2016 European Control Conference
June 29 - July 1, 2016. Aalborg, Denmark

Regularized Least Square Support Vector Machines for Order and Structure Selection of LPV-ARX Models

Manas Mejari, Dario Piga and Alberto Bemporad

Abstract— **Least Squares Support Vector Machine (LS-SVM) is a computationally efficient kernel-based regression approach which has been recently applied to nonparametric identification of Linear Parameter Varying (LPV) systems. In contrast to parametric LPV identification approaches, LS-SVM based methods obviate the need to parameterize the scheduling dependence of the LPV model coefficients in terms of a-priori specified basis functions. However, an accurate selection of the underlying model order (in terms of number of input lags, output lags and input delay) is still a critical issue in the identification of LPV systems in the LS-SVM setting. In this paper, we address this issue by extending the LS-SVM method to sparse LPV model identification, which, besides non-parametric estimation of the model coefficients, achieves data-driven model order selection via convex optimization. The main idea of the proposed method is to first estimate the coefficients of an over-parameterized LPV model through LS-SVM. The estimated coefficients are then scaled by polynomial weights, which are shrunk towards zero to enforce sparsity in the final LPV model estimate. The properties of the proposed approach are illustrated via simulation examples.**

I. Introduction

Identification and control of *Linear Parameter-Varying* (LPV) systems has attracted attention of several researchers in the past two decades. LPV systems can be seen as an extension of *Linear Time-Invariant* (LTI) systems, with a linear dynamic relation between the input and the output signals. Unlike LTI systems, the input/output relation can change over time, according to a measurable time-varying signal, the so called *scheduling vector* p. In this way, LPV models can accurately describe the dynamic behaviour of a large class of nonlinear and time-varying systems like aircrafts [10], automobiles [4], [11] and distillation columns [1], preserving input/output linearity. The nonlinear and time-varying dynamics are embedded in the scheduling variables.

Motivated by the need of accurate and low-complexity LPV models, significant efforts have been spent in the last years for developing efficient approaches for identification of discrete-time LPV models, both in state-space [23], [6], [21] and input-output (IO) representation [2], [9], [12]. The reader is referred to [17] for an overview on LPV identification.

As known, a challenging issue in identification is the choice of the model structure. In terms of LPV identification, choosing the model structure requires to specify both the model order (in terms of number of output lags, input lags, and input delay) and the nonlinear dependence of the model coefficients on p. In the input-output case, a first possible choice is to use an over-parameterized model, by approximating the p-dependent coefficients as the sum of a large set of a-priori specified nonlinear basis functions (e.g., polynomials). Then, sparse parametric estimation like the *LASSO* [16], the *Non-Negative Garrote* (NNG) [3] and *SPARSEVA* [14] can be used to select the p-dependent nonlinear basis functions. Application of the NNG and SPARSEVA for sparse identification of LPV-ARX models is discussed in [20] and [18], respectively. However, although sparse estimators extend the fixed parametric nonlinearity modeling, the adequate a-priori selection of the basis functions remains an open-problem.

In order to deal with the basis functions selection problem, kernel-based methods for LPV identification have been recently proposed in [8], [19], [5]. The main idea behind kernel methods is to introduce a feature function mapping the scheduling vector p to a high-dimensional space. The feature maps are not fixed a priori and can be potentially infinite-dimensional. Only the inner products between the feature maps is specified by the user in terms of nonlinear kernel functions, such as *radial basis functions* or *polynomial kernels*. However, although these methods offer strong flexibility in modeling the nonlinear dependence on the scheduling vector p, they do not address the problem of selecting the LPV model order. To the best of our knowledge, the only contribution addressing the issue of LPV model order selection in a kernel-based setting is [13], where *Least-Squares Support Vector Machines* (LS-SVM) (i.e., a specific kernel-based method developed in [15]) are reformulated in order to achieve data-driven model order selection along with non-parametric identification of the p-dependent LPV model coefficients. Specifically, the problem is formulated by using an extra penalty term aiming at minimizing the maximum absolute value of the LPV model coefficient functions over a set of grid points in the scheduling vector space. However, the method in [13] suffers from the following drawbacks: 1. since it is necessary to grid the scheduling space, the number of grid points grow exponentially with the dimension of p, thus increasing the computation load of the optimization problem; 2. the LPV model coefficient functions are enforced to be null only at the chosen gridding points, but nothing can be said outside these points.

The main contribution of this paper is to present a new method for data-driven order selection along with non-parametric identification of the p-dependent LPV model

*This work was partially supported by the European Commission under project H2020-SPIRE-636834 "DISIRE - Distributed In-Situ Sensors Integrated into Raw Material and Energy Feedstock" (http://spire2030.eu/disire/).

