LPV model order selection in an LS-SVM setting

Dario Piga and Roland Tóth

Abstract— In parametric identification of Linear Parameter-Varying (LPV) systems, the scheduling dependencies of the model coefficients are commonly parameterized in terms of linear combinations of a-priori selected basis functions. Such functions need to be adequately chosen, e.g., on the basis of some first-principles or expert's knowledge of the system, in order to capture the unknown dependencies of the model coefficient functions on the scheduling variable and, at the same time, to achieve a low-variance of the model estimate by limiting the number of parameters to be identified. This problem together with the well-known model order selection problem (in terms of number of input lags, output lags and input delay of the model structure) in system identification can be interpreted as a trade-off between bias and variance of the resulting model estimate. The problem of basis function selection can be avoided by using a non-parametric estimator of the coefficient functions in terms of a recently proposed Least-Square Support-Vector-Machine (LS-SVM) approach. However, the selection of the model order still appears to be an open problem in the identification of LPV systems via the LS-SVM method. In this paper, we propose a novel reformulation of the LPV LS-SVM approach, which, besides of the non-parametric estimation of the coefficient functions, achieves data-driven model order selection via convex optimization. The properties of the introduced approach are illustrated via a simulation example.

I. INTRODUCTION

Since its introduction in [12], the *Linear Parameter-Varying* (LPV) paradigm has become a promising tool for modeling and control of many real systems (see, e.g, the book [10]). Motivated by the need of accurate and low-complexity LPV models for control design purposes, significant efforts have been spent in the last years for developing efficient identification approaches for LPV systems. In the current literature, the existing LPV identification approaches have been mainly formulated in *discrete time* (DT) and they are categorized by the used model structure. In particular, identification schemes for LPV models in a *state-space* (SS) representation can be found in [14], [5], [7], [21], while identification of LPV *input-output* (IO) models is addressed in [1], [17], [6], [3]. An overview of the available LPV identification approaches can be found in the book [16], where the relation between state-space and input-output representation of LPV systems is also discussed.

The simplest representation of LPV systems considered in the discrete-time IO identification framework is given by an

D. Piga and R. Tóth are with the Control Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, Postbus 513, 5600 MB Eindhoven, The Netherlands. Email: `D.Piga@tue.nl` `R.Toth@tue.nl`

autoregressive with exogenous input (ARX) structure, which is defined, in the *single-input single-output* (SISO) case, as

$$y(t) = \sum_{i=1}^{n_{\mathrm{a}}} a_i^{\mathrm{o}}(p(t)) y(t-i) + \sum_{j=0}^{n_{\mathrm{b}}} b_j^{\mathrm{o}}(p(t)) u(t-j) + e_{\mathrm{o}}(t),$$
(1)

where $t \in \mathbb{Z}$ denotes the discrete time, $e_{\mathrm{o}}(t) \in \mathbb{R}$ is a zero-mean white noise, $u(t) \in \mathbb{R}$ and $y(t) \in \mathbb{R}$ are the measured input and output signals of the system, respectively, and $p(t) \in \mathbb{R}^{n_{\mathrm{P}}}$ is the so-called *scheduling variable*, which ranges in a compact set $\mathbb{P} \subseteq \mathbb{R}^{n_{\mathrm{P}}}$ and, according to the literature (see, e.g., [16]), it is assumed to be measurable. The coefficients a_i^{o} and b_j^{o} are functions of the scheduling signal p and they describe the varying linear dynamical relation between the input and the output signals. For clarity of exposition, in this paper we assume that $p(t)$ is scalar (i.e., $n_{\mathrm{P}} = 1$) and the coefficients $a_i^{\mathrm{o}}(p(t))$ and $b_j^{\mathrm{o}}(p(t))$ have a static dependence on p, i.e., $a_i^{\mathrm{o}}(p(t))$ and $b_j^{\mathrm{o}}(p(t))$ only depend on the value of p at time t. The following model structure is used to describe the LPV-ARX system (1):

$$y(t) = \sum_{i=1}^{n_{\mathrm{a}}} a_i(p(t)) y(t-i) + \sum_{j=0}^{n_{\mathrm{b}}} b_j(p(t)) u(t-j) + e(t), \quad (2)$$

where $e(t)$ denotes the residual term. In the parametric identification of LPV systems, the dependence of the co-efficient functions a_i and b_j on the scheduling parameter p is commonly parameterized in terms of a linear combination of a finite number of *a-priori* chosen basis functions $\psi_s : \mathbb{P} \to \mathbb{R}$ in the variable $p(t)$. More specifically,

$$a_i(p(t)) = \sum_{s=1}^{n_\alpha} a_{i,s} \psi_s(p(t)), \qquad i = 1, \dots, n_{\mathrm{a}}, \quad (3a)$$

$$b_j(p(t)) = \sum_{s=1}^{n_\alpha} b_{j,s} \psi_s(p(t)), \qquad j = 0, \dots, n_{\mathrm{b}}, \quad (3b)$$

where $\{\psi_s\}_{s=1}^{n_\alpha} : \mathbb{P} \to \mathbb{R}$ is a set of known basis functions in the scheduling parameter $p(t)$, while $\{a_{i,s} \in \mathbb{R}\}_{s=1}^{n_\alpha}$ and $\{b_{j,s} \in \mathbb{R}\}_{s=1}^{n_\alpha}$ are unknown constant parameters.

