Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Extending Regularized Least Squares Support Vector Machines for Order Selection of Dynamical Takagi-Sugeno Models

Matthias Kahl, Andreas Kroll

Department of Measurement and Control, Institute for System Analytics and Control, University of Kassel, D-34109 Kassel, Germany (e-mail: {matthias.kahl,andreas.kroll}@mrt.uni-kassel.de).

Abstract: In this paper, the problem of order selection for nonlinear dynamical Takagi-Sugeno (TS) fuzzy models is adressed. It is solved by reformulating the TS model in its Linear Parameter Varying (LPV) form and applying an extension of a recently proposed Regularized Least Squares Support Vector Machine (R-LSSVM) technique for LPV models. For that, a nonparametric formulation of the TS identification problem is proposed which uses data-dependent basis functions. By doing so, the partition of unity of the TS model is preserved and the scheduling dependencies of the model are obtained in a nonparametric manner. For the local order selection, a regularization approach is used which forces the coefficient functions of insignificant values of the lagged input and output towards zero.

Keywords: Nonlinear system identification, Model structure selection, Takagi-Sugeno models, NARX, Nonparametric methods, LPV system identification

1. INTRODUCTION

In recent work on linear parameter varying (LPV) system identification, nonparametric identification approaches based on least squares support vector machines (LSSVM) are investigated (Tòth et al., 2011; Bachnas et al., 2014; Mejari et al., 2016). The LSSVM approach exploits the reproducing kernel theory to provide a nonparametric approximation of functions based on a series expansion with kernel functions evaluated on the observed data points. Inspired by this approach, in this contribution the identification of Takagi-Sugeno (TS) models (Takagi and Sugeno, 1985) is investigated using a data-dependent formulation in order to provide a convex framework for local order selection of these models.

In order to find a suitable local model structure, a 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 of TS models. However, if every possible subset has to be evaluated by exhaustive search, the resulting high-dimensional combinatorial optimization problem quickly becomes 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 obtaining sparse local models, the original combinatorial optimization problem can be approximated by lasso-like convex relaxation (Tibshirani, 1996). For TS models 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 an l_1 penalty into a nonlinear optimization problem. Additionally, they applied a sparse estimator for parameter estimation of the local models. 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. Additionally, as the output of a TS model is nonlinear in the parameters of the basis functions, a nonlinear optimization problem has to be solved in order to partition the scheduling space, thus one may stack in local minima.

In this contribution, the partitioning problem is relaxed by using data-dependent basis functions preserving convexity. In order to determine the consequent parameters of the model the utilization of either explicitly regularized global estimation or implicitly regularized local regression are investigated. Furthermore, the local order selection is solved by exploiting the LPV formulation of TS models and applying the regularization approach proposed by Mejari et al. (2016).

2. PROBLEM FORMULATION

The aim of the empirical multivariate modeling problem consists in finding the unknown functional relationship between n input variables $\boldsymbol{x} \in \mathbb{R}^n$ and the output variable $y \in \mathbb{R}$:

$$y(k) = f(\boldsymbol{x}(k)) + v(k), k = 1, \dots, N \quad (1)$$

based on a data set $D^N = \{\boldsymbol{x}(k), y(k)\}_{k=1}^N$ with N elements, containing measurements of input variables and the output of the process under consideration. In order to model dynamical systems, $\boldsymbol{x}(k)$ contains delayed values of the measured system inputs $\boldsymbol{u}(k) \in \mathbb{R}^{n_u}$ and the measured output $y(k) \in \mathbb{R}$. In this contribution, single-input single-output (SISO) systems are considered. But, the approaches can directly be applied to multi-input systems. The term $v(k)$ is assumed to be an independent and identically distributed random variable with zero mean and finite variance σ^2. The function $f : \mathbb{R}^n \to \mathbb{R}$ is the nonlinear function representing the system behavior to be modeled. In this contribution, the TS model class is considered assuming f to be smooth.

2.1 Takagi-Sugeno Model Class

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

$$\hat{y}(k) = \sum_{i=1}^{c} \phi_i(\boldsymbol{z}(k)) \cdot \hat{y}_i(k). \quad (2)$$

As local model type ARX models of the form

$$\hat{y}_i(k) = \sum_{r=1}^{n} \theta_{i,r,\mathrm{LM}} \cdot \varphi_r(k) \quad (3)$$

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

$$\begin{aligned}\boldsymbol{\varphi}(k) =& [1\; y(k-1) \dots y(k-n_y),\\ & u(k-T_\tau) \dots u(k-n_u-T_\tau)]^\top,\end{aligned} \quad (4)$$

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

$$\boldsymbol{\theta}_{i,\mathrm{LM}} = [\theta_{i,0}\; \boldsymbol{\theta}_{i,y}^\top\; \boldsymbol{\theta}_{i,u}^\top]^\top \in \mathbb{R}^n, \quad (5)$$

and T_τ is a discrete dead time. $\theta_{i,0} \in \mathbb{R}$ is the offset of the local model, $\boldsymbol{\theta}_{i,y} \in \mathbb{R}^{n_y}$ is the parameter vector corresponding to $y(k)$, and $\boldsymbol{\theta}_{i,u} \in \mathbb{R}^{n_u}$ corresponds to $u(k)$.