The authors are with the IMT School for Advanced Studies Lucca, 55100 Lucca, Italy. E-mails: {manas.mejari, dario.piga, alberto.bemporad} @imtlucca.it

coefficients, overcoming the main drawbacks of [13]. This is achieved with the following three-step approach:

S1. estimate the coefficients for an over-parameterized LPV model using the non-parametric LS-SVM approach proposed in [19];
S2. scale the estimated coefficients using scheduling-variable dependent polynomial weights, which are then penalised to shrink the previously estimated model coefficients towards zero;
S3. re-estimate the non-null model coefficients.

In this way, an accurate model of the system is determined with low bias and less variance in the estimated coefficients. Overall, the proposed method can be seen as a reformulation of the Non-Negative Garrotte approach in a nonparametric framework. The advantage of this approach over [13] is that it is applicable for identification of LPV systems which are dependent on multi-dimensional scheduling variables, as the scheduling space does not need to be gridded. The polynomial weights also provide flexibility in reshaping and correcting the LPV model coefficients obtained in stage S1.

The paper is organized as follows: In Section II, a short review of LPV-ARX model structure and problem statement are given. The standard LS-SVM method for LPV identification is described in Section III. In Section IV, the mathematical details of the proposed regularized LS-SVM approach for model order selection are discussed. The proposed method is tested on two academic examples and simulation results are presented in Section V. Finally, conclusions are given in Section VI.

II. Problem formulation

The analysis driven in this paper is dedicated to LPV models in an I/O form. For clarity of exposition, we consider *single-input single-output* (SISO) LPV systems, described by the *autoregressive with exogenous input* (ARX) structure:

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

where, $t \in \mathbb{N}$ denotes the discrete time; $u(t) \in \mathbb{R}$ and $y(t) \in \mathbb{R}$ are the measured input and output signals of the system, respectively; $e^{\circ}(t)$ is an additive zero-mean white noise; $p(t) : \mathbb{N} \to \mathbb{P}$ is the measured n_p-dimensional scheduling vector (which may include also past observations of the scheduling signals) and $\mathbb{P} \subseteq \mathbb{R}^{n_p}$ is a compact set where $p(t)$ is assumed to take values. The p-dependent coefficient functions a_i° and b_j°, as well as the parameters n_a° and n_b° defining the dynamical order of the system, are unknown and they have to be estimated from an N-length observed sequence $\mathcal{D}_N = \{u(t), y(t), p(t)\}_{t=1}^{N}$ of data generated by the system in (1).

The following model structure is therefore suitable to describe the true LPV data-generating system in (1):

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

with $e(t)$ denoting the residual term. The parameters n_a and n_b defining the dynamical order of the model in (2) are chosen large enough so that $n_a > n_a^{\circ}$ and $n_b > n_b^{\circ}$ (i.e., the true system belongs to the chosen model class).

In the following section, we briefly describe the LPV LS-SVM identification method in [19], which is used for non-parametric estimation of the model coefficients $\{a_i\}_{i=1}^{n_a}$ and $\{b_j\}_{j=0}^{n_b}$. The main advantage of the LPV LS-SVM identification method proposed in [19] is that it obviates the need to specify the underlying dependency of coefficients$\{a_i\}_{i=1}^{n_a}$ and $\{b_j\}_{j=0}^{n_b}$ on the scheduling vector p.

III. LS-SVM for LPV Identification

Let us consider the LPV-ARX model introduced in (2), which is rewritten in the compact form

$$y(t) = \sum_{i=1}^{n_g} c_i(p(t))x_i(t) + e(t), \quad (3)$$

where $x_i(t)$ and $c_i(p(t))$ denote the i-th component of the $n_g = n_\mathrm{a} + n_\mathrm{b} + 1$-dimensional vector $x(t)$ and $c(p(t))$, respectively, defined as

$$x(t) = [y(t-1) \ \dots \ y(t-n_\mathrm{a}) \ u(t) \ \dots \ u(t-n_\mathrm{b})]^\top,$$
$$c(p(t)) = [a_1(p(t)) \ \dots \ a_{n_a}(p(t)) \ b_0(p(t)) \ \dots \ b_{n_b}(p(t))]^\top.$$

The p-dependent coefficient functions $c_i(p(t))$ are written as

$$c_i(p(t)) = \rho_i^\top \phi_i(p(t)) \quad i = 1, \dots, n_g. \quad (4)$$

where $\rho_i \in \mathbb{R}^{n_\mathrm{H}}$ is an unknown vector of parameters and $\phi_i : \mathbb{P} \to \mathbb{R}^{n_\mathrm{H}}$ (with $i = 1, \dots, n_a + n_b + 1$) maps the observed scheduling variable $p(t)$ to an n_H-dimensional space, commonly referred to as the *feature space*. Unlike the LPV parametric identification approaches, neither the maps ϕ_i nor the dimension n_H of the vectors $\rho_i^\top$ and $\phi_i(p(t))$ are explicitly specified by the user. Potentially, $\rho_i^\top$ 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 (3) is rewritten in the linear regression form:

$$y(t) = \sum_{i=1}^{n_g} \rho_i^\top \phi_i(t)x_i(t) + e(t), \quad (5)$$

where $\phi_i(t)$ is used as a shorthand notation for $\phi_i(p(t))$.