The choice of the number and of the type of the basis functions ψ_s is a critical issue in LPV identification. In fact, in order to adequately describe the unknown dependence of the coefficient functions a_i and b_j on the scheduling variable p, and thus to avoid structural bias, a large set of basis functions ψ_s is typically chosen. This leads to an over-parametrization of a_i and b_j. As a consequence, the estimated parameters $\{a_{i,s} \in \mathbb{R}\}_{s=1}^{n_\alpha}$ and $\{b_{j,s} \in \mathbb{R}\}_{s=1}^{n_\alpha}$ tend to have a large variance. This problem is known as bias/variance trade-off and emerges not only in selecting the basis set $\{\psi_s\}_{s=1}^{n_\alpha}$, but also in choosing the order of the LPV model, namely,

n_a, n_b and a possible delay of the input channel. In fact, an under-parameterized model structure with low values of n_a and n_b might not adequately explain the dynamic behavior of the system, while an over-parameterized LPV model described by a large set of coefficient functions $a_i(p(t))$ and $b_j(p(t))$ leads to a high variance on the final estimates of $a_i(p(t))$ and $b_j(p(t))$. A possible way to overcome this problem is offered by the sparse estimation methods, like the *Non-Negative Garrote* (NNG) [2], the LASSO [15] and the SPARSEVA [11]. The underlying idea of the sparse estimators is to consider a model structure able to describe a large set of possible dynamics. A penalization term (typically the ℓ_1-norm of the parameters) is then included in the identification process in order to enforce sparsity of the coefficient estimates $\{a_{i,s} \in \mathbb{R}\}_{s=1}^{n_\alpha}$ and $\{b_{j,s} \in \mathbb{R}\}_{s=1}^{n_\alpha}$. In this way, the model "best suited" for the approximation of the underlying system is chosen directly from the data. Application of the SPARSEVA and the NNG approach for sparse identification of LPV-ARX models is discussed in [18] and [20], respectively.

Alternative approaches to parametric identification of time-varying and LPV systems are presented in [8], [4] and [19], where the identification problem is formulated in the *Least-Squares Support-Vector-Machine* (LS-SVM) setting [13]. More precisely, in [19] the underlying dependence of the coefficient functions a_i and b_j on p is not a-priori parameterized. In this way, the problem of bias/variance trade-off is partially overcome since the basis set $\{\psi_s\}_{s=1}^{n_\alpha}$ does not need to be a-priori specified and the dependence of the coefficient functions a_i and b_j on p is directly reconstructed from data. However, the problem of selecting the model order (i.e., the parameters n_a, n_b and a possible delay in the input channel) is not yet addressed in the LPV LS-SVM framework. This paper aims at improving the original LS-SVM based LPV identification scheme introduced in [19], by presenting a method for selecting the model order via an LS-SVM based identification approach. This allows to jointly reconstruct the scheduling-parameter dependencies and the LPV model order directly from data, with no prior parametrization of the p-dependent functions a_i and b_j.

The paper is organized as follows: the proposed approach for LPV model order selection in the LS-SVM setting is discussed in Section II, where a quadratic programming problem is formulated to select the dynamical order of the LPV model under no assumptions on the underlying dependence of the coefficient functions a_i and b_j on p. A simulation example is presented in Section III to demonstrate the effectiveness of the method.

II. LPV MODEL ORDER SELECTION

An extension of the LPV LS-SVM approach, which is capable of joint selection of the model order together with the non-parametric estimation of the model coefficients $a_i(p(t))$ and $b_j(p(t))$, is presented in this section. The symbol $\mathbb{I}_n^m$ will be adopted throughout the paper to denote the set of indexes $\{n,\ n+1,\ \dots,\ m\}$.

A. Primal formulation

Consider the LPV model structure introduced in (2), which is rewritten as

$$y(t) = \sum_{i=1}^{n_\mathrm{a}} a_i(p(t))x_i(t) + \sum_{j=0}^{n_\mathrm{b}} b_j(p(t))x_{j+n_\mathrm{a}+1}(t) + e(t),$$
(4)

where $x_i(t)$ denotes the i-th component of the $n_\mathrm{f} = n_\mathrm{a} + n_\mathrm{b} + 1$-dimensional vector $x(t)$ defined as

$$x(t) = [y(t-1)\ \dots\ y(t-n_\mathrm{a})\ u(t)\ \dots\ u(t-n_\mathrm{b})]^\top.$$

Similar to the parametric case (see (3)), let us write the p-dependent functions $a_i(p(t))$ and $b_j(p(t))$ in (4) as

$$a_i(p(t)) = \theta_i^\top \phi_i(p(t)) \qquad\qquad i = 1,\dots,n_\mathrm{a}, \quad \text{(5a)}$$
$$b_j(p(t)) = \theta_{j+n_\mathrm{a}+1}^\top \phi_{j+n_\mathrm{a}+1}(p(t)) \quad j = 0,\dots,n_\mathrm{b}, \quad \text{(5b)}$$

where each $\theta_i^\top \in \mathbb{R}^{n_\mathrm{H}}$ is a vector of parameters and ϕ_i (with $i = 1,\dots,n_\mathrm{a} + n_\mathrm{b} + 1$) is a nonlinear map from $\mathbb{P}$ to an n_H-dimensional space, commonly referred as the *feature space*. Unlike in parametric identification of LPV models, neither the maps ϕ_i nor the dimension n_H of the vectors $\theta_i^\top$ and $\phi_i(p(t))$ are a-priori specified. Potentially, $\theta_i^\top$ and $\phi_i(p(t))$ can be infinite-dimensional vectors (i.e, $n_\mathrm{H} = \infty$). Based on the previously introduced notation, the LPV model (4) can be written in the compact form

$$y(t) = \sum_{i=1}^{n_\mathrm{f}} \theta_i^\top \phi_i(t)x_i(t) + e(t),$$
(6)

where $\phi_i(t)$ is used as shorthand notation to indicate $\phi_i(p(t))$. Based on a finite record of input, output and scheduling parameter measurements, our aim is to enforce sparsity in the estimate of the functions $a_i(p(t))$ (with $i = 1,\dots,n_\mathrm{a}$) and $b_j(p(t))$ (with $j = 0,\dots,n_\mathrm{b}$). In this way, the model structure of system (1) can be detected and then used to estimate the nonzero functions $a_i(p(t))$ and $b_j(p(t))$ characterizing the model in (2). To this aim, let us grid the set $\mathbb{P}$ into M points $\mathcal{P} = \{m_k\}_{k=1}^M$. Such points will be referred in the sequel as nodes of $\mathbb{P}$. The idea underlying the proposed method is based on the minimization of the multicriteria objective function