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

$$\phi_i(\boldsymbol{z}(k)) = \frac{\mu_i(\boldsymbol{z}(k))}{\sum_{j=1}^{c} \mu_j(\boldsymbol{z}(k))}, \quad (6)$$

with the membership functions $\mu_i(\boldsymbol{z}(k))$. In this contribution, Gaussian membership functions are used

$$\mu_i(\boldsymbol{z}(k)) = \exp\left(-\frac{1}{2} \frac{\| \boldsymbol{z}(k) - \boldsymbol{v}_i \|_2^2}{\sigma_i^2}\right), \quad (7)$$

where $\boldsymbol{v}_i \in \mathbb{R}^{n_z}$ represents the partition's prototype and $\sigma_i \in \mathbb{R}_+$ specifies the width of the Gaussian function aggregated in the parameter vector $\boldsymbol{\theta}_{i,\mathrm{MF}} = [\boldsymbol{v}_i^\top\; \sigma_i]^\top$, so that $\mu_i(\boldsymbol{z}(k)) = \mu_i(\boldsymbol{\theta}_{i,\mathrm{MF}}, \boldsymbol{z}(k))$. For simplicity an Euclidean norm is considered.

As the TS model interpolates between the local ARX models (which are assumed to have identical structure) it can be written in LPV-ARX form. Inserting (3) to (2) yields:

$$\hat{y}(k) = \sum_{i=1}^{c} \sum_{r=1}^{n} \phi_i(\boldsymbol{z}(k)) \cdot \theta_{i,r,\mathrm{LM}} \cdot \varphi_r(k), \quad (8)$$

and by introducing the coefficient functions

$$\tilde{\theta}_r(\boldsymbol{z}(k)) = \sum_{i=1}^{c} \phi_i(\boldsymbol{z}(k)) \cdot \theta_{i,r,\mathrm{LM}}, \quad (9)$$

the TS-fuzzy model is a special case of an LPV-ARX model:

$$\hat{y}(k) = \sum_{r=1}^{n} \tilde{\theta}_r(\boldsymbol{z}(k)) \cdot \varphi_r(k), \quad (10)$$

where the same fuzzy basis function for each regressor is used to parametrize the coefficient functions as common in the TS framework.

2.2 Structure Selection of TS models

The structure selection problem of data-driven TS models consists of three parts: i) the choice of appropriate scheduling variables $\boldsymbol{z}(k)$ and input variables $\boldsymbol{u}(k)$, ii) the partitioning of the scheduling space by an appropriate parametrization of the fuzzy basis functions $\phi_i(\boldsymbol{z}(k))$ of a predefined type as well as the choice of the number of local models c, and iii) the selection of a suitable local model structure characterized by n_y, n_u, and T_τ. 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 ii). Alternatively, heuristic construction strategies were proposed like grid partitioning, data-point-based methods, clustering-based approaches or heuristic tree construction algorithms like LOLIMOT each with individual advantages and drawbacks (see, e.g., Nelles, 2001). In order to solve iii), the approaches mentioned in the introduction can be applied. In this contribution, i) is assumed to be given by the modeling exercise. In order to solve ii) and iii) the approach proposed in section 3 is used by extending the approach proposed in Mejari et al. (2016) to TS models.

3. IDENTIFICATION APPROACH

3.1 Regularized LSSVM for Order Selection of LPV Models

The approach proposed in Mejari et al. (2016) is based on the method developed in Tòth et al. (2011) where an componentwise LSSVM is used to describe the nonlinear dependence of the coefficient functions of an LPV model. The approach proposed in Mejari et al. (2016) incorporates an additional regularization step in order to select the dynamical order of the model. The approach consists of three 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. The LSSVM formulation for the estimation of an LPV model starts from:

$$\begin{aligned}\underset{\boldsymbol{\rho},\boldsymbol{e}}{\operatorname{argmin}}\; \mathcal{I}(\boldsymbol{\rho}, \boldsymbol{e}) &= \frac{1}{2} \sum_{r=1}^{n} \boldsymbol{\rho}_r^\top \boldsymbol{\rho}_r + \frac{\lambda}{2} \sum_{k=1}^{N} e^2(k)\\ \text{s.t. } e(k) &= y(k) - \sum_{r=1}^{n} \boldsymbol{\rho}_r^\top \boldsymbol{\phi}_r(\boldsymbol{z}) \varphi_r(k),\end{aligned} \quad (11)$$

