2015 IEEE 54th Annual Conference on Decision and Control (CDC)
December 15-18, 2015. Osaka, Japan

An IV-SVM-based Approach for Identification of State-Space LPV Models under Generic Noise Conditions

Syed Z. Rizvi†, Javad Mohammadpour†, Roland Tóth‡, and Nader Meskin§

Abstract— **This paper presents a nonparametric identification method for state-space linear parameter-varying (LPV) models using a modified support vector machine (SVM) approach. While most LPV identification schemes in the state-space form fall under the general category of parametric methods, regularization-based SVMs provide a viable alternative to model scheduling dependencies, without the need of specifying the dependency structure and with an attractive bias-variance trade-off. In this paper, a solution is proposed for nonparametric identification of LPV state-space models in terms of least-squares SVMs (LS-SVM) and is then extended in a way that the proposed estimation is robust to errors in the noise model estimation. The so-called instrumental variables (IV) method has been used in linear system identification for quite some time, and has recently seen its application in the identification of both nonlinear and LPV systems in the input-output (IO) form. The IV method reduces the bias in estimated LPV state-space models in case the noise model is not estimated properly or is unknown. In the proposed method of this paper, the attractive bias-variance trade-off properties of LS-SVMs are combined with statistical properties of IV-based methods to give robust estimates of the functional dependencies. Numerical examples are provided to compare the performances of the proposed IV-based technique with the LS-SVM-based LPV model identification methods.**

I. Introduction

Linear parameter-varying (LPV) models offer an efficient framework for modeling nonlinear systems by representing the nonlinear model as a linear dynamic relation of the input and output variables where the relations themselves are dependent on the measurable time-varying signals, commonly known as the *scheduling variables*. This way, the scheduling signals take into account the varying operating conditions of the system. This simplicity of LPV models allows for the application of several linear control techniques to nonlinear systems represented by LPV models and opens the door for the application of powerful LPV control synthesis tools. Naturally then, LPV identification has attracted a lot of attention in the past decade [1], with different identification schemes developed for both input-output (IO) and LPV state-space models [2]–[5].

Most identification methods in the literature for state-space LPV models fall under the category of parametric approach, where the scheduling dependencies of the state space model matrices are assumed to be parameterized with respect to *known basis functions* [6]. This leads to over-parametrization of the model coefficients, causing a large variance in the estimates. On the other hand, an inappropriate selection of these functions is known to cause a structural bias [3]. For LPV state-space and bilinear models, parametric methods are mostly based on various subspace approaches. These methods usually require a high computational demand due to the enormous dimensions of the data matrices involved. The authors in [7] proposed a tractable way to reduce this curse of dimensionality for LPV state-space models with affine parameter dependence. A few more subspace-based methods were later published in [2], [8] among others.

Nonparametric methods provide an alternative that can avoid the bias-variance trade-off by obtaining nonparametric reconstruction of the scheduling dependencies in LPV models. With the emergence of kernel-based techniques, a new avenue of nonparametric identification, classification and data processing has appeared in the last two decades. Kernels are functions that enable us to perform linear operations in high-dimensional feature spaces, often mapping nonlinear dependencies efficiently using the so-called *kernel trick* [9]. This has sprouted the use of kernel-based techniques for solving various problems under the umbrella of LPV system identification [3], [10]. The identification approaches in [3], [11]–[13] reported efficient kernel-based methods employing *least-squares support vector machines* (LS-SVM) for LPV-IO and state-space models; the results showed consistent estimates with an attractive bias-variance trade-off. An iterative mixed parametric method for LPV state-space identification was proposed recently in [14], in which the authors described the C matrix using a nonparametric LS-SVM model, while the A matrix was described by a parametric model.

While regularization-based SVMs provide an attractive bias-variance trade-off for efficient nonparametric identification, these methods suffer from severe restrictions imposed on noise, where the obtained estimates would be truly unbiased *w.r.t.* the noise only when writing the estimation problem into a regression form and the resulting noise

*This work was made possible by the NPRP grant # NPRP 5-574-2-233 from the Qatar National Research Fund (QNRF), a member of Qatar Foundation. The statements made here are solely the responsibility of the authors.

†Syed Z. Rizvi and Javad Mohammadpour are with the Complex Systems Control Laboratory (CSCL), College of Engineering, The University of Georgia, Athens, GA 30602, USA. Corresponding author email: javadm@uga.edu

‡Roland Tóth is with the Control Systems Group, Dept. of Electrical Engineering, Eindhoven University of Technology, P.O.Box 513, 5600 MB Eindhoven, The Netherlands. Email: r.toth@tue.nl

§Nader Meskin is with the Dept. of Electrical Engineering, Qatar University, Doha, Qatar. Email: nader.meskin@qu.edu.qa

process is white [11]. In order to obtain unbiased estimates for more generic noise conditions, *instrumental variables* (IV) have been employed in LTI identification theory in [15], and recently in the context of nonparametric identification for nonlinear ARX models in [11] and LPV IO models in [11].

In this paper, we present an LS-SVM-based nonparametric identification method for multi-input multi-output (MIMO) state-space LPV models. We further derive an IV-based SVM optimization problem to model the coefficient dependencies on the scheduling variables in the presence of colored noise, often correlated with the scheduling variables. The paper is arranged as follows. The problem is formulated in Section II. An LS-SVM identification algorithm is formulated in Section III. Section IV discusses the conditions on the instruments and on the LPV structure of the noise. Section V formulates the proposed IV-based algorithm. Performance of the proposed method is demonstrated on a numerical example in Section VI. Concluding remarks are made in Section VII.