$$\mathcal{J}(\theta,e) = \frac{1}{2}\sum_{i=1}^{n_\mathrm{f}} \theta_i^\top \theta_i + \frac{\lambda}{2}\sum_{t=1}^{N} e^2(t) + \mu \left\|[\overline{\eta}_1\ \dots\ \overline{\eta}_{n_\mathrm{f}}]^\top\right\|_1,$$
(7)

where λ and μ are positive constants (regularization parameters) and $\overline{\eta}_i$ denotes the maximum absolute value of the function $\theta_i^\top \phi_i$ over the nodes $\mathcal{P} = \{m_k\}_{k=1}^M$, i.e.,

$$\overline{\eta}_i = \max_{k=1,\dots,M} \left|\theta_i^\top \phi_i(m_k)\right|.$$
(8)

Note that three criteria are considered in the definition of the objective function $\mathcal{J}(\theta,e)$. More precisely, the term $\sum_{t=1}^{N} e^2(t)$ aims at minimizing the prediction error, while the regularization term $\sum_{i=1}^{n_\mathrm{f}} \theta_i^\top \theta_i$ is added in $\mathcal{J}(\theta,e)$ to prevent overfitting. In fact, since the dimension n_H of the

parameter vector θ_i is not specified and it can be potentially infinite, penalizing the 2-norm of θ_i is essential to achieve an accurate estimate of the functions $a_i(p(t))$ and $b_j(p(t))$ in terms of the bias/variance trade-off. The latter term $\left\| \begin{bmatrix} \overline{\eta}_1 & \dots & \overline{\eta}_{\mathrm{nf}} \end{bmatrix}^\top \right\|_1$ is introduced to enforce sparsity in the estimate of the model. More specifically, the regularization term $\left\| \begin{bmatrix} \overline{\eta}_1 & \dots & \overline{\eta}_{\mathrm{nf}} \end{bmatrix}^\top \right\|_1$ in (7) aims at shrinking $\theta_i^\top \phi_i$ to the zero function in order to minimize the number of non-zero coefficient functions $a_i(p(t))$ and $b_j(p(t))$ characterizing the chosen LPV model structure.

By introducing the slack variables $r = \{r_i\}_{i=1}^{n_f}$, the considered identification problem can be formulated in terms of a quadratic programm:

$$\min_{\theta,e,r} \frac{1}{2}\sum_{i=1}^{n_f}\theta_i^\top\theta_i \; + \; \frac{\lambda}{2}\sum_{t=1}^{N}e^2(t) \; + \mu\sum_{i=1}^{n_f}r_i, \tag{9a}$$

$$\text{s.t. } e(t) = y(t) - \sum_{i=1}^{n_f}\theta_i^\top \phi_i(t)x_i(t), \quad t \in \mathbb{I}_1^N, \tag{9b}$$

$$- r_i \le \theta_i^\top \phi_i(m_k) \le r_i, \; i \in \mathbb{I}_1^{n_f}, \; k \in \mathbb{I}_1^M. \tag{9c}$$

Note that the parameters θ_i minimizing problem (9) cannot be computed since it would require an explicit representation of the feature maps $\{\phi_i(t)\}_{i=1}^{n_f}$. In order to estimate both the parameters θ_i and the feature maps $\phi_i(t)$ together, the dual formulation of (9) is considered.

B. Dual formulation

Define the *Lagrangian* $\mathcal{L}(\theta,e,r,\alpha,\beta^+,\beta^-)$ associated with problem (9) as

$$\mathcal{L}(\theta,e,r,\alpha,\beta^+,\beta^-) = \frac{1}{2}\sum_{i=1}^{n_f}\theta_i^\top\theta_i + \frac{\lambda}{2}\sum_{t=1}^{N}e^2(t) + \mu\sum_{i=1}^{n_f}r_i$$
$$- \sum_{t=1}^{N}\alpha_t\left(e(t) - y(t) + \sum_{i=1}^{n_f}\theta_i^\top\phi_i(t)x_i(t)\right)$$
$$- \sum_{i=1}^{n_f}\sum_{k=1}^{M}\beta_{i,k}^+\left(r_i - \theta_i^\top\phi_i(m_k)\right)$$
$$- \sum_{i=1}^{n_f}\sum_{k=1}^{M}\beta_{i,k}^-\left(r_i + \theta_i^\top\phi_i(m_k)\right), \tag{10}$$

where $\alpha = \{\alpha_t\}_{t=1}^N$, $\beta^+ = \{\beta_{i,k}^+\}_{i=1,k=1}^{n_f,M}$, $\beta^- = \{\beta_{i,k}^-\}_{i=1,k=1}^{n_f,M}$ are the *Lagrangian multipliers* associated with the constraints defining the feasible set of problem (9). Define the *Lagrange dual function* $g : \mathbb{R}^{Mn_f+N} \to \mathbb{R}$ as the minimum of $\mathcal{L}(\theta,e,r,\alpha,\beta^+,\beta^-)$ over (θ,e,r), i.e.,

$$g(\alpha,\beta^+,\beta^-) = \inf_{\theta,e,r}\mathcal{L}(\theta,e,r,\alpha,\beta^+,\beta^-). \tag{11}$$