In the LS-SVM formulation, the following *quadratic programming* (QP) problem with a regularized ℓ_2 loss function is considered to estimate the LPV model (5) from the data observations $\mathcal{D}_N$:

$$\min_{\rho_i, e} \quad \frac{1}{2}\sum_{i=1}^{n_g} \rho_i^\top \rho_i + \frac{\lambda}{2}\sum_{t=1}^{N} e^2(t) \quad (6a)$$

$$\text{s.t.} \quad e(t) = y(t) - \sum_{i=1}^{n_g} \rho_i^\top \phi_i(t)x_i(t), \ \ t \in \mathbb{I}_1^N \quad (6b)$$

where $\mathbb{I}_1^N$ denotes the set of indexes $\{1, \dots, N\}$, and $\lambda > 0$ is a tuning hyper-parameter. The term $\sum_{t=1}^{N} e^2(t)$ in the

cost function (6a) aims at minimizing the prediction error, while the regularization term $\rho_i^\top \rho_i$ is added in (6a) to prevent overfitting. The hyper-parameter λ should be then tuned to balance the bias/variance tradeoff. Note that the parameters ρ_i minimizing (6) cannot be computed since this would require an explicit representation of the feature maps $\{\phi_i(t)\}_{i=1}^{n_g}$. Thus, the dual formulation of Problem (6) is considered. The Lagrangian $\mathcal{L}(\rho, e, \alpha)$ associated with the primal Problem (6) is given by:

$$\mathcal{L}(\rho, e, \alpha) = \frac{1}{2}\sum_{i=1}^{n_g} \rho_i^\top \rho_i + \frac{\lambda}{2}\sum_{t=1}^{N} e^2(t) - \sum_{t=1}^{N} \alpha_t \left(e(t) - y(t) + \sum_{i=1}^{n_g} \rho_i^\top \phi_i(t) x_i(t) \right), \quad (7)$$

where $\alpha = [\alpha_1 \cdots \alpha_N]^\top \in \mathbb{R}^N$ is the vector of Lagrange multipliers associated with the equality constraints (6b). The optimal solution of the primal QP problem (6) is then achieved when the following KKT conditions are satisfied:

$$\frac{\partial \mathcal{L}}{\partial \rho_i} = 0 \rightarrow \rho_i = \sum_{t=1}^{N} \alpha_t \phi_i(t) x_i(t), \quad (8a)$$

$$\frac{\partial \mathcal{L}}{\partial e(t)} = 0 \rightarrow e(t) = \frac{1}{\lambda}\alpha_t, \quad (8b)$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_t} = 0 \rightarrow e(t) = y(t) - \sum_{i=1}^{n_g} \rho_i^T \phi_i(t) x_i(t). \quad (8c)$$

Substituting equations (8a) and (8b) into (8c) leads to:

$$y(t) = \sum_{i=1}^{n_g} \left(\sum_{t=1}^{N} \alpha_t x_i(t) \phi_i^T(t) \right) \phi_i(t) x_i(t) + \frac{1}{\lambda}\alpha_t \quad t \in \mathbb{I}_1^N \quad (9)$$

Equations in (9) can be written in the matrix form

$$Y = \left(\Omega + \lambda^{-1} I_N\right) \alpha, \quad (10)$$

where, $Y = [y(1) \cdots y(N)]^\top$, I_N is the identify matrix of size N, and Ω is the Kernel Matrix whose (j,k)-th entry is given by: $[\Omega]_{j,k} = \sum_{i=1}^{n_g} [\Omega^i]_{j,k}$ with

$$[\Omega^i]_{j,k} = x_i(j)\phi_i^T(p(j))\phi_i(p(k))x_i(k) \quad (11)$$

$$= x_i(j) K_i(p(j), p(k)) x_i(k) \quad (12)$$

where, K_i is a positive definite *kernel function* defining the inner product $\phi_i^\top(p(j))\phi_i(p(k))$. Specification of the kernel instead of the feature maps ϕ is called *kernel trick* [22] and it obviates the need to specify the feature maps explicitly, thus allowing the identification of the coefficient functions c_i by only specifying the kernel functions. A typical choice of kernel, which provides uniformly effective representation of a large class of smooth functions, is the *Radial Basis Function* (RBF) function:

$$K_i(p(j), p(k)) = \exp\left(-\frac{(p(j) - p(k))^2}{\sigma_i^2} \right), \quad i = 1, \ldots, n_g, \quad (13)$$