II. Problem Formulation

Consider an LPV system represented by the following discrete-time state-space model

$$\begin{aligned} x(k+1) &= A(p_k)x(k) + B(p_k)u(k) + v(k), \\ y(k) &= C(p_k)x(k) + z(k), \end{aligned} \tag{1}$$

where $k \in \mathbb{Z}$ denotes the discrete time instant, and matrices $A(p_k) \in \mathbb{R}^{n\times n}$, $B(p_k) \in \mathbb{R}^{n\times n_u}$, and $C(p_k) \in \mathbb{R}^{n_y\times n}$ are functions of time-varying scheduling variables $p(k) \in \mathbb{R}^{n_p}$, denoted as p_k for better readability. Variables $v(k) \in \mathbb{R}^n, z(k) \in \mathbb{R}^{n_y}$ are zero-mean, quasi-stationary stochastic noise processes, not necessarily white, but independent from $u(k)$. We aim at employing nonlinear kernel functions under the LS-SVM framework in order to estimate the functional dependencies of the state-space matrices on the scheduling variables. We can formulate the problem by writing the state-space LPV model (1) as follows

$$\begin{aligned} x(k+1) &= W_x\varphi_x^\top(k) + \varepsilon_v(k), \\ y(k) &= W_y\varphi_y^\top(k) + \varepsilon_z(k) \end{aligned} \tag{2}$$

where $\varepsilon_v(k), \varepsilon_z(k)$ are residual errors on the states and outputs, $W_x = [W_1 \; W_2] \in \mathbb{R}^{n\times 2n_H}$ and $W_y = W_3 \in \mathbb{R}^{n_y\times n_H}$ are weighting matrices, and $\varphi_x^\top(k) \in \mathbb{R}^{2n_H\times 1}$ and $\varphi_y^\top(k) \in \mathbb{R}^{n_H\times 1}$ are unknown regressors given by

$$\begin{aligned} \varphi_x^\top(k) &= [(\Phi_1(p_k)x(k))^\top \; (\Phi_2(p_k)u(k))^\top]^\top, \\ \varphi_y^\top(k) &= \Phi_3(p_k)x(k). \end{aligned} \tag{3}$$

In addition, n_H represents the dimension of a possibly infinite-dimensional feature space. From (2)-(3), we can gauge that $A(p_k) = W_1\Phi_1(p_k), B(p_k) = W_2\Phi_2(p_k)$, and $C(p_k) = W_3\Phi_3(p_k)$. We assume that the states are available for measurement, a possibility often encountered in applications like chemical processes. Estimating dependencies without state measurements is the subject of our ongoing research and is not covered in this paper. The problem, therefore, reduces to finding the dependency of $W_x\varphi_x^\top(k)$ and $W_y\varphi_y^\top(k)$ on p_k given the data $\{u(k), x(k), y(k), p(k)\}_{k=1}^N$, where N is the number of samples.

III. An LS-SVM Approach for LPV-SS Identification

The estimates of the LPV state-space matrix functions can be obtained by minimizing the following cost function

$$\begin{aligned} \mathcal{J}(W_x, W_y, \varepsilon_v, \varepsilon_z) &= \frac{1}{2}\left(\|W_x\|_\mathrm{F}^2 + \|W_y\|_\mathrm{F}^2\right) \\ &+ \frac{1}{2}\left(\sum_{k=1}^N \varepsilon_v^\top(k)\Gamma\varepsilon_v(k) + \sum_{k=1}^N \varepsilon_z^\top(k)\Psi\varepsilon_z(k)\right), \end{aligned} \tag{4}$$

where $\|\cdot\|_\mathrm{F}$ denotes the Frobenius norm, and $\Gamma = \mathrm{diag}(\gamma_1, \cdots, \gamma_n)$ and $\Psi = \mathrm{diag}(\psi_1, \cdots, \psi_{n_y})$ are diagonal weighting matrices on the residual errors $\varepsilon_v(k)$ and $\varepsilon_z(k)$ in (2); these weighting matrices are known as the regularization parameters. The above optimization can be solved by introducing *Lagrangian multipliers* and substituting the inner product $\Phi_i^\top\Phi_i$ using an *a priori* chosen nonlinear kernel function as shown in [3]. We define the Lagrangian as

$$\begin{aligned} \mathcal{L}(W_x, W_y, \boldsymbol{\alpha}, \boldsymbol{\beta}, \varepsilon_v, \varepsilon_z) &= \mathcal{J}(W_x, W_y, \varepsilon_v, \varepsilon_z) \\ &- \sum_{j=1}^N \alpha_j^\top\left\{W_x\varphi_x^\top(j) + \varepsilon_v(j) - x(j+1)\right\} \\ &- \sum_{j=1}^N \beta_j^\top\left\{W_y\varphi_y^\top(j) + \varepsilon_z(j) - y(j)\right\}, \end{aligned} \tag{5}$$

where $\alpha_j \in \mathbb{R}^n$, $\beta_j \in \mathbb{R}^{n_y}$ are the Lagrange multipliers at the discrete time j. In order to solve for the global optimum, saddle points are obtained by solving the *Karush-Kuhn-Tucker* (KKT) conditions as follows