Since $\mathcal{L}(\theta,e,r,\alpha,\beta^+,\beta^-)$ is a convex quadratic function of (θ,e,r), the Lagrange dual function $g(\alpha,\beta^+,\beta^-)$ can be computed from the (necessary and sufficient) optimality

conditions:

$$\frac{\partial\mathcal{L}}{\partial\theta_i} = 0 \to \theta_i = \sum_{t=1}^{N}\alpha_t\phi_i(t)x_i(t) + \sum_{k=1}^{M}\left(\beta_{i,k}^- - \beta_{i,k}^+\right)\phi_i(m_k), \tag{12a}$$

$$\frac{\partial\mathcal{L}}{\partial e(t)} = 0 \to e(t) = \frac{1}{\lambda}\alpha_t, \tag{12b}$$

$$\frac{\partial\mathcal{L}}{\partial r_i} = 0 \to \mu = \sum_{k=1}^{M}\left(\beta_{i,k}^+ + \beta_{i,k}^-\right), \tag{12c}$$

where the conditions in (12a) and (12c) are need to be satisfied for all $i \in \mathbb{I}_1^{n_f}$, while (12b) holds for all $t \in \mathbb{I}_1^N$. From substitution of conditions (12) into the definition of the *Lagrangian* $\mathcal{L}(\theta,e,r,\alpha,\beta^+,\beta^-)$, the following expression of the Lagrange dual function is obtained:

$$g(\alpha,\beta^+,\beta^-) = -\frac{1}{2\lambda}\alpha^\top\alpha + \alpha^\top Y - \frac{1}{2}\sum_{i=1}^{n_f}\begin{bmatrix}\alpha \\ \beta_i^- - \beta_i^+\end{bmatrix}^\top\Omega_i\begin{bmatrix}\alpha \\ \beta_i^- - \beta_i^+\end{bmatrix} \tag{13}$$

with

$$\beta_i^+ = \begin{bmatrix}\beta_{i,1}^+ & \dots & \beta_{i,M}^+\end{bmatrix}^\top, \beta_i^- = \begin{bmatrix}\beta_{i,1}^- & \dots & \beta_{i,M}^-\end{bmatrix}^\top, \tag{14}$$

and

$$\Omega_i = \begin{bmatrix} X_i\big(\Phi_i^{(N)}\big)^\top\Phi_i^{(N)}X_i & X_i\big(\Phi_i^{(N)}\big)^\top\Phi_i^{(M)} \\ \big(\Phi_i^{(M)}\big)^\top\Phi_i^{(N)}X_i & \big(\Phi_i^{(M)}\big)^\top\Phi_i^{(M)} \end{bmatrix}. \tag{15}$$

In (15), X_i is a diagonal matrix of size N whose diagonal entries are $x_i(1),\dots,x_i(N)$, while $\Phi_i^{(N)}$, $\Phi_i^{(M)}$, β_i^+ and β_i^- are defined as

$$\Phi_i^{(N)} = \begin{bmatrix} \phi_i(p(1)) & \phi_i(p(2)) & \dots & \phi_i(p(N)) \end{bmatrix}, \tag{16a}$$

$$\Phi_i^{(M)} = \begin{bmatrix} \phi_i(m_1) & \phi_i(m_2) & \dots & \phi_i(m_M) \end{bmatrix}, \tag{16b}$$

Note that the obtained Lagrange dual function $g(\alpha,\beta^+,\beta^-)$ is a concave function since the matrix Ω_i is positive semidefinite. As a matter of fact, Ω_i is defined by the inner product

$$\begin{bmatrix} X_i\big(\Phi_i^{(N)}\big)^\top \\ \big(\Phi_i^{(M)}\big)^\top \end{bmatrix}\begin{bmatrix} \Phi_i^{(N)}X_i & \Phi_i^{(M)} \end{bmatrix}.$$

The *Lagrange dual problem* associated with (9) is then given by the convex optimization problem:

$$\min_{\alpha,\beta^+,\beta^-} - g(\alpha,\beta^+,\beta^-) \tag{17a}$$

$$\text{s.t. } \mu = \sum_{k=1}^{M}\left(\beta_{i,k}^+ + \beta_{i,k}^-\right) \tag{17b}$$

$$\beta_{i,k}^+, \beta_{i,k}^- \ge 0, \; i \in \mathbb{I}_1^{n_f}, \; k \in \mathbb{I}_1^M. \tag{17c}$$

Note that, since the primal problem (9) is a convex quadratic problem with linear equality constraints, strong duality holds for (9). This means that the solution of (9) can be evaluated from its dual formulation. Specifically, once the Lagrangian multipliers are computed by solving the (convex) optimization (17), the parameters θ_i minimizing (9) are obtained from the optimality conditions (12a). However, since we started with the assumption that the feature maps ϕ_i

are a-priori unknown and their dimension n_H is potentially infinite, the parameters θ_i can not be explicitly computed neither via (12a) nor via the solution of the primal problem (9). On the other hand, thanks to the structure of the matrices Ω_i (eq. (15)) characterizing the Lagrange dual function $g(\alpha, \beta^+, \beta^-)$ in (13), the dual problem (17) can be solved without a prior specification of the feature maps $\Phi_i^{(N)}$ and $\Phi_i^{(M)}$, as discussed in the sequel.

Let us rewrite the matrix Ω_i in the block form:

$$\Omega_i = \begin{bmatrix} \Omega_i^{(N,N)} & \Omega_i^{(N,M)} \\ \left(\Omega_i^{(N,M)}\right)^\top & \Omega_i^{(M,M)} \end{bmatrix}, \quad (18)$$

where the (j,k)-th entry of the matrices $\Omega_i^{(N,N)}, \Omega_i^{(N,M)}$ and $\Omega_i^{(M,M)}$ is given by

$$\left[\Omega_i^{(N,N)}\right]_{j,k} = x_i(j)\underbrace{\phi_i^\top(p(j))\phi_i(p(k))}_{K_i(p(j),p(k))}x_i(k), \quad (19a)$$

$$\left[\Omega_i^{(N,M)}\right]_{j,k} = x_i(j)\underbrace{\phi_i^\top(p(j))\phi_i(m_k)}_{K_i(p(j),m_k)}, \quad (19b)$$

$$\left[\Omega_i^{(M,M)}\right]_{j,k} = \underbrace{\phi_i^\top(m_j)\phi_i(m_k)}_{K_i(m_j,m_k)}. \quad (19c)$$

