Preprints of the 16th IFAC Symposium on System Identification
The International Federation of Automatic Control
Brussels, Belgium. July 11-13, 2012

Nonparametric Identification of LPV Models Under General Noise Conditions: An LS-SVM Based Approach

V. Laurain [*,***] **R. Tóth** [**] **W-X. Zheng** [***] **M. Gilson** [*]

[*] *Centre de Recherche en Automatique de Nancy (CRAN), Université de Lorraine, CNRS, 2, rue Jean Lamour 54506 Vandoeuvre Cedex, France, email:* vincent.laurain@cran.uhp-nancy.fr
[**] *Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands, email:* r.toth@tudelft.nl.
[***] *School of Computing and Mathematics, University of Western Sydney, Penrith NSW 2751, Australia.*

Abstract: Parametric identification approaches in the Linear Parameter-Varying (LPV) setting require optimal prior selection of a set of functional dependencies, used in the parametrization of the model coefficients, to provide accurate model estimates of the underlying system. Consequently, data-driven estimation of these functional dependencies has a paramount importance, especially when very limited *a priori* knowledge is available. Existing over-parametrization and nonparametric methods dedicated to nonlinear estimation offer interesting starting points for this problem, but need reformulation to be applied in the LPV setting. Moreover, most of these approaches are developed under quite restrictive auto-regressive noise assumptions. In this paper, a nonparametric Least-Squares Support Vector Machine (LS-SVM) approach is extended for the identification of LPV polynomial models. The efficiency of the approach in the considered noise setting is shown, but the drawback of the auto-regressive noise assumption is also demonstrated by a challenging LPV identification example. To preserve the attractive properties of the approach, but to overcome the drawbacks in the estimation of polynomial LPV models in a general noise setting, a recently developed Instrumental Variable (IV)-based extension of the LS-SVM method is applied. The performance of the introduced IV and the original LS-SVM approaches are compared in an identification study of an LPV system with unknown noise dynamics.

1. INTRODUCTION

Estimation of *linear parameter-varying* (LPV) polynomial models in an *input-output* (IO) setting has received a significant attention recently in the identification literature (see, *e.g.*, Bamieh and Giarré [2002], Laurain et al. [2010], Hsu et al. [2008], Cerone and Regruto [2008], Tóth [2010]). The signal relations in these LPV models are considered to be linear just as in the *linear-time invariant* (LTI) case, but the parameters are assumed to be functions of a measurable time-varying signal, the so-called *scheduling variable* $p : \mathbb{Z} \to \mathbb{P}$. Here the compact set $\mathbb{P} \subset \mathbb{R}^{n_\mathrm{P}}$ denotes the *scheduling space*. This particular structure of LPV models allows an efficient extension of the LPV prediction error framework for data-driven modeling of both nonlinear and time-varying behaviors (Tóth [2010]).

Most available LPV parametric identification methods, even subspace or set-membership approaches, are analyzed under the assumption that the nonlinear dependencies of the model coefficients on p are appropriately parameterized (Bamieh and Giarré [2002], Laurain et al. [2010], Tóth [2010]). Such an assumption is crucial not only from the view point of stochastic consistency, but also due to issues of numerical and structural bias. The adequate selection of the p-dependent functions in a LPV model parametrization is a challenging problem in general, for which the most often applied solution in the parametric case leads to a (heavy) over-parametrization of the model coefficients with a rich set of dependencies (Tóth et al. [2009]). Hence the variance of the estimates can be seriously large, even if the order of the actual model is low.

This problem is similar to the classical model structure or regressor selection problem encountered in statistics and LTI identification (Ljung [1999]). It also plays an important role in many nonlinear system identification problems where only a limited *a priori* knowledge on the model structure and hence on the involved nonlinearites is available. To cope with this problem, especially when the length of the available data-record is relatively low, an elegant solution leads through the regularization of the optimization criterion. This allows to reduce the number of estimated non-zero parameters and hence the variance on the possible expanse of bias. Following this idea, many approaches have been introduced for linear regression problems from sparse estimators to shrinkage methods. Among these methods, *least-squares support vector machine*'s (LS-SVM's), originating form statistical learning theory (Vapnik [1998]), have had a significant impact on *nonlinear block model* identification (see Suykens et al. [2002], Falck

et al. [2010], Giri and Bai [2010]), due to the so-called *kernel-trick*, which allows a significant reduction of the variance without restricting the representation capability of the parametrization. An LS-SVM idea based identification approach has been also recently introduced in the LPV case (Tóth et al. [2011]). The attractive properties of this approach have been demonstrated in terms of computational load and achieved bias and variance compared to other regularized LPV methods (Hsu et al. [2008], Tóth et al. [2009]).