with $\boldsymbol{e} = [e(1)\cdots e(N)]^\top \in \mathbb{R}^N$, the unknown parameter vector $\boldsymbol{\rho}_r \in \mathbb{R}^{n_H}$, the feature maps $\boldsymbol{\phi}_r : \mathbb{R} \to \mathbb{R}^{n_H}$, and the regularization parameter $\lambda \in \mathbb{R}_+$. The Lagrangian dual problem associated with (11) is:

$$\mathcal{L}(\boldsymbol{\rho}, \boldsymbol{e}, \boldsymbol{\alpha}) = \mathcal{I}(\boldsymbol{\rho}, \boldsymbol{e}) - \sum_{k=1}^{N} \alpha_k \left(e(k) - y(k) + \sum_{r=1}^{n} \boldsymbol{\rho}_r^\top \boldsymbol{\phi}_r(\boldsymbol{z}(k))\varphi_r(k) \right), \tag{12}$$

with $\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_N]^\top \in \mathbb{R}^N$ being the Lagrangian multipliers. Setting derivatives with respect to $\boldsymbol{\rho}$, $\boldsymbol{e}$, and $\boldsymbol{\alpha}$ to zero and eliminating $\boldsymbol{\rho}$ and $\boldsymbol{e}$ gives the following set of equations

$$y(k) = \sum_{r=1}^{n} \left(\sum_{k=1}^{N} \alpha_k \varphi_r(k) \boldsymbol{\phi}_r^\top(\boldsymbol{z}(k)) \right) \boldsymbol{\phi}_r(\boldsymbol{z}(k))\varphi_r(k) + \frac{1}{\lambda}\alpha_k. \tag{13}$$

Instead of specifying $\boldsymbol{\phi}(\boldsymbol{z}(k))$, in the LSSVM setting the inner product $\boldsymbol{\phi}_r^\top(\boldsymbol{z}(k))\boldsymbol{\phi}_r(\boldsymbol{z}(k))$ is defined by an a priori chosen positive definite kernel function $K_r = \boldsymbol{\phi}_r^\top(\boldsymbol{z}(k))\boldsymbol{\phi}_r(\boldsymbol{z}(k))$ which enables the description of the nonlinear scheduling dependencies of the model coefficient functions. 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):

$$K_r(\boldsymbol{z}(k), \boldsymbol{z}(m)) = \exp\left(-\frac{1}{2} \frac{\| \boldsymbol{z}(k) - \boldsymbol{z}(m) \|_2^2}{\sigma_r^2} \right), \tag{14}$$

with the hyper-parameter σ_r specifying its width. Based on K_r, the $N \times N$ kernel matrix $\boldsymbol{\Omega}$ is constructed whose (k,m)-th element is given by

$$\begin{aligned} [\boldsymbol{\Omega}]_{k,m} &= \sum_{r=1}^{n} [\boldsymbol{\Omega}_r]_{k,m}, \\ [\boldsymbol{\Omega}_r]_{k,m} &= \varphi_r(k)\left(K_r(\boldsymbol{z}(k), \boldsymbol{z}(m))\right)\varphi_r(m) \end{aligned} \tag{15}$$

The solution of the LSSVM estimation problem in terms of the Lagrangian multipliers is given by

$$\boldsymbol{\alpha} = \left(\boldsymbol{\Omega} + \lambda^{-1}\boldsymbol{I}_N\right)^{-1}\boldsymbol{Y}, \tag{16}$$

with $\boldsymbol{Y} = [y(1)\dots y(N)]^\top \in \mathbb{R}^N$, and the coefficient functions to be estimated are obtained as

$$\hat{\vartheta}_r(\cdot) = \boldsymbol{\rho}_r^\top \boldsymbol{\phi}_r(\cdot) = \sum_{k=1}^{N} \alpha_k K_r(\boldsymbol{z}(k), \cdot)\varphi_r(k). \tag{17}$$

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

$$\begin{aligned} \underset{\{\boldsymbol{w}_r\}_{r=1}^n}{\operatorname{argmin}} \sum_{k=1}^{N} &\left(y(k) - \sum_{r=1}^{n} \boldsymbol{w}_r^\top \boldsymbol{\zeta}(\boldsymbol{z}(k))\hat{\vartheta}_r(\boldsymbol{z}(k))\varphi_r(k) \right)^2 \\ &+ \gamma \sum_{r=1}^{n} \| \boldsymbol{w}_r \|_\infty, \end{aligned} \tag{18}$$

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

$$\boldsymbol{w}_r^\top \boldsymbol{\zeta}(\boldsymbol{z}(k))\hat{\vartheta}_r(\boldsymbol{z}(k)) = \bar{\vartheta}_r(\boldsymbol{z}(k)) \tag{19}$$

represents the scaled versions of the original coefficient functions introduced for the regularization. The regularization term $\gamma\sum_{r=1}^{n} \| \boldsymbol{w}_r \|_\infty$, i.e. the sum of the infinity norms $l_{1,\infty}$, forces all elements of the vector $\boldsymbol{w}_r$ either to be equal to zero or non-zero.