In (19), $K_i(\bullet, \bullet)$ is a positive definite *kernel function* K_i : $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ satisfying the Mercer's condition [9] and defining the inner products $\phi_i^\top(p(j))\phi_i(p(k))$, $\phi_i^\top(p(j))\phi_i(m_k)$ and $\phi_i^\top(m_j)\phi_i(m_k)$. Specification of the kernels instead of the maps ϕ_i is called the *kernel trick* [13], which allows the characterization of the Lagrange dual function $g(\alpha, \beta^+, \beta^-)$ and consequently the computation of the solution of problem (17) without explicitly defining the feature maps ϕ_i. *Radial basis functions* (RBFs) are typically chosen as kernels, i.e.,

$$K_i(p(j), p(k)) = \exp\left(-\frac{(p(j) - p(k))^2}{\sigma^2}\right), \ i \in \mathbb{I}_1^{n_f}, \quad (20)$$

where $\sigma > 0$ is a hyper-parameter chosen by the user to control the width of the RBF.

Once the Lagrangian multipliers α, β^+, β^- are computed through (17), the coefficients functions $a_i(\bullet)$ and $b_j(\bullet)$ are estimated from (5) and the optimality conditions (12a), i.e.,

$$a_i(\bullet) = \theta_i^\top \phi_i(\bullet) = \sum_{t=1}^N \alpha_t K_i(p(t), \bullet)x_i(t) + \\ + \sum_{k=1}^M \left(\beta_{i,k}^- - \beta_{i,k}^+\right)K_i(m_k, \bullet), i \in \mathbb{I}_1^{n_a},$$

$$b_j(\bullet) = \theta_j^\top \phi_j(\bullet) = \sum_{t=1}^N \alpha_t K_j(p(t), \bullet)x_j(t) + \\ + \sum_{k=1}^M \left(\beta_{j,k}^- - \beta_{j,k}^+\right)K_j(m_k, \bullet), j \in \mathbb{I}_{n_a+1}^{n_f},$$

Note that only the combined estimate $\theta_i^\top \phi_i(\bullet)$ can be computed by means of the defined kernel function K_i. On the other hand, the estimate of the parameter vectors θ_i is never accessible separately.

III. SIMULATION EXAMPLE

The ability of the proposed approach for joint selection of the model order together with the non-parametric estimation of the model coefficients $a_i(p(t))$ and $b_j(p(t))$ is shown in this section by means of an extensive Monte-Carlo study based on a simulation example. The considered LPV data-generating system is described by the difference equation

$$y(t) = a_1^o(p(t))y(t-1) + a_2^o(p(t))y(t-2) + \\ + b_5^o(p(t))u(t-5) + e_o(t), \quad (21)$$

where $e_o(t)$ is a white noise with Gaussian distribution $\mathcal{N}(0, \sigma_e^2)$ and standard deviation $\sigma_e = 0.1$. The scheduling-dependent coefficients $a_1^o(p(t))$, $a_2^o(p(t))$, and $b_5^o(p(t))$ are described by the nonlinear functions:

$$a_1^o(p(t)) = \begin{cases} -0.5, & \text{if } p(t) > 0.5; \\ -p(t), & \text{if } -0.5 \leq p(t) \leq 0.5; \\ 0.5, & \text{if } p(t) < -0.5; \end{cases} \quad (22a)$$

$$a_2^o(p(t)) = \sin(2\pi p(t)), \quad (22b)$$

$$b_5^o(p(t)) = p^3(t). \quad (22c)$$

For estimation purposes, a data set $\mathcal{D}_N = \{u(t), y(t), p(t)\}_{t=1}^N$ with $N = 500$ is gathered from the system with u and p being white-noise processes, independent of other, with uniform distribution $\mathcal{U}(-1, 1)$. To analyze the statistical properties of the estimation, a Monte-Carlo study of 100 runs is performed. At each run, new realizations of u, p and e_o are considered to generate the data set for estimation. The average of the *signal-to-noise ratio* (SNR) over the generated data is equal to 12 dB, where the SNR is defined, at each run, as

$$10\log_{10}\left(\frac{\sum_{t=1}^N w_o^2(t)}{\sum_{t=1}^N e_o^2(t)}\right), \quad (23)$$

with $w_o(t)$ denoting the noise-free system output.

The following over-parameterized LPV model is used to describe the data-generating system:

$$y(t) = \sum_{i=1}^{n_a} a_i(p(t))y(t-i) + \sum_{j=0}^{n_b} b_j(p(t))u(t-j) + e(t), \quad (24)$$

with $n_a = n_b = 10$. The dependence of $a_i(p(t))$ and $b_j(p(t))$ on p is not specified. First, the coefficient functions $a_i(p(t))$ and $b_j(p(t))$ are estimated through the LPV LS-SVM approach proposed in [19], corresponding to the minimization of the objective function $\mathcal{J}(\theta, e)$ in (7) for $\mu = 0$. Radial basis functions are used as kernels. The values of the hyper-parameters λ (in (7)) and σ (characterizing the RBF K_i in (20)) are chosen via cross-validation based optimization, that is by maximizing (with a grid search) the *best fit rate* (BFR) w.r.t. a validation data set of length $N_V = 200$, where the BFR is defined as

$$\text{BFR} = \max\left\{0, 1 - \frac{\|y(t) - \hat{y}(t)\|_2}{\|y(t) - \overline{y}\|_2}\right\}, \quad (25)$$

with $\hat{y}(t)$ being the simulated model output and $\overline{y}$ is the sample mean of the output over the validation data set. The

TABLE I

AVERAGE AND STANDARD DEVIATION (OVER THE 100 MONTE-CARLO RUNS) OF THE MAXIMUM ABSOLUTE VALUE $\overline{a}_i$ OF THE COEFFICIENTS FUNCTIONS $a_i(p(t))$ ESTIMATED BY MEANS OF THE LS-SVM AND THE REGULARIZED LS-SVM (RLS-SVM) APPROACH.