$$\frac{\partial\mathcal{L}}{\partial W_x} = 0 \Rightarrow W_x = \sum_{j=1}^N \alpha_j\varphi_x(j), \tag{6a}$$

$$\frac{\partial\mathcal{L}}{\partial W_y} = 0 \Rightarrow W_y = \sum_{j=1}^N \beta_j\varphi_y(j), \tag{6b}$$

$$\frac{\partial\mathcal{L}}{\partial \alpha_j} = 0 \Rightarrow \varepsilon_v(j) = x(j+1) - W_x\varphi_x^\top(j), \tag{6c}$$

$$\frac{\partial\mathcal{L}}{\partial \beta_j} = 0 \Rightarrow \varepsilon_z(j) = y(j) - W_y\varphi_y^\top(j), \tag{6d}$$

$$\frac{\partial\mathcal{L}}{\partial \varepsilon_v(j)} = 0 \Rightarrow \alpha_j = \Gamma\varepsilon_v(j), \tag{6e}$$

$$\frac{\partial\mathcal{L}}{\partial \varepsilon_z(j)} = 0 \Rightarrow \beta_j = \Psi\varepsilon_z(j). \tag{6f}$$

Substituting the relations (6a)-(6f) into (2), we can eliminate the primal decision variables and write

$$\begin{aligned} x(k+1) &= W_x\varphi_x^\top(k) + \varepsilon_v(k) \\ &= \underbrace{\left\{\sum_{j=1}^N \alpha_j\varphi_x(j)\right\}}_{W_x}\varphi_x^\top(k) + \underbrace{\Gamma^{-1}\alpha_k}_{\varepsilon_v(k)}, \end{aligned} \tag{7}$$

$$y(k) = W_y \varphi_y^\top(k) + \varepsilon_z(k) = \underbrace{\left\{\sum_{j=1}^{N} \beta_j \varphi_y(j)\right\}}_{W_y} \varphi_y^\top(k) + \underbrace{\Psi^{-1}\beta_k}_{\varepsilon_z(k)}. \quad (8)$$

Replacing the inner product $\Phi_i^\top(p_j)\Phi_i(p_k)$ by a kernel function $\bar{k}^i(p_j, p_k)$, we define kernel matrices Ω and Ξ as

$$[\Omega]_{j,k} = \varphi_x(j)\varphi_x^\top(k) = \sum_{i=1}^{2} z_i^\top(j)\bar{k}^i(p_j, p_k)z_i(k), \quad (9)$$

$$[\Xi]_{j,k} = \varphi_y(j)\varphi_y^\top(k) = x^\top(j)\bar{k}^3(p_j, p_k)x(k), \quad (10)$$

where $z_i(k) = x(k)$ and $z_i(k) = u(k)$ for $i = 1, 2$, respectively; the function $\bar{k}^i(\cdot,\cdot)$ denotes a nonlinear kernel function. While a wide variety of kernel functions exist in the literature to choose from, commonly used kernels include the *Radial Basis Function* (RBF), polynomial and sigmoid kernels among many others [9]. A typical RBF kernel, also known as the Gaussian kernel, is represented by

$$\bar{k}^i(p_j, p_k) = \exp\left(-\frac{\|p_j - p_k\|_2^2}{2\sigma_i^2}\right), \quad (11)$$

where σ_i is a free kernel parameter and $\|\cdot\|_2$ represents the Euclidean norm. By using (9) and (10), we can write (7)-(8) in a more compact form as

$$X = \boldsymbol{\alpha}\Omega + \Gamma^{-1}\boldsymbol{\alpha},$$
$$Y = \boldsymbol{\beta}\Xi + \Psi^{-1}\boldsymbol{\beta},$$

where $\Omega \in \mathbb{R}^{N\times N}$ and $\Xi \in \mathbb{R}^{N\times N}$ are the kernel matrices, $\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_N] \in \mathbb{R}^{n\times N}$ and $\boldsymbol{\beta} = [\beta_1 \cdots \beta_N] \in \mathbb{R}^{n_y\times N}$ are Lagrange multipliers, and $X = [x(1) \cdots x(N)] \in \mathbb{R}^{n\times N}$ and $Y = [y(1) \cdots y(N)] \in \mathbb{R}^{n_y\times N}$ contain the states and outputs for the N samples. The solution to the above equations can be obtained as follows

$$\mathrm{vec}(\boldsymbol{\alpha}) = \left(I_N \otimes \Gamma^{-1} + \Omega^\top \otimes I_n\right)^{-1}\mathrm{vec}(X), \quad (12)$$

$$\mathrm{vec}(\boldsymbol{\beta}) = \left(I_N \otimes \Psi^{-1} + \Xi^\top \otimes I_{n_y}\right)^{-1}\mathrm{vec}(Y), \quad (13)$$

where $\otimes$ denotes Kronecker product and $\mathrm{vec}(\cdot)$ denotes vectorization function, which stacks subsequent columns in a matrix below one another; matrix I_N represents identity matrix of dimension N. The solutions (12)-(13) are obtained using the solution to the classical Sylvester equation. Once estimated, the estimate of the state-space matrices can be calculated by using (6a)-(6b) as