However, a major drawback of regularization based approaches has been the severe restrictions on the noise, most commonly in terms of an auto-regressive noise process with the same dynamics as the deterministic system. To cope with colored noise, some modified variants of the LS-SVM algorithm has been developed in the nonlinear identification setting (Espinoza et al. [2006], Falck et al. [2010]), but the validity of the used assumptions on the noise and the effect of violating them have never been discussed. In practice, noise often does not satisfy completely all prior assumptions, hence bias effects can be unpredictable especially in a nonparametric context. In the parametric framework, one of the most efficient methods to handle noise modeling error is the *instrumental variable* (IV) technique (Söderström and Stoica [1983]). It has been applied to many types of linear models, like LPV models (Laurain et al. [2010]), and more recently has been introduced in the LS-SVM framework (Laurain et al. [2011]).

The contribution of the current paper is to extend the LPV LS-SVM identification of polynomial IO models (see Tóth et al. [2011]) to the realistic assumption of general priori unknown noise conditions. This is achieved based on the recently introduced IV modification of the LS-SVM approach. The significance of this contribution is highlighted by an analysis of the original LS-SVM scheme under realistic (general) noise conditions, demonstrating the clear need for improvement.

The paper is organized as follows. In Section 2, the identification problem as well as the assumed model structure are introduced. The LPV LS-SVM approach is briefly summarized in Section 3. This is followed in Section 4 by an analysis of the approach via a representative example where the noise conditions violate the underlying assumption. The recently proposed IV scheme for LS-SVM methods is introduced in Section 5 and its application to LPV models is proposed in Section 6. In Section 7, both methods are compared based on their statistical properties and prediction capabilities. Finally, the conclusions are presented in Section 8.

2. PROBLEM DESCRIPTION

To set the preliminaries for the upcoming discussion, the concept of the data-generating system and the considered model structure are briefly introduced in this section.

2.1 The considered system

The LPV data-generating system, considered in this paper, is defined as

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

where $k\in\mathbb{Z}$ is the discrete time, $u:\mathbb{Z}\to\mathbb{R}$ and $y:\mathbb{Z}\to\mathbb{R}$ denote the input and the output signals respectively, $p:\mathbb{Z}\to\mathbb{P}$ is the so called *scheduling variable* with range $\mathbb{P}\subseteq\mathbb{R}^{n_\mathrm{p}}$, v_o is a bounded stochastic noise process and, without loss of generality, $a_0^\mathrm{o}=1$. Furthermore (to keep the notation simple), the nonlinear functions $a_i^\mathrm{o}, b_j^\mathrm{o}:\mathbb{P}\to\mathbb{R}$, describing the coefficients of the dynamic relation (1), are assumed to have *static dependence* on p. The latter means, that these functions depend only on $p(k)$, *i.e.*, the instantaneous value of p. In practice however – as it is shown in Section 7 – the results of this paper can be straightforwardly extended to the case of causal dynamic dependence (dependence on both the instantaneous and past values of p).

In the sequel, the focus of the discussion is going to be on the identification of the nonlinear functions $\{a_i^\mathrm{o}\}_{i=1}^{n_\mathrm{a}}$ and $\{b_j^\mathrm{o}\}_{j=0}^{n_\mathrm{b}}$, *i.e.*, the estimation/learning of these dependencies on p. Thus, the orders n_a and n_b of the polynomial operators in (1) as well as the scheduling variable p are assumed to be *a priori* known. Moreover, both u and p are considered to be deterministic (noise-free) which is the most commonly used assumption in the LPV identification setting (Tóth [2010]). Furthermore, we also leave the noise v_o unspecified for the time being.

2.2 The considered model

The idea upon which the intended nonparametric method is based, is to assume that the nonlinear p-dependent coefficient functions $f_i(p_k)$ (corresponding to a_i and b_j) lay in a high (possibly infinite) dimensional plane:

$$f_i(p(k))=\sum_{i=1}^{n_\mathrm{g}}\rho_i^\top\phi_i(p(k)), \quad (2)$$

where each $\phi_i:\mathbb{R}^\cdot\to\mathbb{R}^{n_\mathrm{H}}$ denotes an undefined, potentially infinite ($n_\mathrm{H}=\infty$) dimensional function, so-called *feature map*, and $\rho_i\in\mathbb{R}^{n_\mathrm{H}}$ is the associated parameter vector.

