4$ yielding the correct dynamical order and dead time in 43 of the 50 Monte-Carlo runs. For 12 dB, only γ is adjusted to $\gamma = 2.0 \cdot 10^4$ and σ and λ are kept the same as for 18 dB yielding the correct dynamical order and dead time in 39 out of 50 runs.

4.2 LPV and TS Re-Estimation Results

The obtained LPV models M_{LPV} are compared to the parametric models M_{TS1} to M_{TS4}. It has to be mentioned that in all cases the scheduling variable is assumed to be known. Furthermore, all hyper-parameters for the TS modeling are pre-fixed as they were outside the scope of this investigation (that is $\nu = 1.2$, $c = 5$, and $\sigma = 0.3$).

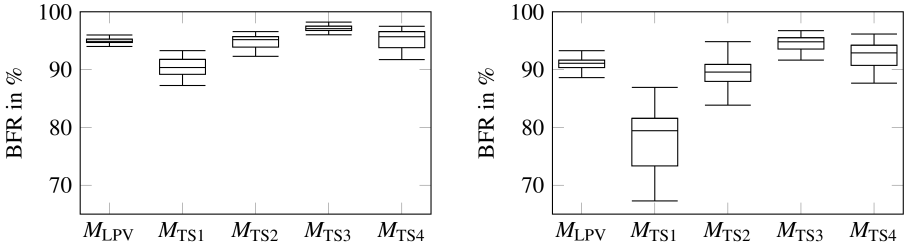

Figure 1: Box plots of the BFR on the validation data set of the five estimated models for an SNR of 18 dB (left) and 12 dB (right).

Figure 1 shows box plots of the BFR on the validation data sets of the five estimated models in the case the correct model structure was found by the R-LSSVM approach. Further results are listed in Table 1. It can be seen that the LPV model shows good results for both noise levels due to the inherent l_2 regularization preventing overfitting. The over-parametrized TS model M_{TS1} clearly suffers from overfitting, whereas the TS model with the correct dynamical order M_{TS2} shows comparable results to the LPV model. Although, a slight decrease of the median and a higher variation of the goodness of fit can be seen for lower SNR, probably, due to the sensitivity of the nonlinear optimization to local minima in this example. An improvement of the results is obtained by further exploiting the LSSVM results by approach 2 and 3. Especially, applying the pre-filtering approach yields the best results in this simulation example. The average fit can be improved by approach 3. But the variation of the results is higher as for the LSSVM model. In practice, a combination of the proposed re-estimation approaches may be promising.

Table 1: Performance comparison of selected models.

Model	**dim**(θ)	**average CPU sec**	**mean BFR @ 18 dB [%]**	**std BFR @ 18 dB [%]**	**mean BFR @ 12 dB [%]**	**std BFR @ 12 dB [%]**
M_{LPV}	N	3.6	94.9	0.6	90.9	1.1
M_{TS1}	55+5	181.3	90.4	1.7	77.4	7.0
M_{TS2}	15+5	35.8	94.7	1.5	89.1	2.8
M_{TS3}	15+5	19.5	96.8	1.3	94.4	1.5
M_{TS4}	15+5	0.4	95.0	2.2	92.4	2.3

Regarding the computational effort, it has to be mentioned that 1 run of the R-LSSVM approach requires 3.6 s. The nonlinear optimization of the TS model M_{TS2} with the correct dynamical order requires 35.8 s whereas it takes 181.3 s for the over-parametrized model M_{TS1}. Especially, it is obvious that the R-LSSVM approach may also reduce the computational burden for dynamical order selection by solving a surrogate problem compared to a wrapper approach for dynamical order selection, where a nonlinear optimization would be performed multiple times.

4.3 Discussion on Generalizability to Real-World Applications

In this example, the applicability of the R-LSSVM approach in the TS framework and the potential of re-estimation approaches was evaluated. The test system was chosen out of the set defined by the model class in order to evaluate the ability of the approach to identify the true system structure. In a real world application, the system will hardly be within the model class. But, the regularization approach for (local) order selection used in this contribution simply requires an estimate of the coefficient functions. Thus, the characteristics of the system under consideration should be at least smooth for both, the TS model and the LSSVM model, offering a wide range of application in real modeling tasks. The application to systems with multiple inputs, e.g., like the combustion engine considered in Kahl et al (2015), is straight forward, simply by augmenting the regression vector φ appropriately. However, in order to deal with the large number of training samples ($N \approx 46000$) in this example, a modification of the LSSVM, like fixed size LSSVM (De Brabanter et al, 2010), has to be used. Regarding the noise model, the ARX assumption might not be valid for the data-generating system. If e.g. the disturbance in a technical systems stems from measurement noise, the output error (OE) assumption would be more realistically. However, in this case, the optimization problems become nonlinear due to the recursion of the predicted output values leaving the convex framework.

5 Conclusions and Outlook

In this contribution, the dynamical order selection problem of a Takagi-Sugeno fuzzy model is solved by applying recently proposed LSSVM techniques for LPV models. It is shown that a TS model can be viewed as a special case of LPV models and that the order of the local models can be selected by using the R-LSSVM approach without prior specification of the antecedent part of the fuzzy model. This is done by regularizing the complete coefficient function describing the parameter dependencies on the scheduling variables instead of shrinking individual local parameters towards zero. In this way, the nonlinear parameter dependencies of the system can be estimated in a non-parametric but convex setting and solving of a nonlinear optimization problem to find a suitable partitioning of a TS model is avoided. Especially, for dynamical order selection, this is an important aspect. The reported simulation example has shown the capability of the R-LSSVM approach to find the correct dynamical order in most cases also for high noise levels. Additionally, the re-estimation step to obtain a parametrized TS model is found to be more accurate by pre-filtering the estimation data with the obtained LPV model.