$$\hat{A}(\cdot) = W_1\Phi_1(\cdot) = \sum_{k=1}^{N} \alpha_k x^\top(k)\bar{k}^1(p_k, \cdot), \quad (14a)$$

$$\hat{B}(\cdot) = W_2\Phi_2(\cdot) = \sum_{k=1}^{N} \alpha_k u^\top(k)\bar{k}^2(p_k, \cdot), \quad (14b)$$

$$\hat{C}(\cdot) = W_3\Phi_3(\cdot) = \sum_{k=1}^{N} \beta_k x^\top(k)\bar{k}^3(p_k, \cdot), \quad (14c)$$

giving us a nonparametric estimate of the state-space matrices. It is noteworthy that the parameter matrices W_i or the basis functions $\Phi_i(\cdot)$ are not accessible explicitly. What we estimate via nonlinear kernel functions is $W_i\Phi_i(\cdot)$.

Remark 1: It is noteworthy here that since the objective function (4) and constraints (2) constitute a convex primal problem with linear equality constraints, strong duality holds, and there exists no duality gap between the dual solution (12)-(13) and the solution to the primal problem.

IV. Noise Model Estimation and Instrumental Variables

Consider that there exists W_{x0}, W_{y0} such that the following holds true for the true data-generating LPV system (1): $x(k+1) = W_{x0}\varphi_x^\top(k) + v_0(k)$, $y(k) = W_{y0}\varphi_y^\top(k) + z_0(k)$. It was shown in [11] that LS-SVM-based optimization can provide consistent estimates under the assumption that W_{x0}, W_{y0} are bounded smooth functions and the condition that there is an absence of correlation between $\varphi_x(k)$ and $v_0(k)$, *e.g.*, when v_0 is a white noise process. The same holds true for correlation between $\varphi_y(k)$ and $z_0(k)$. This is not often the case in practice with sensor and actuator noise; noise is mostly colored and often correlated with the scheduling variables. To address the identification problem in the presence of colored noise, one could increase the complexity of the noise model; this, however, would lead to an increase in the complexity of the estimation and might in many cases, lead to a non-convex optimization problem. The so-called instrumental variable (IV) methods have provided a viable alternative by providing estimates that are robust to noise model estimation errors. Suppose that the colored noise $v(k)$, $z(k)$ have the following LPV structure

$$v(k) = \sum_{i=0}^{\infty} f_i(p_k, k)\cdot e(k-i), \quad (15a)$$

$$z(k) = \sum_{i=0}^{\infty} g_i(p_k, k)\cdot e(k-i), \quad (15b)$$

where $e(k)$ represents zero-mean white noise. The IV-based solutions provide unbiased estimates as long as the scheduling variables p_k are independent from $e(k)$ and the LPV filters (15a)-(15b) represent monic *infinite impulse response* (IIR) filters that are asymptotically stable. Recently, a nonparametric IV solution has been derived for nonlinear ARX models in [11] that provides unbiased estimates irrespective of the noise model. The idea follows the IV for linear systems in [15], which states that one can circumvent the problem of biased estimates due to noise modeling errors by introducing so-called instruments ζ_x and ζ_y such that

$$\mathbb{E}\{v_0(k)\zeta_x(k)\} = \mathbb{E}\{z_0(k)\zeta_y(k)\} = 0, \ \forall k \in \mathbb{Z}. \quad (16)$$

Consequently, we can introduce the instrument in the LS-SVM optimization problem derived in the previous section.

V. IV-SVM Modification

A. The IV-SVM scheme

We now modify the cost function (4) such that the condition (16) is met; we do this by introducing instruments

$\zeta_x(k),\ \zeta_y(k) \in \mathbb{R}^{1\times 2n_H}$. While the regressors $\varphi_x(k), \varphi_y(k)$ depend on the past samples of states and inputs, the instruments can be chosen by the user. The modified cost function can be written as

$$\mathcal{I}(W_x,W_y,\varepsilon_v,\varepsilon_z) = \frac{1}{2}\left(\|W_x\|_{\mathrm{F}}^2 + \|W_y\|_{\mathrm{F}}^2\right) + \frac{1}{2}\left(\sum_{k=1}^{N}\|\Gamma\varepsilon_v(k)\zeta_x(k)\|_{\mathrm{F}}^2 + \sum_{k=1}^{N}\|\Psi\varepsilon_z(k)\zeta_y(k)\|_{\mathrm{F}}^2\right), \tag{17}$$

The Lagrangian will now be defined as

$$\mathcal{L}(W_x,W_y,\boldsymbol{\alpha},\boldsymbol{\beta},\varepsilon_v,\varepsilon_z) = \mathcal{I}(W_x,W_y,\varepsilon_v,\varepsilon_z) - \sum_{j=1}^{N}\alpha_j^\top\left\{W_x\varphi_x^\top(j) + \varepsilon_v(j) - x(j+1)\right\} - \sum_{j=1}^{N}\beta_j^\top\left\{W_y\varphi_y^\top(j) + \varepsilon_z(j) - y(j)\right\}. \tag{18}$$