	$\overline{a}_1$	$\overline{a}_2$	$\overline{a}_3$	$\overline{a}_4$	$\overline{a}_5$	$\overline{a}_6$	$\overline{a}_7$	$\overline{a}_8$	$\overline{a}_9$	$\overline{a}_{10}$
True value	0.5	1	0	0	0	0	0	0	0	0
Mean (LS-SVM)	$8.3 \cdot 10^{-1}$	1.1	$4.4 \cdot 10^{-1}$	$4.4 \cdot 10^{-1}$	$4.8 \cdot 10^{-1}$	$4.6 \cdot 10^{-1}$	$4.7 \cdot 10^{-1}$	$4.6 \cdot 10^{-1}$	$4.8 \cdot 10^{-1}$	$3.9 \cdot 10^{-1}$
Mean (RLS-SVM)	$3.2 \cdot 10^{-1}$	$6.3 \cdot 10^{-1}$	$2.7 \cdot 10^{-4}$	$3.8 \cdot 10^{-4}$	$1.1 \cdot 10^{-4}$	$9.3 \cdot 10^{-5}$	$3.4 \cdot 10^{-5}$	$7.4 \cdot 10^{-4}$	$3.1 \cdot 10^{-4}$	$6.2 \cdot 10^{-5}$
std (LS-SVM)	$2.1 \cdot 10^{-1}$	$1.0 \cdot 10^{-1}$	$2.1 \cdot 10^{-1}$	$2.1 \cdot 10^{-1}$	$2.5 \cdot 10^{-1}$	$2.2 \cdot 10^{-1}$	$2.3 \cdot 10^{-1}$	$2.1 \cdot 10^{-1}$	$2.1 \cdot 10^{-1}$	$1.9 \cdot 10^{-1}$
std (RLS-SVM)	$1.2 \cdot 10^{-1}$	$1.4 \cdot 10^{-1}$	$8.6 \cdot 10^{-5}$	$1.1 \cdot 10^{-4}$	$5.2 \cdot 10^{-5}$	$3.8 \cdot 10^{-5}$	$7.3 \cdot 10^{-6}$	$2.1 \cdot 10^{-4}$	$1.0 \cdot 10^{-4}$	$1.1 \cdot 10^{-5}$

TABLE II

AVERAGE AND STANDARD DEVIATION (OVER THE 100 MONTE-CARLO RUNS) OF THE MAXIMUM ABSOLUTE VALUE $\overline{b}_j$ OF THE COEFFICIENTS FUNCTIONS $b_j(p(t))$ ESTIMATED BY MEANS OF THE LS-SVM AND THE REGULARIZED LS-SVM (RLS-SVM) APPROACH.

	$\overline{b}_0$	$\overline{b}_1$	$\overline{b}_2$	$\overline{b}_3$	$\overline{b}_4$	$\overline{b}_5$	$\overline{b}_6$	$\overline{b}_7$	$\overline{b}_8$	$\overline{b}_9$	$\overline{b}_{10}$
True value	0	0	0	0	0	1	0	0	0	0	0
Mean (LS-SVM)	$3.3 \cdot 10^{-1}$	$3.0 \cdot 10^{-1}$	$3.5 \cdot 10^{-1}$	$3.2 \cdot 10^{-1}$	$3.3 \cdot 10^{-1}$	1.06	$2.8 \cdot 10^{-1}$	$4.6 \cdot 10^{-1}$	$4.8 \cdot 10^{-1}$	$4.1 \cdot 10^{-1}$	$4.2 \cdot 10^{-1}$
Mean (RLS-SVM)	$7.4 \cdot 10^{-5}$	$3.4 \cdot 10^{-5}$	$1.4 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	$5.5 \cdot 10^{-4}$	$3.5 \cdot 10^{-1}$	$6.1 \cdot 10^{-5}$	$4.8 \cdot 10^{-4}$	$8.2 \cdot 10^{-5}$	$2.8 \cdot 10^{-4}$	$8.7 \cdot 10^{-5}$
std (LS-SVM)	$1.7 \cdot 10^{-1}$	$1.3 \cdot 10^{-1}$	$2.0 \cdot 10^{-1}$	$2.1 \cdot 10^{-1}$	$1.6 \cdot 10^{-1}$	$2.1 \cdot 10^{-1}$	$1.5 \cdot 10^{-1}$	$2.1 \cdot 10^{-1}$	$2.1 \cdot 10^{-1}$	$1.6 \cdot 10^{-1}$	$1.5 \cdot 10^{-1}$
std (RLS-SVM)	$1.8 \cdot 10^{-5}$	$8.3 \cdot 10^{-6}$	$1.4 \cdot 10^{-4}$	$7.7 \cdot 10^{-5}$	$2.9 \cdot 10^{-5}$	$1.8 \cdot 10^{-1}$	$2.4 \cdot 10^{-5}$	$1.0 \cdot 10^{-4}$	$5.2 \cdot 10^{-5}$	$9.9 \cdot 10^{-5}$	$3.9 \cdot 10^{-5}$

obtained values of λ and σ have been: $\lambda = 600$ and $\sigma = 0.4$. The estimates of the functions a_1, a_2 and b_5 are depicted in Fig. 1(a), Fig. 2(a) and Fig. 3(a), where the mean of the estimated functions over the 100 Monte-Carlo runs is plotted together with their standard deviation.

The regularized LS-SVM approach described in Section II is now applied to select the order of the considered LPV model. The interval $\mathbb{P} = [-1\ 1]$ is gridded into 7 nodes $\mathcal{P} = \{m_k\}_{k=1}^7 = \{-1, -0.66, -0.33, 0, 0.33, 0.66, 1\}$. In order to provide a fair comparison w.r.t. the original LS-SVM approach, the values of λ and σ optimized for the LS-SVM method (i.e., $\lambda = 600$ and $\sigma = 0.4$), are used. The regularization parameter μ is computed by means of an exhaustive search aiming at the maximization of the BFR w.r.t. the validation set. The obtained value of μ has been 350.