As the $l_{1,\infty}$-norm induces a bias in the estimated coefficient functions, in a final step, the non-zero coefficient functions are re-estimated with the approach proposed in Tòth et al. (2011), that is computing (16), in order to obtain unbiased estimates.

3.2 Order Selection of Takagi-Sugeno Models

The approach described in section 3.1 can been used to determine the (local) dynamical order of a TS model in its LPV form (10) as it has already be investigated in Kahl and Kroll (2018). However, due to the application of the kernel trick in the LSSVM framework, the parameter vectors $\boldsymbol{\rho}_r$ of the parametric LPV model in (11), which we think of as TS model (8), are neither accessible, nor can a partition of unity be preserved in this setting. To circumvent this drawbacks in the TS framework, we investigate a nonparametric TS formulation (in the primal weight space) for local order selection consisting of the following three steps:

- S1 Partition the scheduling space by means of data-based FBFs
- S2 Estimate local model parameters of an over-parametrized TS model by either use of explicit global or implicit local estimation
- S3 Penalize the coefficient functions of the resultant model to force sparseness in the number of lagged inputs/outputs

Nonparametric Takagi-Sugeno Partitioning (S1) In the first step, each membership function μ_i is centered on each training data point $\boldsymbol{z}$ of the scheduling variable, such that:

$$\mu_i(\boldsymbol{z}(k)) = \exp\left(-\frac{1}{2} \frac{\| \boldsymbol{z}(k) - \boldsymbol{z} \|_2^2}{\sigma_i^2} \right) \tag{20}$$

and

$$\phi_i(\boldsymbol{z}(k)) = \frac{\mu_i(\boldsymbol{z}(k))}{\sum_{j=1}^{N} \mu_j(\boldsymbol{z}(k))}. \tag{21}$$

By doing so, the scheduling space partitioning is given by means of the training data at hand, such that the data points are assigned to the local models by the corresponding basis function ϕ_i. σ_i is treated as hyperparameter and can be determined by cross validation. This approach is strongly related to piecewise affine (PWA) system identification (see, e.g., Garulli et al. (2012) for a survey) and has been generalized to local linear model networks, which can be interpreted as TS models, in Münker and Nelles (2017). The approach proposed in Münker and Nelles (2017) first solves a local regression problem to obtain an estimate of the local model parameters. Afterwards, the parameters are clustered and assigned to the corresponding local models to obtain a compact rule base. In this contribution, the clustering step applied in (Münker and Nelles, 2017) is omitted in order to avoid the solution of a non-convex optimization problem in steps S1 to S3. However, the clustering can be applied once the correct model order is found, which is out of the scope of this paper.

Estimation of Local Model Parameters (S2) Clearly, the obtained model lacks sparseness and at least $n \cdot N$ local model parameters have to be estimated and thus regularization is required to prevent overfitting. In fact, (11) is the constrained version of the ridge regression problem in the feature space. In the TS model identification setting, for given $\boldsymbol{\theta}_{i,\mathrm{MF}}$, $i = 1, \ldots, c$ the ridge regression problem is

$$\underset{\boldsymbol{\theta}_{\mathrm{LM}}}{\operatorname{argmin}} \frac{1}{2} \parallel \boldsymbol{Y} - \boldsymbol{\Lambda}\boldsymbol{\theta}_{\mathrm{LM}} \parallel_2^2 + \frac{\lambda}{2} \parallel \boldsymbol{\theta}_{\mathrm{LM}} \parallel_2^2, \tag{22}$$

where $\lambda \in \mathbb{R}_+$, $\boldsymbol{\theta}_{\mathrm{LM}}^\top = \left[\boldsymbol{\theta}_{\mathrm{LM},1}^\top \ldots \boldsymbol{\theta}_{\mathrm{LM},c}^\top\right] \in \mathbb{R}^{n \cdot c}$, $\boldsymbol{\varphi} = [\boldsymbol{\varphi}(1) \ldots \boldsymbol{\varphi}(N)]^\top \in \mathbb{R}^{N \times n}$, and $\boldsymbol{\Lambda} = [\boldsymbol{\Gamma}_1\boldsymbol{\varphi} \ldots \boldsymbol{\Gamma}_c\boldsymbol{\varphi}]^\top \in \mathbb{R}^{N \times c \cdot n}$ is the extended regression matrix with $\boldsymbol{\Gamma}_i = \mathrm{diag}(\phi_i(\boldsymbol{z}(1)) \ldots \phi_i(\boldsymbol{z}(N))) \in \mathbb{R}^{N \times N}$. The solution is given by