Like before, the stationary points for the Lagrangian are obtained by solving the KKT conditions as follows

$$\frac{\partial\mathcal{L}}{\partial W_x} = 0 \Rightarrow W_x = \sum_{j=1}^{N}\alpha_j\varphi_x(j), \tag{19a}$$

$$\frac{\partial\mathcal{L}}{\partial W_y} = 0 \Rightarrow W_y = \sum_{j=1}^{N}\beta_j\varphi_y(j), \tag{19b}$$

$$\frac{\partial\mathcal{L}}{\partial \alpha_j} = 0 \Rightarrow x(j+1) = W_x\varphi_x^\top(j) + \varepsilon_v(j), \tag{19c}$$

$$\frac{\partial\mathcal{L}}{\partial \beta_j} = 0 \Rightarrow y(j) = W_y\varphi_y^\top(j) + \varepsilon_z(j), \tag{19d}$$

$$\frac{\partial\mathcal{L}}{\partial \varepsilon_v(j)} = 0 \Rightarrow \varepsilon_v(j) = \left(2\zeta_x(j)\zeta_x^\top(j)\Gamma^\top\Gamma\right)^{-1}\alpha_j, \tag{19e}$$

$$\frac{\partial\mathcal{L}}{\partial \varepsilon_z(j)} = 0 \Rightarrow \varepsilon_z(j) = \left(2\zeta_y(j)\zeta_y^\top(j)\Psi^\top\Psi\right)^{-1}\beta_j. \tag{19f}$$

Next, like before, we eliminate the primal decision variables by substituting (19a)-(19f) into (2) and obtain the following:

$$x(k+1) = W_x\varphi_x^\top(p_k) + \varepsilon_v(k) = \underbrace{\left\{\sum_{j=1}^{N}\alpha_j\varphi_x(p_j)\right\}}_{W_x}\varphi_x^\top(p_k) + \underbrace{\left(2\zeta_x(k)\zeta_x^\top(k)\Gamma^\top\Gamma\right)^{-1}\alpha_k}_{\varepsilon_v(k)}, \tag{20}$$

$$y(k) = W_y\varphi_y^\top(k) + \varepsilon_z(k) = \underbrace{\left\{\sum_{j=1}^{N}\beta_j\varphi_y(j)\right\}}_{W_y}\varphi_y^\top(k) + \underbrace{\left(2\zeta_y(k)\zeta_y^\top(k)\Psi^\top\Psi\right)^{-1}\beta_k}_{\varepsilon_z(k)}. \tag{21}$$

The kernel matrices Ω and Ξ are defined as before, and (20)-(21) can be written in compact form as

$$X = \boldsymbol{\alpha}\Omega + (\Gamma^\top\Gamma)^{-1}\boldsymbol{\alpha}Z_x, \tag{22}$$

$$Y = \boldsymbol{\beta}\Xi + (\Psi^\top\Psi)^{-1}\boldsymbol{\beta}Z_y, \tag{23}$$

where

$$Z_x = \mathrm{diag}\Big(\left(2\zeta_x(1)\zeta_x^\top(1)\right)^{-1}, \cdots, \left(2\zeta_x(N)\zeta_x^\top(N)\right)^{-1}\Big), \tag{24}$$

$$Z_y = \mathrm{diag}\Big(\left(2\zeta_y(1)\zeta_y^\top(1)\right)^{-1}, \cdots, \left(2\zeta_y(N)\zeta_y^\top(N)\right)^{-1}\Big). \tag{25}$$

The solution to (22)-(23) can be obtained as follows

$$XZ_x^{-1} = \boldsymbol{\alpha}\Omega Z_x^{-1} + (\Gamma^\top\Gamma)^{-1}\boldsymbol{\alpha},$$
$$\mathrm{vec}(\boldsymbol{\alpha}) = \underbrace{\Big(I_N \otimes (\Gamma^\top\Gamma)^{-1} + (\Omega Z_x^{-1})^\top \otimes I_n\Big)^{-1}}_{D_x^{IV}}\mathrm{vec}(XZ_x^{-1}), \tag{26}$$

$$YZ_y^{-1} = \boldsymbol{\beta}\Xi Z_y^{-1} + (\Psi^\top\Psi)^{-1}\boldsymbol{\beta},$$
$$\mathrm{vec}(\boldsymbol{\beta}) = \underbrace{\Big(I_N \otimes (\Psi^\top\Psi)^{-1} + (\Xi Z_y^{-1})^\top \otimes I_{n_y}\Big)^{-1}}_{D_y^{IV}}\mathrm{vec}(YZ_y^{-1}). \tag{27}$$

Remark 2: By choosing the regularization parameters as $\gamma_1 = \cdots = \gamma_n$ and $\psi_1 = \cdots = \psi_{n_y}$, one can, albeit at the cost of introducing conservatism in the estimation, simplify the solution to (22)-(23) by avoiding the calculation of inverses in D_x^{IV} and D_y^{IV}. Equations (22)-(23) can then be written in a simplified manner with a simpler solution as

$$X = \boldsymbol{\alpha}\Omega + \boldsymbol{\alpha}Z_x\gamma_1^{-2} \implies \boldsymbol{\alpha} = X\underbrace{\left(\Omega + Z_x\gamma_i^{-2}\right)^{-1}}_{\hat{D}_x^{IV}}, \tag{28}$$

$$Y = \boldsymbol{\beta}\Xi + \boldsymbol{\beta}Z_y\psi_1^{-2} \implies \boldsymbol{\beta} = Y\underbrace{\left(\Xi + Z_y\psi_i^{-2}\right)^{-1}}_{\hat{D}_y^{IV}}. \tag{29}$$

Thus, calculation of the inverses $D_x^{IV} \in \mathbb{R}^{Nn\times Nn}$ and $D_y^{IV} \in \mathbb{R}^{Nn_y\times Nn_y}$ is replaced by calculating the inverses of matrices with much smaller dimensions, *i.e.* $\hat{D}_x^{IV} \in \mathbb{R}^{N\times N}$, $\hat{D}_y^{IV} \in \mathbb{R}^{N\times N}$.