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

Once the kernel matrix Ω is defined, the Lagrange multipliers α are computed from (10), and the estimates of the coefficients c_i are obtained by substituting (8a) into (4):

$$\hat{c}_i(\cdot) = \rho_i^\top \phi_i(\cdot) = \sum_{t=1}^{N} \alpha_t K_i(p(t), \cdot) x_i(t), i \in \mathbb{I}_1^{n_g} \quad (14)$$

IV. Model order selection

In this section, we present an extension of the LS-SVM method, which aims at estimating a sparse LPV model structure, thus minimizing the number of non-zero p-dependent coefficients $\{c_i\}_{i=1}^{n_g}$. The key idea of the proposed method is to start with an over-parameterized LPV model, whose coefficient functions $\{c_i\}_{i=1}^{n_g}$ are estimated through the LS-SVM approach presented in Section III. Then, the estimated coefficients $\{\hat{c}_i\}_{i=1}^{n_g}$ are reshaped, multiplying them by polynomial weights depending on the scheduling variable p. A regularization term aiming at shrinking the polynomial weights towards zero is considered in order to enforce sparsity in the final estimated model.

Let us introduce the polynomial weights

$$w_i(p(t)) = \mathbf{w}_i^\top \varphi(p(t)), \quad i = 1, \ldots, n_g, \quad (15)$$

where $\varphi(p(t))$ is a vector of monomials in the variable $p(t)$ and $\mathbf{w}_i \in \mathbb{R}^{n_\mathrm{w}}$ is the (unknown) vector of parameters. The monomials in the vector $\varphi(p(t))$ are specified a-priori.

Given the estimate of the coefficients obtained from LS-SVM (see eq. (14)) and their corresponding values at the training points $\{p(t)\}_{t=1}^{N}$, i.e.,

$$\hat{c}_i(p(t)) = \sum_{k=1}^{N} \alpha_k K_i(p(k), p(t)) x_i(k), \quad i \in \mathbb{I}_1^{n_g} \quad (16)$$

we scale $\hat{c}_i(p(t))$ with the polynomial weights $w_i(p(t))$, i.e.,

$$\tilde{c}_i(p(t)) = w_i(p(t))\hat{c}_i(p(t)) = \mathbf{w}_i^\top \varphi(p(t))\hat{c}_i(p(t)), \quad i \in \mathbb{I}_1^{n_g},$$

where $\tilde{c}_i$ are the scaled LPV coefficients.

With the scaled coefficients, the considered LPV-ARX model becomes:

$$y(t) = \sum_{i=1}^{n_g} \mathbf{w}_i^\top \varphi(p(t))\hat{c}_i(p(t)) x_i(t) + e(t). \quad (17)$$

Now, to enforce the sparsity in the estimate of the (scaled) coefficients $\tilde{c}_i(p(t))$, a group LASSO term penalizing the l_∞-norm of the polynomial coefficient vectors $\mathbf{w}_i$ is minimized together with the residual error, leading to the following convex optimization problem:

$$\min_{\{\mathbf{w}_i\}_{i=1}^{n_g}} \sum_{t=1}^{N} \left(y(t) - \sum_{i=1}^{n_g} \mathbf{w}_i^\top \varphi(p(t))\hat{c}_i(p(t)) x_i(t) \right)^2 + \mu \sum_{i=1}^{n_g} \|\mathbf{w}_i\|_\infty \quad (18)$$

Note that the group LASSO term in (18) penalizes the mixed $\ell_{1,\infty}$-norm (i.e., sum of the infinity norms) of the parameter vectors $\mathbf{w}_i$, with $i = 1, \ldots, n_g$. The infinity norm is considered as a norm of the group so that, at the

solution, the vector $\mathbf{w}_i$ is enforced to be either identically to zero or full. Indeed, only the component of the vector $\mathbf{w}_i$ with largest absolute value affects the objective function in (18). Note that, when $\mathbf{w}_i$ is zero, the polynomial weight $w_i(\cdot) = \mathbf{w}_i^\top \varphi(\cdot)$ is null. Thus, the corresponding scaled coefficient function $\tilde{c}_i$ is also null. The hyper-parameter $\mu \geq 0$ is tuned by the user to balance the tradeoff between minimizing the fitting error and minimizing the number of non-zero functions $\{\tilde{c}_i\}_{i=1}^{n_g}$ defining the LPV-ARX model with scaled coefficients reported in (17).