The maximum absolute value $\overline{a_i}$ and $\overline{b_j}$ of the coefficients functions $a_i(p(t))$ and $b_j(p(t))$ estimated via the original LS-SVM approach and its regularized version is reported in Table I and Table II, which show the average and the standard deviation of $\overline{a_i}$ and $\overline{b_j}$ over the 100 Monte-Carlo runs. It is important to highlight that $\overline{a_i}$ and $\overline{b_j}$ are the maximum of $|a_i(.)|$ and $|b_j(.)|$ over the whole interval $\mathbb{P} = [-1\ \ 1]$, and not only over the chosen nodes $\mathcal{P}$. Results in Table I and Table II show that the regularized LS-SVM approach correctly detects the LPV model structure. In fact, the only coefficient functions with an (average) maximum absolute value greater than a threshold of 10^{-3} are a_1, a_2 and b_5, which are exactly the nonzero coefficient functions defining the considered data-generating system in (21). Note that, because of the ℓ_1-penalty term $\mu \| [\overline{\eta}_1\ \dots\ \overline{\eta}_{\mathrm{nf}}]^\top \|_1$ introduced in (9) to shrink the coefficient functions a_i and b_j to zero, the estimates of a_1, a_2 and b_5 are biased. In fact, the estimated maximum values of $|a_1(.)|$, $|a_2(.)|$ and $|b_5(.)|$ over the interval $\mathbb{P}$ are 0.32, 0.63 and 0.35, respectively, while the corresponding true values are 0.5, 1 and 1. Nevertheless, it is important to point out that the nonzero coefficient functions should be estimated in two stages. First, the proposed regularized LS-SVM approach should be used to select the order of the LPV model. Then,

the zero coefficient functions should be discarded in the description of the LPV model (24) and a lower-complexity LPV model should be re-identified without considering the regularization term $\mu \| [\overline{\eta}_1\ \dots\ \overline{\eta}_{\mathrm{nf}}]^\top \|_1$. According to this two-stage approach, the nonzero coefficient functions a_1, a_2 and b_5 are then re-estimated. The obtained results are shown in Fig.1(b), Fig. 2(b) and Fig. 3(b), where the mean estimated functions a_1, a_2 and b_5 together with the standard deviation intervals over the 100 Monte-Carlo runs are plotted. Results in Fig.1, Fig. 2 and Fig. 3 show that, as expected, detecting the LPV model order is beneficial, in terms of variance reduction, in the final estimate of the coefficient functions.

IV. CONCLUSIONS

This paper has addressed the problem of identifying LPV-ARX models in the LS-SVM framework, where the underlying dependence of the coefficient functions on the scheduling parameter is not a-priori parameterized. The method discussed in the paper provides a systematic way to estimate the dynamical order of LPV-ARX models by using only a set of measured data generated by the plant, and without exploiting any a-priori information on the underlying behavior of the true LPV system. The reported simulation has shown the effectiveness of the proposed approach to detect the LPV model order and thus to provide a lower-variance estimate (w.r.t. to the original LPV LS-SVM identification method introduced in [19]) of the coefficient functions describing the LPV model.

REFERENCES

[1] B. A. Bamieh and L. Giarré. Identification of linear parameter-varying models. *Int. Journal of Robust and Nonlinear Control*, 12(9):841–853, 2002.

[2] L. Breiman. Better subset regression using the nonnegative garrote. *Technometrics*, pages 373–384, 1995.

[3] V. Cerone, D. Piga, and D. Regruto. A convex relaxation approach to set-membership identication of LPV systems. *Automatica*, 49(9):2853–2859, 2013.

[4] T. Falck, H. Ohlsson, L. Ljung, J. Suykens, and B. De Moor. Segmentation of times series from nonlinear dynamical systems. In *Proc. of the 18th IFAC World Congress*, pages 13209–13214, Milan, Italy, 2011.

[5] F. Felici, J. W. van Wingerden, and M. Verhaegen. Subspace identification of MIMO LPV systems using a periodic scheduling sequence. *Automatica*, 43:1684–1697, 2007.