B. Selection of the Instruments

To ensure unbiased estimates, the instruments $\zeta_x(k)$ and $\zeta_y(k)$ should be uncorrelated to the noise signals $v(k)$ and $z(k)$; in other words, condition (16) should hold true. As described in [15], the chosen instrument should be correlated with the regression variables, in this case, $\varphi_x(k)$ and $\varphi_y(k)$, but should be uncorrelated with the noise processes. Most instruments used in practice are generated by passing the past inputs through a filter; this is true in the LTI case where instruments can be generated using a *least squares* (LS) estimated model, while estimated LS-SVM-based nonlinear models have been used to generate instruments in the nonlinear [11] and LPV [11] cases in the input-output form. While the choice of an instrument generally remains an open problem, and depends highly on the structure of the system and the noise model, following the general principles outlined in [15] and later adopted in [11], we choose the instruments as

$$\zeta_x^\top(k) = [(\Phi_1(p_k)\xi_1(k))^\top\ (\Phi_2(p_k)\xi_2(k))^\top]^\top, \tag{30}$$

$$\zeta_y^\top(k) = \Phi_3(p_k)\xi_1(k), \tag{31}$$

where $\xi_i(k) = \hat{x}(k)$ and $\xi_i(k) = u(k)$ for $i = 1, 2$, respectively. Variable $\hat{x}(k)$ denotes noise-free states. Since we usually do not have access to noise-free measurements,

Algorithm 1 IV-SVM algorithm for state-space LPV identification

Initialize: iter = 0

1: Given $\mathcal{D} = \{u(k), x(k), y(k), p(k)\}_{k=1}^{N}$, obtain an initial estimate $\hat{A}(p_k)$, $\hat{B}(p_k)$, and $\hat{C}(p_k)$ by obtaining Lagrange multipliers $\boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)}$ using LS-SVM solution (12)-(13) as proposed in Section III; this model is labeled as $\mathcal{M}^0$.

2: Simulate the developed model in Step 1 to get estimates $\hat{x}(k)$.

3: iter = 1

while $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ do not converge according to (34) **do**

4: Define the instruments $\zeta_x(k), \zeta_y(k)$ as (30)-(31); compute (32)-(33) to obtain the matrices Z_x, Z_y.

5: Estimate the Lagrange multipliers $\boldsymbol{\alpha}^{\text{iter}}, \boldsymbol{\beta}^{\text{iter}}$ using IV-SVM solution (26)-(27) and solve (14a)-(14c) to obtain the model $\mathcal{M}^{\text{iter}}$.

6: Using $\mathcal{M}^{\text{iter}}$, generate the estimates $\hat{x}(k)$.

7: iter $\leftarrow$ iter + 1.

end while

$\hat{x}(k)$ can be considered to be estimates of $x(k)$. The inner-product $\zeta_i(k)\zeta_i^\top(k)$ $(i = x, y)$ in (24)-(25) can now be replaced by kernel functions, giving us Z_x and Z_y as follows:

$$Z_x = \text{diag}\Bigg(\Big(2\sum_{i=1}^{2} \xi_i^\top(1)\bar{k}^i(p_1,p_1)\xi_i(1)\Big)^{-1}, \cdots, \Big(2\sum_{i=1}^{2} \xi_i^\top(N)\bar{k}^i(p_N,p_N)\xi_i(N)\Big)^{-1}\Bigg), \quad (32)$$

$$Z_y = \text{diag}\Bigg(\Big(2\xi_1^\top(1)\bar{k}^3(p_1,p_1)\xi_1(1)\Big)^{-1}, \cdots, \Big(2\xi_1^\top(N)\bar{k}^3(p_N,p_N)\xi_1(N)\Big)^{-1}\Bigg). \quad (33)$$

In case the chosen kernel function $\bar{k}^i$ is an RBF kernel, $\bar{k}^i(p_j,p_j) = 1$; however, we refrain from making this simplification in (32)-(33) since the derived expression is generic and may pertain to the choice of any kernel function, not necessarily RBF. The estimates of the state-space matrices $\hat{A}(\cdot)$, $\hat{B}(\cdot)$ and $\hat{C}(\cdot)$ can then be computed by using (14a)-(14c) in which $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the Lagrange multipliers obtained from the IV-SVM update (26)-(27), and $x(k)$ and $u(k)$ denote the recorded noisy states and inputs, respectively. The estimates $\hat{x}(k)$ used for the calculation of Z_x and Z_y can be considered to be those obtained from the developed SVM-based model, and can be improved iteratively until the solutions $\boldsymbol{\alpha}, \boldsymbol{\beta}$ converge according to the following criterion:

$$\mathcal{O}^{\text{iter}} = \left\|\boldsymbol{\alpha}^{\text{iter}} - \boldsymbol{\alpha}^{\text{iter}-1}\right\|_{\text{F}} + \left\|\boldsymbol{\beta}^{\text{iter}} - \boldsymbol{\beta}^{\text{iter}-1}\right\|_{\text{F}} \leq \epsilon, \quad (34)$$

where ϵ is a small number chosen for the stopping criterion of the iterative procedure. Detailed steps needed to implement the IV-SVM routine are described in Algorithm 1.