Summarizing, the shape of the p-dependent LPV model coefficients $\{c_i\}_{i=1}^{n_g}$ is initially obtained through a standard LS-SVM dedicated to LPV identification. Then, polynomial weights $\{w_i(p(t))\}_{i=1}^{n_g}$ are used to reshape the estimated model coefficients $\{\hat{c}_i\}_{i=1}^{n_g}$ and, at the same time, to shrink the LPV model coefficients towards zero, thus reducing the complexity of the estimated LPV model (by minimizing the number of nonzero model coefficients $\{\tilde{c}_i\}_{i=1}^{n_g}$).

As the final estimate of the scaled coefficients $\{\tilde{c}_i\}_{i=1}^{n_g}$ will be biased because of the regularization term $\sum_{i=1}^{n_g} \|\mathbf{w}_i\|_\infty$ (see eq. (18)), an LPV model with reduced complexity, containing only the coefficients $\{\tilde{c}_i\}_{i=1}^{n_g}$ which have been detected to be nonzero, should be re-identified by using the non-regularized LS-SVM approach discussed in Section III.

V. Simulation examples

This section shows the effectiveness of the proposed method on two simulation examples. To study the statistical properties of the estimation, Monte-Carlo simulations of 100 runs are performed for each example. At each Monte-Carlo run, a new data set of inputs, scheduling variables and noises is generated. The output used in the training phase is corrupted by an additive zero-mean white noise e^{o} with Gaussian distribution. The effect of the noise e^{o} on the output signal is quantified through the Signal-to-Noise Ratio (SNR), defined as

$$\mathrm{SNR} = 10 \log \frac{\sum_{t=1}^{N} \left(y(t) - e^{\mathrm{o}}(t)\right)^2}{\sum_{t=1}^{N} \left(e^{\mathrm{o}}(t)\right)^2}, \tag{19}$$

Radial basis functions are used as kernels to define the inner product among the feature maps ϕ_i. The values of the hyper-parameters λ, μ (eq. (18)) and σ_i (characterizing the RBF K_i in (13)) are chosen through a cross-calibration procedure, that is by maximizing (with a grid search) the *Best Fit Rate* (BFR) w.r.t. a calibration data set of length N_{C}. The BFR is defined as

$$\mathrm{BFR} = \max \left\{ 1 - \sqrt{\frac{\sum_{t=1}^{N_{\mathrm{C}}} \left(y(t) - \hat{y}(t)\right)^2}{\sum_{t=1}^{N_{\mathrm{C}}} \left(y(t) - \bar{y}\right)^2}}, 0 \right\} \tag{20}$$

with $\hat{y}(t)$ being the simulated model output and $\bar{y}$ being the sample mean of the output over the calibration set. In order to speed up the calibration procedure, the parameters σ_i are equal among each others (i.e., $\sigma_i = \sigma$ for all $i = 1, \ldots, n_g$). The results obtained after the training and calibration phase are validated on a noiseless data sequence. The BFR is used to assess the quality of the estimated models.

All computations are carried out on a i5 1.7GHz Intel core processor with 4 GB of RAM running MATLAB R2013a. The CVX package [7] is used to solve Problem (18).

A. Example 1

The aim of the academic example reported in this subsection is threefold: 1. showing the capabilities of the presented regularization approach in detecting the correct structure of the underlying LPV data-generating systems; 2. showing that the variance of estimate obtained by non-regularized LS-SVM can be reduced by first detecting the model structure; 3. comparing the performance, in terms of computational time, of the presented approach w.r.t. to the regularization approach proposed in [13]. Consider the LPV data-generating system taken from [13]:

$$y(t) = a_1^{\mathrm{o}}(p(t))y(t-1) + a_2^{\mathrm{o}}(p(t))y(t-2) + b_5^{\mathrm{o}}(p(t))u(t-5) + e^{\mathrm{o}}(t). \tag{21}$$

The p-dependent coefficients $a_1^{\mathrm{o}}(p(t))$, $a_2^{\mathrm{o}}(p(t)$ and $b_5^{\mathrm{o}}(p(t))$ are described by the nonlinear functions:

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

$$a_2^{\mathrm{o}}(p(t)) = \sin(2\pi p(t)); \;\; b_5^{\mathrm{o}}(p(t)) = p^3(t)$$

The system is estimated from a training data set $\mathcal{D}_N$ of length $N = 500$, while a calibration data set of length $N_{\mathrm{C}} = 200$ is used to tune the hyper-parameters λ, μ and σ. To gather data, the input u and the scheduling parameter p are chosen to be white-noise processes independent of each other with uniform distribution $\mathcal{U}(-1, 1)$. The standard deviation of the noise e^{o} is 0.3. The average of the SNR over the over 100 Monte-Carlo runs is 7 dB. The identification problem is formulated in the LS-SVM setting by using over-parameterized LPV model structure (2) with $n_a = n_b = 10$.