$$\hat{\boldsymbol{\theta}}_{\mathrm{LM}} = \left(\boldsymbol{\Lambda}^\top\boldsymbol{\Lambda} + \lambda \boldsymbol{I}_{nc}\right)^{-1} \boldsymbol{\Lambda}^\top \boldsymbol{Y}, \tag{23}$$

where $\boldsymbol{I}_{nc}$ is the $n \cdot c \times n \cdot c$ identity matrix. As for $c = N$ local models $n \cdot N$ local model parameters have to be estimated the solution can also be derived from its dual problem. Converting (22) into a constrained minimization problem

$$\begin{aligned} \underset{\boldsymbol{\theta}_{\mathrm{LM}},\boldsymbol{e}}{\operatorname{argmin}}\ \mathcal{I}(\boldsymbol{\theta}_{\mathrm{LM}}, \boldsymbol{e}) = \frac{\lambda}{2} \parallel \boldsymbol{e} \parallel_2^2 + \frac{1}{2} \parallel \boldsymbol{\theta}_{\mathrm{LM}} \parallel_2^2 \\ \text{s.t. } \boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{\Lambda}\boldsymbol{\theta}_{\mathrm{LM}}, \end{aligned} \tag{24}$$

the corresponding Lagrangian is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}_{\mathrm{LM}}, \boldsymbol{e}, \boldsymbol{\alpha}) = &\frac{\lambda}{2} \parallel \boldsymbol{e} \parallel_2^2 + \frac{1}{2} \parallel \boldsymbol{\theta}_{\mathrm{LM}} \parallel_2^2 \\ &+ \boldsymbol{\alpha} \left(\boldsymbol{e} - \boldsymbol{Y} + \boldsymbol{\Lambda}\boldsymbol{\theta}_{\mathrm{LM}} \right). \end{aligned} \tag{25}$$

Proceeding as for the LSSVM, i.e. setting derivatives with respect to the primal variables to zero and eliminating the primal variables by substitution, the dual problem is obtained as

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\ \frac{1}{2}\boldsymbol{\alpha} \left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \lambda \boldsymbol{I}_N\right) \boldsymbol{\alpha}^\top - \boldsymbol{\alpha}\boldsymbol{Y} \tag{26}$$

with $\boldsymbol{I}_N$ is the $N \times N$ identity matrix. The solution is given by

$$\boldsymbol{\alpha} = \left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \lambda \boldsymbol{I}_N\right)^{-1} \boldsymbol{Y}. \tag{27}$$

By using the dual, only N parameters $\boldsymbol{\alpha}$ have to estimated. With the centers of the FBFs given by the training data, the local model parameters in the primal space are given by

$$\boldsymbol{\theta}_{\mathrm{LM}} = \boldsymbol{\alpha}\boldsymbol{\Lambda}. \tag{28}$$

Alternatively, the local estimation approach can be used to obtain the consequent parameters as stated in Münker and Nelles (2017). In local estimation, the parameters of each local model are estimated independently of the remaining ones by using the basis functions to weight each least squares problem

$$\underset{\boldsymbol{\theta}_{i,\mathrm{LM}}}{\operatorname{argmin}} \parallel \boldsymbol{\Gamma}_i^{1/2} \left(\boldsymbol{Y} - \boldsymbol{\varphi}\boldsymbol{\theta}_{i,\mathrm{LM}}\right) \parallel_2^2. \tag{29}$$

The solution is given by

$$\boldsymbol{\theta}_{i,\mathrm{LM}} = \left(\boldsymbol{\varphi}^\top \boldsymbol{\Gamma}_i \boldsymbol{\varphi}\right)^{-1} \boldsymbol{\varphi}^\top \boldsymbol{\Gamma}_i \boldsymbol{Y}, \quad i = 1, \ldots, c. \tag{30}$$

As for local estimation the aggregation of the local models is not taken into account, it does not provide an optimal TS model in terms of minimal modeling error. However, the number of effective parameters is reduced and thus some implicit regularization is provided (Nelles, 2001).