VI. Simulation Results

The following numerical example of a second order discrete-time state-space LPV model is considered.

$$x(k+1) = A(p_k)x(k) + B(p_k)u(k) + v(k),$$

$$A(p_k) = \begin{bmatrix} \frac{1}{6}p_k^2 & p_k^3 + p_k \\ 0 & \frac{1}{8}p_k^3 \end{bmatrix}, B(p_k) = \begin{bmatrix} \text{sat}(p_k) \\ \sin(q) + \cos(q) \end{bmatrix},$$

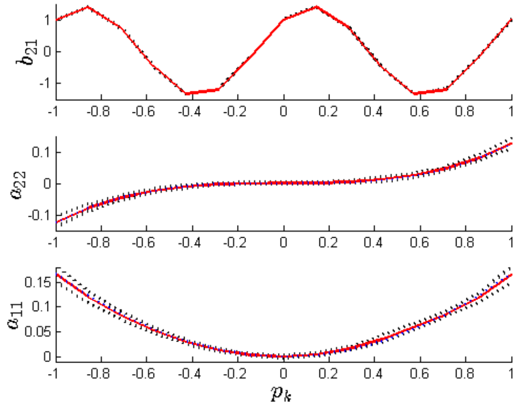

Fig. 1. Functional dependencies of elements $\{a_{11}\}$ and $\{a_{22}\}$, $\{b_{21}\}$ on the scheduling variable p_k as estimated by IV-SVM; solid red lines denote the mean values of the estimates over the Monte-Carlo simulations, while dotted black lines denote the standard deviation. Original functional dependencies are shown by dotted blue line.

TABLE I

Monte-Carlo simulation results for output BFR

	Mean (BFR %)	Std. (BFR %)
LS-SVM	86.22	0.8247
IV-SVM	97.72	0.1596

where $q = 2\pi p_k$, and

$$\text{sat}(p_k) = \begin{cases} -0.5, & p_k < -0.5 \\ 0.5, & p_k > 0.5 \\ 0 & \text{otherwise.} \end{cases}$$

The signal $v(k)$ represents a colored noise correlated with the scheduling variable p_k and given by

$$v(k+1) = \begin{bmatrix} 0.95p_k & 0 \\ 0 & 0.95p_k \end{bmatrix} v(k) + \begin{bmatrix} -0.3p_k^3 \\ -0.15p_k \end{bmatrix} e(k),$$

where $e(k) \sim \mathcal{N}(0, \sigma_e^2)$ is a white noise sequence. Given the measurements of states $x(k)$, we are interested only in the estimation of $A(p_k)$ and $B(p_k)$, and hence, only Lagrange multipliers $\boldsymbol{\alpha}$ are sought; estimation of $C(p_k)$ would follow the same procedure for the estimation of $\boldsymbol{\beta}$ as outlined in the previous section. A total of 1000 samples of scheduling variables $p_k \in [-1, 1]$ are generated such that $p_k = \sin(0.8k)$. Uniformly distributed random inputs $u \in [-1, 1]$ are generated. White noise $e(k)$ is generated with standard deviation $\sigma_e = 0.15$, and $v(k)$ is calculated. This results in an average SNR of 12dB over the two states. Data is divided into 700 and 300 samples for estimation and validation, respectively. An RBF function is chosen for the kernel functions $\bar{k}^i$ with $\sigma_1 = \sigma_2 = 0.7$ for both the LS-SVM and the IV-SVM cases. Regularization parameters $\gamma_{1,2}$ are tuned to 2400 and 3200 for the two cases by searching over a grid space of hyperparameters in order to maximize the following *best fit rate* (BFR) in each case

$$\text{BFR} := 100\% \cdot \max\left(\frac{\|x - \hat{x}\|_2}{\|x - \bar{x}\|_2}, 0\right),$$

TABLE II

MONTE-CARLO SIMULATION RESULTS: FUNCTIONAL DEPENDENCIES AS ESTIMATED BY THE PROPOSED IV-SVM-BASED METHOD. BEST FIT RATE (BFR) STATISTICS (%).

Function	Mean	Std.	Function	Mean	Std.
$a_{11}(p_k)$	79.775	0.015	$a_{12}(p_k)$	99.099	0.052
$a_{21}(p_k)$	96.012	0.078	$a_{22}(p_k)$	86.501	0.044
$b_{11}(p_k)$	97.315	0.101	$b_{21}(p_k)$	97.508	0.025

where $\hat{x}$ represents simulated states of the estimated model and $\bar{x}$ denotes sample mean value of the states.

The proposed LS-SVM and IV-SVM algorithms are run and the Lagrange multipliers are estimated in each case. The performance of the algorithms is assessed on noise-less validation data set and average BFR values are calculated for the two states. BFR statistics over 50 runs of Monte-Carlo simulations are tabulated in Table I.