First, the estimates of the coefficients $\{a_i\}_{i=1}^{n_a}$ and $\{b_j\}_{j=0}^{n_b}$ are obtained using the non-regularized LS-SVM approach described in Section III. The chosen hyper-parameters, maximizing the BFR w.r.t. the calibration data set, are: $\lambda = 600$ and $\sigma = 0.4$. Then, second-order polynomials $\{w_i(p(t))\}_{i=1}^{n_g}$ are used as weights to re-shape the estimated coefficients $\{\hat{a}_i\}_{i=1}^{n_a}$ and $\{\hat{b}_j\}_{j=0}^{n_b}$, and thus to detect the LPV model structure by solving Problem (18). The chosen hyper-parameter μ, maximizing the BFR w.r.t. the calibration data set, is $\mu = 5$.

Table I and II show the maximum absolute values (over the training points $\{p(t)\}_{t=1}^{N}$) of the coefficient functions a_i and b_j obtained from the non-regularized LS-SVM approach, along with ones computed through the regularized LS-SVM version (denoted as R-LS-SVM) proposed in this paper. The obtained results show that the proposed regularized LS-SVM approach is able to detect the correct underlying structure of the LPV data-generating system. Indeed, the only coefficients a_i and b_j which have been detected to have an (average) maximum absolute value larger than a threshold of 10^{-6} are:

a_1, a_2 and b_5. Results in Table I and II also show that, as expected, the estimate of the nonzero coefficients obtained by the regularized LS-SVM is biased, because of the regularization penalty $\sum_{i=1}^{n_g} \|\mathrm{w}_i\|_\infty$ in eq. (18) (note that this a well-known problem affecting also parametric regularization methods like the LASSO). Therefore, the nonzero coefficient functions are re-estimated without the regularization term. Precisely, the coefficient functions which are detected to be null are discarded in the description of the LPV model (2) and a lower-complexity LPV model is re-identified through non-regularized LPV LS-SVM approach. The estimates of the nonzero coefficient a_1 is plotted in Fig. 1, which shows the mean estimate, along with the standard deviation intervals computed over the 100 Monte-Carlo runs. The estimate of the same coefficients obtained with the non-regularized LPV LS-SVM approach is also plotted in the same figure. The obtained results show that detecting the LPV model structure is beneficial, in terms of variance reduction, in the final estimate of the coefficient functions.

For the sake of comparison, the regularization method of [13] is also run, by gridding the scheduling variable space into 20 equidistant points. In terms of model quality, the method in [13] is also able to detect the true LPV model structure. However, the benefits of the proposed method w.r.t. [13] can be appreciated in terms of computational time. For fixed values of the hyper-parameters λ, μ and σ, the average CPU time (over the 100 Monte-Carlo runs) required by the proposed identification algorithm is 2 s. On the other hand, the average overall time required by the method of [13] to solve the same identification problem is 24 s (12x slower than the method proposed in this paper).

B. Example 2. Multidimensional scheduling variable

The aim of the academic example considered in this section is to show the effectiveness of the proposed regularization scheme in the identification of LPV systems with multidimensional scheduling signals. This represents one of the main advantage of the proposed method over [13]. In fact, the method in [13] requires to grid the scheduling space, thus limiting its applicability to the identification of LPV systems with one/two-dimensional scheduling variables. The considered LPV data-generating system is described by the difference equation:

$$\begin{aligned} y(t) = a_1^{\mathrm{o}}(p(t))y(t-1) + a_2^{\mathrm{o}}(p(t))y(t-2) + \\ + b_4^{\mathrm{o}}(p(t))u(t-4) + b_5^{\mathrm{o}}(p(t))u(t-5) + e^{\mathrm{o}}(t), \end{aligned} \tag{22}$$