Local Order Selection (S3) With the partitioning given by the training data points of $\boldsymbol{z}(k)$ (thus, $\boldsymbol{z}(k)$ is assumed to be known) and the estimates of the local model parameters determined either by the explicitly regularized global approach (23) (alternatively, using (27) with (28)) or the implicitly regularized local estimation approach (30), the coefficient functions of the TS model are given by (9). With $\tilde{\theta}_r(\boldsymbol{z}(k))$ given, the regularization approach (18), with $\hat{\vartheta}_r(\boldsymbol{z}(k)) = \tilde{\theta}_r(\boldsymbol{z}(k))$, can be applied in order to determine the local order of the TS model.

4. NUMERICAL EXAMPLE

In order to evaluate the performance of the proposed approach, a slightly modified version of the case study in (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 (7) are used for partitioning. The prototypes are chosen to be $\{v_i\} = \{-3; -1; 0; 1; 3\}$ and the parameters specifying the width of each Gaussian is $\sigma_i = 0.3 \forall i$. The i-th local model is defined by the following difference equation obtained from discretization of the respective continuous second-order lag elements:

$$\begin{aligned} y_i(k) = &\underbrace{(2 - 2D_i\omega_0 T_s)}_{\theta_{i,1,\mathrm{LM}}} y(k-1) \\ &- \underbrace{(2D_i\omega_0 T_s - \omega_0^2 T_s^2 - 1)}_{\theta_{i,2,\mathrm{LM}}} y(k-2) \\ &+ \underbrace{K_i \omega_0^2 T_s^2}_{\theta_{i,3,\mathrm{LM}}} u(k-2), \end{aligned} \tag{31}$$

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

$$y(k) = f(\boldsymbol{\varphi}(k), z(k)) + e(k), \tag{32}$$

where $\boldsymbol{\varphi}(k) = [y(k-1), y(k-2), u(k-2)]$, $z(k) = u(k-2)$, and the Gaussian distributed additive zero-mean white noise $e(k)$. 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.

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 $u(k) \sim \mathcal{U}(-5, 5)$. The average Signal-to-Noise Ratio (SNR) over the 50 Monte-Carlo runs is 18 dB, corresponding to a standard deviation of the noise of 0.5. The SNR is defined as

$$\mathrm{SNR} = 10\ \mathrm{dB} \cdot \log_{10} \left(\frac{\sum_{k=1}^{N} (y_0(k))^2}{\sum_{k=1}^{N} (y(k) - y_0(k)^2} \right), \tag{33}$$

with $y_0(k)$ being the noise-free system output. The models are evaluated in a simulation of the validation data set 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:

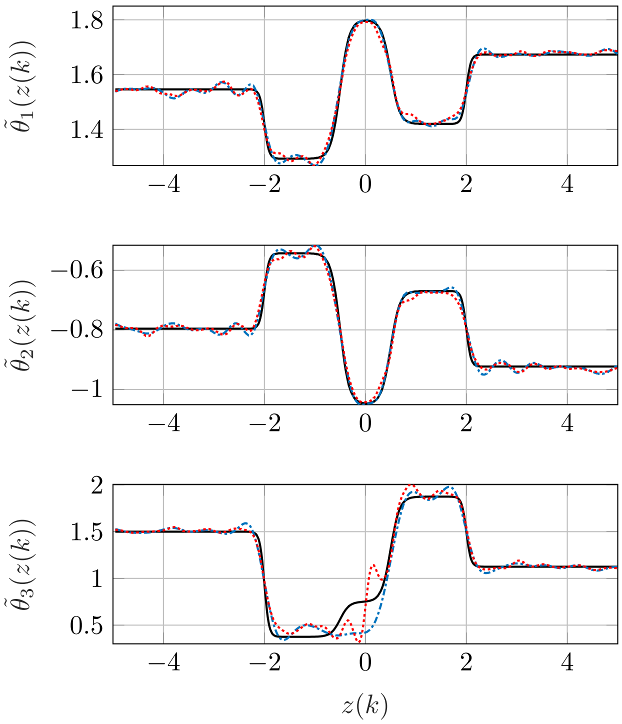

Fig. 1. Coefficient functions of the true system (black), and of estimated TS model for global (blue), and for local parameter identification (red)

$$\text{BFR} = 100\% \cdot \max\left\{1 - \sqrt{\frac{\sum_{k=1}^{N_\text{v}}(y(k)-\hat{y}(k))^2}{\sum_{k=1}^{N_\text{v}}(y(k)-\bar{y}(k))^2}}, 0\right\}. \tag{34}$$

4.1 Order Selection Results

For the identification an overparametrized TS model with $n_a = n_b = 5$ is considered. σ_i of all fuzzy membership functions (7) 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 of length $N_\text{c} = 1000$ and are fixed in the Monte-Carlo studies. For global estimation, the obtained values are $\sigma_i = 0.25$, $\lambda = 2.125 \cdot 10^{-2}$, and $\gamma = 1.1 \cdot 10^4$. For local estimation, $\sigma_i = 0.1$ is obtained and γ is chosen identical to the global approach yielding the correct dynamical order and dead time in 48 of the 50 Monte-Carlo runs for both, global and local consequent parameter estimation.