The current investigations are made under the assumption of known scheduling variables and examined for a single input single output system. The extension to a system with multiple inputs is straightforward and will be investigated in a real world case study. In order to deal with an over-parametrized scheduling space the LSSVM framework has to be extended with a regularization step shrinking the derivatives of the coefficient functions with respect to the scheduling variables to zero which will be investigated in future work in the context of TS modeling. Furthermore, the combination of the proposed re-estimation approaches will be investigated.

Acknowledgements The authors thank the reviewers for their careful reading and their insightful comments and suggestions to improve this contribution.

References

Babuška R (1998) Fuzzy Modelling for Control, International Series in Intelligent Technologies, Vol. 12. Springer Science + Business Media B.V, Dordrecht (The Netherlands). DOI: 10.1007/978-94-011-4868-9.

Belz J, Nelles O, Schwingshackl D, Rehrl J, Horn M (2017) Order Determination and Input Selection with Local Model Networks. IFAC-PapersOnLine 50(1):7327–7332, Elsevier B.V. DOI: 10.1016/j.ifacol.2017.08.1475.

De Brabanter K, De Brabanter J, Suykens JA, De Moor B (2010) Optimized Fixed-size Kernel Models for Large Data Sets. Computational Statistics & Data Analysis 54(6):1484–1504, Elsevier B.V. DOI: 10.1016/j.csda.2010.01.024.

Gringard M, Kroll A (2017) On Optimal Offline Experiment Design for the Identification of Dynamic TS Models: Multi-step Signals for Uncertainty-minimal Consequent Parameters. In: Proceedings of the 27th Workshop Computational Intelligence, Hoffmann F, Hüllermeier E, Mikut R (eds), KIT Scientific Publishing, Karlsruhe (Germany), pp. 117–138. DOI: 10.5445/KSP/1000074341.

Hong X, Harris CJ (2001) Variable Selection Algorithm for the Construction of MIMO Operating Point Dependent Neurofuzzy Networks. IEEE Transactions on Fuzzy Systems 9(1):88–101, Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/91.917117.

Kahl M, Kroll A, Kästner R, Sofsky M (2015) Application of Model Selection Methods for the Identification of Dynamic Boost Pressure Model. IFAC-PapersOnLine 48(25):829–834, Elsevier B.V. DOI: 10.1016/j.ifacol.2015.12.232.

Kroll A (1996) Identification of Functional Fuzzy Models Using Multidimensional Reference Fuzzy Sets. Fuzzy Sets and Systems 80(2):149–158.

Kroll A (2011) On Choosing the Fuzziness Parameter for Identifying TS Models with Multidimensional Membership Functions. Journal of Artificial Intelligence and Soft Computing Research 1(4):283–300.

Lughofer E, Kindermann S (2010) SparseFIS: Data-driven Learning of Fuzzy Systems with Sparsity Constraints. IEEE Transactions on Fuzzy Systems 18(2):396–411, Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/TFUZZ.2010.2042960.

Luo M, Sun F, Liu H, Li Z (2014) A Novel T-S Fuzzy Systems Identification with Block Structured Sparse Representation. Journal of the Franklin Institute 351(7):3508–3523, Elsevier B.V. DOI: 10.1016/j.jfranklin.2013.05.008.

Mejari M, Piga D, Bemporad A (2016) Regularized Least Square Support Vector Machines for Order and Structure Selection of LPV-ARX Models. In: Proceedings of the 15th European Control Conference, Institute of Electrical and Electronics Engineers (IEEE), New York (USA), pp. 1649–1654. DOI: 10.1109/ECC.2016.7810527.

Nelles O (2001) Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models, 1st edn. Springer, Berlin, Heidelberg (Germany). DOI: 10.1007/978-3-662-04323-3.

Piga D, Tòth R (2013) LPV Model Order Selection in an LS-SVM Setting. In: Proceedings of the 52nd IEEE Conference on Decision and Control, Institute of Electrical and Electronics Engineers (IEEE), New York (USA), pp. 4128–4133. DOI: 10.1109/CDC.2013.6760522.

Takagi T, Sugeno M (1985) Fuzzy Identification of Systems and its Application to Modelling and Control. IEEE Transactions on Systems, Man, and Cybernetics, pp. 116–132, Institute of Electrical and Electronics Engineers (IEEE), New York (USA). DOI: 10.1109/TSMC.1985.6313399.

Tòth R, Laurain V, Zheng WX, Poolla K (2011) Model Structure Learning: A Support Vector Machine Approach for LPV Linear-regression Models. In: Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference, Institute of Electrical and Electronics Engineers (IEEE), New York (USA), pp. 3129–3197. DOI: 10.1109/CDC.2011.6160564.

Wang HO, Tanaka K, Griffin M (1995) Parallel Distributed Compensation of Nonlinear Systems by Takagi-Sugeno Fuzzy Model. In: Proceedings of the IEEE International Conference on Fuzzy Systems, Institute of Electrical and Electronics Engineers (IEEE), New York (USA), pp. 531–538. DOI: 10.1109/FUZZY.1995.409737.