where, $p(t) = [p_1(t)\ p_2(t)\ p_3(t)]^\top \in \mathbb{R}^3$. The unknown functions are: $a_1^{\mathrm{o}}(p(t)) = 0.3p_1^2(t) + 0.2p_2^2(t) - 0.1p_3^2(t)$, $a_2^{\mathrm{o}}(p(t)) = 0.2p_1(t) - 0.3p_2(t) + 0.1p_3(t)$, $b_4^{\mathrm{o}}(p(t)) = 0.2\sin(2\pi p_1(t)) + \sin(2\pi p_2(t))$, $b_5^{\mathrm{o}}(p(t)) = 0.4\cos(2\pi p_2(t)) + 0.3\sin(2\pi p_3(t))$. Training data set $\mathcal{D}_N$ of length $N = 3000$ is used for estimation, while a calibration data set of length $N_{\mathrm{C}} = 1000$ is used to tune the hyper-parameters λ, μ and σ. To generate the data set, the input u and the scheduling signals p_1, p_2 and p_3 are chosen to be white-noise processes independent of each other with uniform distribution $\mathcal{U}(-1, 1)$. The standard deviation of the noise e^{o} is 0.08. The average of the SNR over the over 100 Monte-Carlo runs is 15 dB.

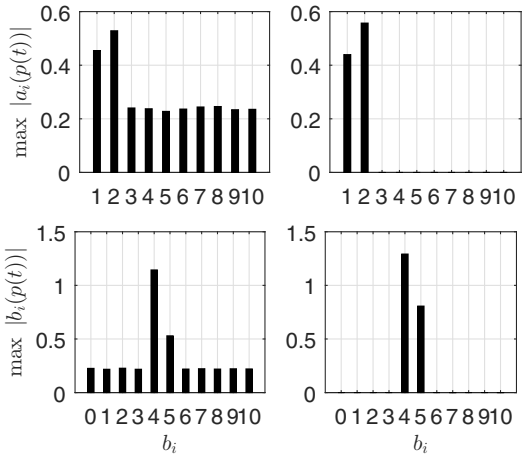

Fig. 2. Example 2. Average (over 100 Monte-Carlo runs) of the maximum absolute value of the LPV model coefficients $\{a_i\}_{i=1}^{10}$ and $\{b_j\}_{j=0}^{10}$ obtained through non-regularized LS-SVM (left panels) and through regularized LS-SVM (right panels).

First, the estimates of the coefficients are obtained using the non-regularized LS-SVM approach described in Section III with hyper-parameters, $\lambda = 900$ and $\sigma = 0.8$. Then, second-order polynomials $\{w_i(p(t))\}_{i=1}^{n_g}$ are used as weights to re-shape the estimated coefficients $\{\hat{a}_i\}_{i=1}^{n_a}$ and $\{\hat{b}_j\}_{j=0}^{n_b}$, and thus to detect the LPV model structure by solving Problem (18). The chosen hyper-parameter μ, is $\mu = 25$. The total computational time (including calibration phase) required to solve the estimation problem is 302 s. For fixed values of λ, σ and μ, the average computational time is 15.1 s (i.e., 20 different combinations of the hyper-parameters λ, σ and μ have been tested in calibration). Fig. 2 depicts the average maximum absolute values of the estimated coefficient functions over 100 Monte-Carlo runs. Results in Fig. 2 show that the proposed regularized LS-SVM detects the true structure of the underlying dynamics, while a non-sparse LPV model is obtained by using standard LS-SVM. This has an impact on the generalization properties of the model to unseen data. In fact, the BFR (in validation) obtained using standard LS-SVM is 0.62, while a BFR equal to 0.88 is achieved by regularized LS-SVM.

VI. Conclusions

This paper has proposed a regularized LS-SVM method for sparse identification of LPV-ARX models. The dependence of the LPV model coefficients on the scheduling variable is not a-priori parameterized, and it is estimated, at the first stage, by using a standard (i.e., non-regularized) LS-SVM based regression approach. The obtained coefficients are then re-shaped by polynomial weights, and a penalty term is minimized in order to shrink the scaled model coefficients towards zero, thus enforcing a sparse structure in the estimated LPV model. The proposed method exploits the flexibility of the LS-SVM to reconstruct the underlying dependence of the LPV model coefficients on the scheduling signal, while the parametric structure of the scaling weights

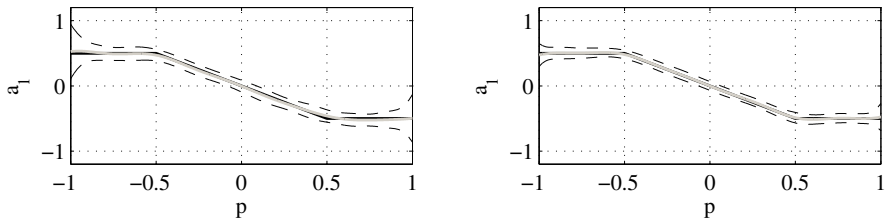

Fig. 1. Example 1. Estimate of the non-zero coefficients obtained via non-regularized LS-SVM (left panels) and after model order selection (right panels). True function (solid black line), mean estimate (solid gray line) and standard deviation intervals (dashed black line) over the 100 Monte Carlo runs.