Elements of the identified state-space matrices using IV-SVM show remarkable accuracy. The BFR values for the functional dependencies of these elements evaluated over the interval $[-1, 1]$ are tabulated in Table II. Figure 1 shows three of these dependencies, namely, elements a_{11}, a_{22}, and b_{21}, with the dashed line showing mean functional value over all Monte-Carlo runs and the dotted black line showing the standard deviation. The improvement in output BFR values for the IV-SVM is reflected in the estimation of these functions as well. In Figure 2, we compare the mean functional values as estimated by the LS-SVM and IV-SVM algorithms. Owing to the colored nature of the noise, and the fact that the considered noise is correlated with the scheduling variables, some functional estimates like a_{22} in the LS-SVM case show a visible bias.

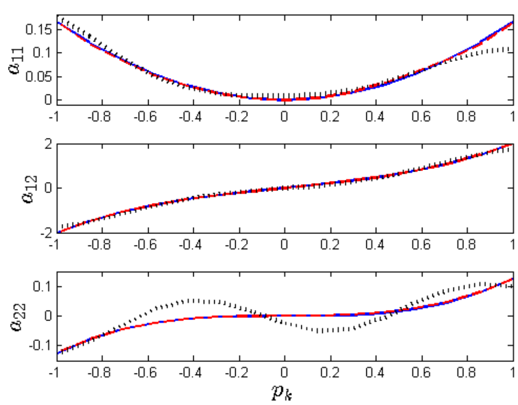

Fig. 2. Functional dependencies of elements $\{a_{11}\}$ and $\{a_{12}\}$, $\{a_{22}\}$ on the scheduling variable p_k as estimated by LS-SVM (dotted black) and IV-SVM (solid red). Original functional dependencies are shown by dashed blue line.

VII. CONCLUDING REMARKS

This paper has presented a nonparametric identification scheme for state-space LPV models under generic noise conditions. The proposed technique makes use of the non-parametric LS-SVM method that has shown encouraging estimation results for input-output LPV models, and further incorporates the so-called "instruments" to induce robustness to noise modeling errors. Instrumental variables have proven to be effective in the LTI context and recently seen their extension to nonlinear ARX and polynomial LPV models. To gauge the performance of the proposed method, we have considered simulation studies with a high level of output noise, which was not only colored but also correlated with the scheduling variables. The results have shown encouraging improvements compared to LS-SVM based estimation methods in terms of providing unbiased estimates. Developing IV-SVM-based identification methods for state-space LPV models without the availability of state measurements is the focus of our ongoing research.

REFERENCES

[1] W. J. Rugh and J. S. Shamma, "Research on gain scheduling," *Automatica*, vol. 36, no. 10, pp. 1401–1425, 2000.

[2] F. Felici, J. V. Wingerden, and M. Verhaegen, "Subspace identification of MIMO LPV systems using a periodic scheduling sequence," *Automatica*, vol. 43, no. 10, pp. 1684–1697, 2007.

[3] 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 Conf. Decision and Control, and European Control Conf.*, Dec 2011, pp. 3192–3197.

[4] V. Verdult, "Nonlinear system identification: a state-space approach," Ph.D. dissertation, University of Twente, The Netherlands, 2002.

[5] R. Tóth, P. Heuberger, and P. V. den Hof, "Prediction-error identification of LPV systems: Present and beyond," in *Control of Linear Parameter Varying Systems with Applications*. Springer, 2012, pp. 27–58.

[6] B. Bamieh and L. Giarre, "Identification of linear parameter varying models," *Int. J. Robust Nonlinear Control*, vol. 12, no. 9, pp. 841–853, 2002.

[7] V. Verdult and M. Verhaegen, "Subspace identification of multivariable linear parameter-varying systems," *Automatica*, vol. 38, no. 5, pp. 805–814, 2002.

[8] J. V. Wingerden and M. Verhaegen, "Subspace identification of MIMO LPV systems: The PBSID approach," in *Proc. of the 47th IEEE Conf. Decision and Control*, 2008, pp. 4516–4521.

[9] B. Schölkopf and A. Smola, *Learning with Kernels*. Cambridge, MA: MIT Press, 2002.

[10] S. Rizvi, J. Mohammadpour, R. Tóth, and N. Meskin, "Parameter set-mapping using kernel-based PCA for linear parameter-varying systems," in *Proc. of the 13th European Control Conf.*, 2014, pp. 2744–2749.

[11] V. Laurain, R. Tóth, D. Piga, and W. Zheng, "An instrumental least squares support vector machine for nonlinear system identification," *Automatica*, vol. 54, pp. 340–347, 2015.

[12] F. Abbasi, J. Mohammadpour, R. Tóth, and N. Meskin, "A support vector machine-based method for LPV-ARX identification with noisy scheduling parameters," in *Proc. of the 13th European Control Conf.*, 2014, pp. 370–375.

[13] S. Rizvi, J. Mohammadpour, R. Tóth, and N. Meskin, "A kernel-based approach to MIMO LPV atate-space identification and application to a nonlinear process system," in *Proc. of the 1st IFAC Workshop on Linear Parameter varying Systems*, Grenoble, France, 2015, to appear.

[14] P. L. dos Santos, T.-P. Azevedo-Perdicoulis, J. Ramos, S. Deshpande, D. Rivera, and J. de Carvalho, "LPV system identification using a separable least squares support vector machines approach," in *Proc. of the 53rd IEEE Conf. Decision and Control*, 2014, pp. 2548–2554.

[15] L. Ljung, *System identification: theory for the user*. NJ: Prentice Hall, 1987.