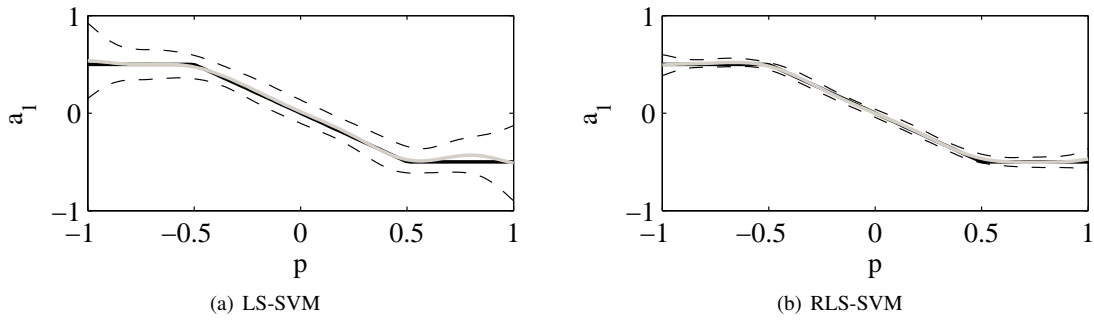(a) LS-SVM   (b) RLS-SVM

Fig. 1. Coefficient function $a_1(p(t))$: (a) Estimate via the LS-SVM approach; (b) Estimate after model order selection. True function (solid black line), mean estimate (solid gray line) and standard deviation intervals (dashed black line) over the 100 Monte Carlo runs.
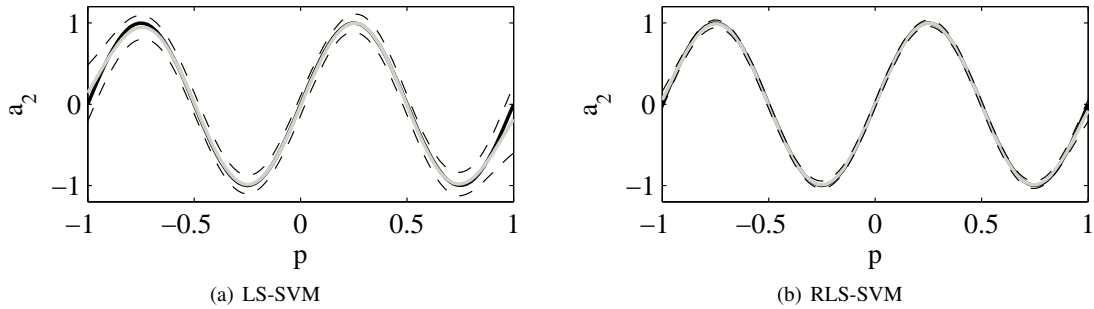

(a) LS-SVM   (b) RLS-SVM

Fig. 2. Coefficient function $a_2(p(t))$: (a) Estimate via the LS-SVM approach; (b) Estimate after model order selection. True function (solid black line), mean estimate (solid gray line) and standard deviation intervals (dashed black line) over the 100 Monte Carlo runs.
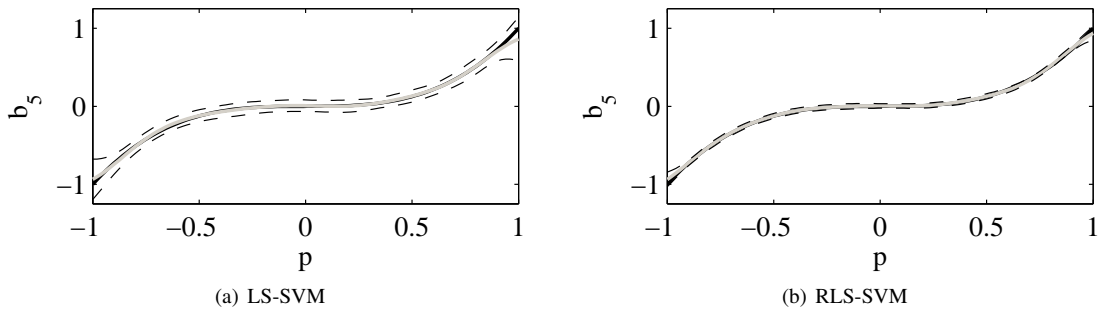

(a) LS-SVM   (b) RLS-SVM

Fig. 3. Coefficient function $b_5(p(t))$: (a) Estimate via the LS-SVM approach; (b) Estimate after model order selection. True function (solid black line), mean estimate (solid gray line) and standard deviation intervals (dashed black line) over the 100 Monte Carlo runs.

[6] V. Laurain, M. Gilson, R. Tóth, and H. Garnier. Refined instrumental variable methods for identification of LPV Box–Jenkins models. *Automatica*, 46(6):959–967, 2010.

[7] P. Lopes dos Santos, J. A. Ramos, and J. L. Martins de Carvalho. Subspace identification of linear parameter-varying systems with innovation-type noise models driven by general inputs and a measurable white noise time-varying parameter vector. *Int. Journal of Systems Science*, 39(9):897–911, 2008.

[8] S. Mehrkanoon, T. Falck, and J. Suykens. Parameter estimation for time-varying dynamical systems using Least Squares Support Vector Machines. In *Proc. of the 16th IFAC Symposium on System Identification*, pages 1300–1305, Brussels, Belgium, 2012.

[9] J. Mercer. Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London, Series A*, 209:415–446, 1909.

[10] J. Mohammadpour and C. W. Scherer (eds.). *Control of linear parameter varying systems with applications*. Springer, 2012.

[11] C.R. Rojas and H. Hjalmarsson. Sparse estimation based on a validation criterion. In *Proc. of the 50th IEEE Conference on Decision and Control*, pages 2825–2830, Orlando, Florida (USA), 2011.

[12] J. S. Shamma and M. Athans. Analysis of gain scheduled control for nonlinear plants. *IEEE Trans. Autom. Control*, 35(8):898–907, 1990.

[13] J. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific Publishing, Singapore, 2002.

[14] M. Sznaier and C. Mazzarro. An LMI approach to control-oriented identification and model (in)validation of LPV systems. *IEEE Transaction on Automatic Control*, 48(9):1619–1624, 2003.

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

[16] R. Tóth. *Modeling and identification of linear parameter-varying systems*. Lecture Notes in Control and Information Sciences, Vol. 403, Springer, Heidelberg, 2010.

[17] R. Tóth, P. S. C. Heuberger, and P. M. J. Van den Hof. Asymptotically optimal orthonormal basis functions for LPV system identification. *Automatica*, 45(6):1359–1370, 2010.

[18] R. Tóth, H. Hjalmarsson, and C. R. Rojas. Order and structural dependence selection of LPV-ARX models revisited. In *Proc. of the 51st IEEE Conference on Decision and Control*, pages 6271–6276, Maui, Hawaii (USA), 2012.

[19] R. Tóth, V. Laurain, W. X. Zheng, and K. Poolla. Model structure learning: A support vector machine approach for LPV linear-regression models. In *Proc. of the 50th IEEE Conference on Decision and Control*, pages 3192–3197, Orlando, Florida (USA), 2011.

[20] R. Tóth, C. Lyzell, M. Enqvist, P.S.C. Heuberger, and P.M.J. Van den Hof. Order and structural dependence selection of LPV-ARX models using a nonnegative garrote approach. In *Proc. of the 48th IEEE Conference on Decision and Control*, pages 7406–7411, Shanghai, China, 2009.

[21] J. W. van Wingerden and M. Verhaegen. Subspace identification of bilinear and LPV systems for open-and closed-loop data. *Automatica*, 45(2):372–381, 2009.