4.2 Global versus Local Consequent Parameter Estimation

In this section, we further investigate the results regarding the obtained coefficient functions and estimates of the local model parameters. Fig. 1 shows the obtained coefficient functions for the determined hyper-parameters and the correct dynamical order for one noise realization. As can be seen, for both approaches the coefficient functions are captured well. But, for both approaches there is a slight deviation in the fit of the coefficient function $\tilde{\theta}_3(z(k))$ in the range $[-0.7, 0.4]$. For the global estimation, there is a smooth but biased approximation in this range.

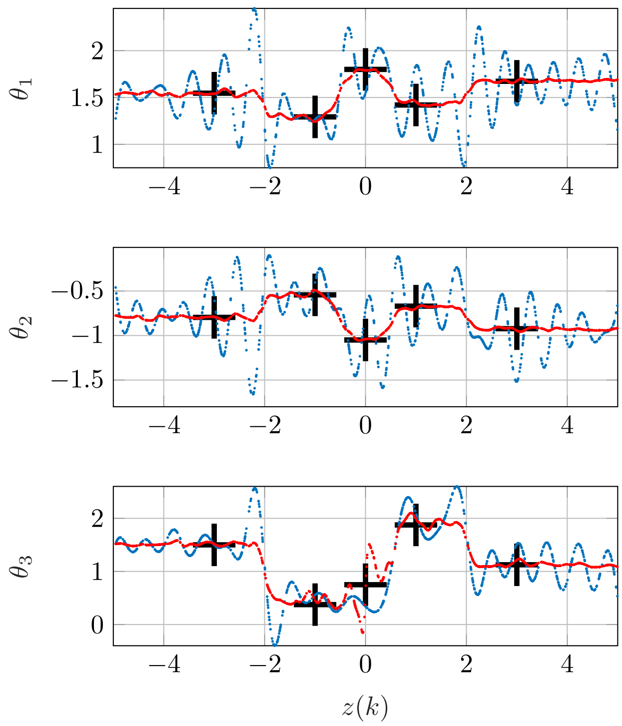

Fig. 2. Local model parameters of the true system (black crosses), and of estimated TS model for global (blue dots), and for local parameter identification (red dots)

For the local estimation, the estimate oscillates around the coefficient function of the underlying system. The corresponding values of the BFR on the validation data set are 92.36 % for the global identification and 93.39 % for local identification approach, both indicating an excellent fit. The estimated local model parameters corresponding to each local model in $z(k)$ are depicted in Fig. 2 and compared to the local parameters of the TS system. As can be seen, the estimates for the local estimation are comparable to the local parameters of the true system, whereas the results for the global estimation show highly oscillatory behavior.

4.3 Discussion

For the determination of the local dynamical order using the approach described in section 3.2 just the coefficient functions $\tilde{\theta}_r(z(k))$ have to be estimated. As both estimation approaches provide good approximations of the coefficient functions, the results for the order selection are similar in this example. The results for the estimates of the local model parameters vastly differ for both estimation approaches. Due to the nonparametric partitioning of the scheduling space, the estimates using the global identification approach show oscillatory behavior although a regularization is used to prevent ill-conditioning. According to Murray-Smith and Johansen (1995), this is due to compensation effects where the negative contribution of one local model is compensated by the positive contribution of the neighboring ones. This significantly harms the interpretability of the model and may be disadvantageous when used for control purposes. Moreover, for a rule-base simplification based on parameter-space clustering, like proposed in Münker and Nelles (2017), the global

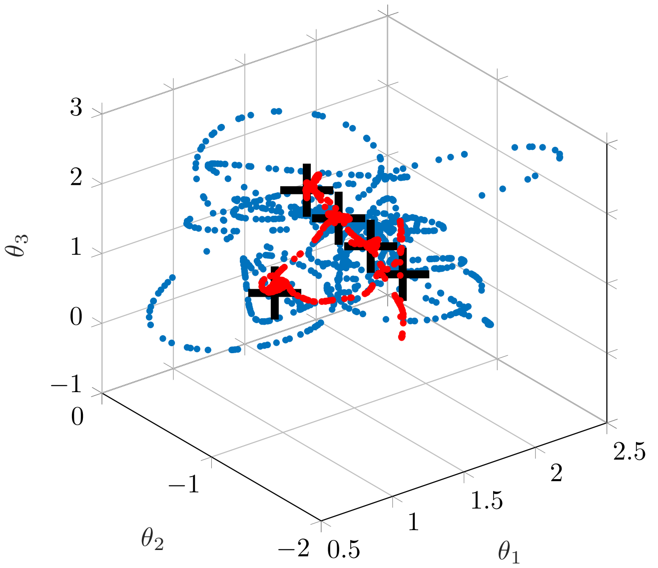

Fig. 3. Local model parameters of the true system (black crosses), of the estimated TS model for global (blue dots), and for local parameter identification (red dots) in the parameter space