TABLE I

EXAMPLE 1. AVERAGE AND STANDARD DEVIATION (OVER 100 MONTE-CARLO RUNS) OF THE MAXIMUM ABSOLUTE VALUE OF THE ESTIMATED LPV MODEL COEFFICIENTS $a_i(p(t))$

Coefficients	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
True Value	0.5	1	0	0	0	0	0	0	0	0
Mean (LS-SVM)	0.6245	1.0498	0.1853	0.1788	0.1858	0.1939	0.1987	0.1891	0.1935	0.1904
Mean (R-LS-SVM)	0.4766	0.9667	2.61e-11	2.25e-11	2.63e-11	3.29e-11	2.47e-11	2.24e-11	2.31e-11	3.02e-11
std (LS-SVM)	8.54e-02	3.56e-02	9.80e-02	9.12e-02	8.61e-02	0.1052	0.1033	8.16e-02	9.06e-02	9.45e-02
std (R-LS-SVM)	6.63e-02	3.81e-02	5.51e-11	3.48e-11	4.01e-11	6.14e-11	3.78e-11	3.17e-11	3.43e-11	5.96e-11

TABLE II

EXAMPLE 1. AVERAGE AND STANDARD DEVIATION (OVER 100 MONTE-CARLO RUNS) OF THE MAXIMUM ABSOLUTE VALUE OF THE ESTIMATED LPV MODEL COEFFICIENTS $b_j(p(t))$

Coefficients	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
True Value	0	0	0	0	0	1	0	0	0	0	0
Mean (LS-SVM)	0.1236	0.1372	0.1292	0.1204	0.1213	1.0133	0.1270	0.1350	0.1333	0.1201	0.1329
Mean (R-LS-SVM)	1.79e-11	1.68e-11	1.78e-11	1.73e-11	1.56e-11	0.8748	1.85e-11	1.99e-11	1.96e-11	1.74e-11	1.71e-11
std (LS-SVM)	6.30e-02	7.10e-02	5.80e-02	5.64e-02	6.32e-2	0.1009	5.62e-02	5.66e-02	5.92e-02	5.66e-02	5.63e-02
std (R-LS-SVM)	3.46e-11	2.66e-11	2.63e-11	3.11e-11	2.56e-11	7.62e-03	2.65e-11	2.71e-11	3.12e-11	2.69e-11	2.64e-11

allows to select the dynamical structure of the model through a group-LASSO based approach. The reported simulation results show the capabilities of the proposed approach and its advantages with respect to an other regularized LS-SVM approach available in the literature.

REFERENCES

[1] A. A. Bachnas, R. Tóth, J. Ludlage, and A. Mesbah. 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, 2014.

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

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

[4] V. Cerone, D. Piga, and D. Regruto. Set-membership LPV model identification of vehicle lateral dynamics. *Automatica*, 47(8):1794–1799, 2011.

[5] R. Duijkers, R. Tóth, D. Piga, and V. Laurain. Shrinking complexity of scheduling dependencies in LS-SVM based LPV system identification. In *53rd IEEE Conference on Decision and Control*, pages 2561–2566, Los Angeles, California, USA, 2014. IEEE.

[6] 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.

[7] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. url: `http://cvxr.com/cvx`, 2014.

[8] K. Hsu, T. L. Vincent, and K. Poolla. Nonparametric methods for the identification of linear parameter varying systems. In *Proc. of the Int. Symposium on Computer-Aided Control System Design*, pages 846–851, San Antonio, Texas, USA, Sept. 2008.

[9] 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.

[10] A. Marcos and G.J. Balas. Development of linear-parameter-varying models for aircraft. *Journal of Guidance Control and Dynamics*, 27(2):218–228, 2004.

[11] C. Novara, F. Ruiz, and M. Milanese. Direct identification of optimal SM-LPV filters and application to vehicle yaw rate estimation. *IEEE Transactions on Control Systems Technology*, 19(1):5–17, 2011.

[12] D. Piga, P. Cox, R. Tóth, and V. Laurain. LPV system identification under noise corrupted scheduling and output signal observations. *Automatica*, 53:329–338, 2015.

[13] D. Piga and R. Tóth. LPV model order selection in an LS-SVM setting. In *52nd IEEE Conference on Decision and Control*, pages 4128–4133, Florence, Italy, 2013.

[14] 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.

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

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

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

[22] V. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, 1998.

[23] V. Verdult and M. Verhaegen. Kernel methods for subspace identification of multivariable LPV and bilinear systems. *Automatica*, 41:1557–1565, 2005.