estimates are not useful (see Fig. 3). On the contrary, the local estimation is robust against the overparametrization and still provides results which are comparable to local linearizations of the system with a higher approximation accuracy of the coefficient functions. Note, in order to obtain smoother results for the local model parameters for local estimation, the bandwidth σ_i can be adjusted. This however would decrease the BFR.

5. CONCLUSIONS AND OUTLOOK

In this contribution, a convex framework for local dynamical order selection of Takagi-Sugeno fuzzy models was proposed. Data-dependent membership functions were used such that the partitioning of the scheduling space was given by the training data at hand. In this way, the nonlinear parameter dependencies of the system can be estimated in a non-parametric but convex setting. Solving 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. In order to obtain estimates of the local model parameters, explicitly regularized global estimation and implicitly regularized local estimation were investigated. Subsequently, the regularization approach proposed in Mejari et al. (2016) was applied, where the LPV form of the TS model is exploited. A simulation example demonstrated the capability of the proposed approach to find the correct dynamical order in 96 % of all cases. However, the values of the estimated local model parameters showed major differences for the used estimation approaches. The local parameter estimation approach was robust against the overparametrization and should be preferred in this setting. The current investigations were carried out under the assumption of known scheduling variables and examined a single input single output system. The extension to systems with multiple inputs is straightforward and will be investigated in a real world case study.

REFERENCES

Bachnas, A., Tóth, R., Ludlage, J., and Mesbah, A. (2014). A review on data-driven linear parameter-varying modeling approaches: A high-purity distillation column case study. *Journal of Process Control*, 24(4), 272–285.

Belz, J., Nelles, O., Schwingshackl, D., Rehrl, J., and Horn, M. (2017). Order determination and input selection with local model networks. *IFAC-PapersOnLine*, 50(1), 7327–7332.

Garulli, A., Paoletti, S., and Vicino, A. (2012). A survey on switched and piecewise affine system identification. *IFAC Proceedings Volumes*, 45(16), 344–355.

Gringard, M. and Kroll, A. (2017). On optimal offline experiment design for the identification of dynamic TS models: Multi-step signals for uncertainty-minimal consequent parameters. In *Proc. 27th Workshop Computational Intelligence*, 117–138. Dortmund, Germany.

Hong, X. and Harris, C.J. (2001). Variable selection algorithm for the construction of MIMO operating point dependent neurofuzzy networks. *IEEE Transactions on Fuzzy Systems*, 9(1), 88–101.

Kahl, M., Kroll, A., Kästner, R., and Sofsky, M. (2015). Application of model selection methods for the identification of dynamic boost pressure model. *IFAC-PapersOnLine*, 48(25), 829–834.

Kahl, M. and Kroll, A. (2018). Structure identification of dynamical takagi-sugeno fuzzy models by using lpv techniques. *Archives of Data Science, Series A (Online First)*, 5(1), A19, 17 S. online.

Lughofer, E. and Kindermann, S. (2010). SparseFIS: Data-driven learning of fuzzy systems with sparsity constraints. *IEEE Transactions on Fuzzy Systems*, 18(2), 396–411.

Luo, M., Sun, F., Liu, H., and Li, Z. (2014). A novel T-S fuzzy systems identification with block structured sparse representation. *Journal of the Franklin Institute*, 351(7), 3508–3523.

Mejari, M., Piga, D., and Bemporad, A. (2016). Regularized least square support vector machines for order and structure selection of LPV-ARX models. In *Proc. 15th European Control Conf.*, 1649–1654. Aalborg, Denmark.

Münker, T. and Nelles, O. (2017). Generalizing piecewise affine system identification to local model networks. In *Proc. 2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–7.

Murray-Smith, R. and Johansen, T.A. (1995). Local learning in local model networks. In *Proc. 4th Int. Conf. on Artificial Neural Networks*, 40–46. Cambridge, U.K.

Nelles, O. (2001). *Nonlinear system identification*. Springer.

Takagi, T. and Sugeno, M. (1985). Fuzzy identification of systems and its application to modelling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1), 116–132.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.

Tòth, R., Laurain, V., Zheng, W.X., and Poolla, K. (2011). Model structure learning: A support vector machine approach for LPV linear-regression models. In *Proc. 50th IEEE Conf. on Decision and Control and European Control Conf.*, 3129–3197. Orlando, Florida (USA).