In the presented LPV context, this formulation manifests in the model structure

$$\mathcal{M}_\rho: \qquad y(k)=\sum_{i=1}^{n_\mathrm{g}}\rho_i^\top\phi_i(p_k)x_i(k)+e(k), \quad (3)$$

where

$$x_i(k)=y(k-i), \qquad i=1,\dots,n_\mathrm{a}, \quad (4a)$$
$$x_{n_\mathrm{a}+1+j}(k)=u(k-j), \qquad j=0,\dots,n_\mathrm{b}. \quad (4b)$$

Additionally, introduce $\rho=[\,\rho_1^\top\ \dots\ \rho_{n_\mathrm{g}}^\top\,]^\top\in\mathbb{R}^{n_\mathrm{g}n_\mathrm{H}}$ and

$$\varphi(k)=\left[\,\phi_1^\top(p_k)x_1(k)\ \dots\ \phi_{n_\mathrm{g}}^\top(p_k)x_{n_\mathrm{g}}(k)\,\right]^\top, \quad (5)$$

such that (3) can be rewritten in the regression form

$$y(k)=\rho^\top\varphi(k)+e(k). \quad (6)$$

The major difficulty here is to correctly model the nonlinearities and to minimize the risk of structural bias while keeping the computational time and the variance of the estimate relatively low. However, minimizing the structural bias by increasing the dimensions of the different ϕ_i feature maps directly increases both the variance of the estimates and the required computational time. To cope with these issues, regularization based solutions have been reported in Tóth et al. [2011] and Tóth et al. [2009]. In the sequel, we are going to concentrate on the nonparametric LS-SVM approach presented in Tóth et al. [2011].

3. THE LS-SVM METHOD FOR LPV MODELS

This section briefly summarizes the LPV LS-SVM which we intended to further extended to the general noise case. The identification criterion – the cost function by which the aim of the estimation is defined – is a regularized least-squares (prediction) error criterion and in this approach is defined in the form of

$$\mathcal{J}(\rho, e) = \frac{1}{2}\|\rho\|_{\ell_2}^2 + \frac{\gamma}{2}\|e(t_k)\|_{\ell_2}^2, \tag{7}$$

where the bias-variance trade-off is controlled by the regularization hyper-parameter $\gamma \in \mathbb{R}_0^+$. Regularization in (7) makes possible to handle large-scale over-parametrizaiton problems like in sparse estimators (Tóth et al. [2009]).

Introduce the output data vector $Y = [\, y(1) \ \dots \ y(N) \,]^\top$ and the regressor matrix $\Phi = [\, \varphi(1) \ \dots \ \varphi(N) \,]^\top$. Without going into details about the assumed noise dynamics, it can be shown that estimates $a_i(\centerdot)$ and $b_j(\centerdot)$ of the nonlinear functions $a_i^{\mathrm{o}}(\centerdot)$ and $b_j^{\mathrm{o}}(\centerdot)$ can be obtained by applying the following three steps (Tóth et al. [2011]):

Step 1: Construct the matrix $\Omega = \Phi\Phi^\top \in \mathbb{R}^{N\times N}$ by using the following property

$$[\Omega]_{j,k} = \sum_{i=1}^{n_{\mathrm{g}}} [\Omega^{(i)}]_{j,k} \tag{8}$$

with

$$\begin{aligned}[\Omega^{(i)}]_{j,k} &= x_i(j)\phi_i^\top(j)\phi_i(k)x_i(k),\\ &= x_i(j)K_i\big(p(j),p(k)\big)x_i(k).\end{aligned}$$

Here, $\{K_i\}_{i=1}^{n_{\mathrm{g}}}$ are positive definite so-called kernel functions. These kernels define the inner products of $\phi_i^\top(j)\phi_i(k)$ and hence characterize the feature maps $\{\phi_i\}_{i=1}^{n_{\mathrm{g}}}$. Specification of the kernel functions instead of ϕ_i is called the *kernel trick* (Vapnik [1998], Schölkopf and Smola [2002]), which allows the identification of the coefficient functions a_i and b_j without explicitly defining the feature maps involved. In other words, this allows the implicit definition of high dimensional feature maps without actually handling the high dimensional objects, resulting in low computational time of the estimation compared to the representation capabilities. A typical type of kernel is, for example, the *Radial Basis Function* (RBF) kernel:

$$K_i(p(j),p(k)) = \exp\left(-\tfrac{\|p(j)-p(k)\|_2^2}{\sigma_i^2}\right). \tag{9}$$

Step 2: Use the matrix Ω to compute the dual optimal solution of (7) as

$$\hat{\alpha} = \big(\Omega + \gamma^{-1}I_N\big)^{-1}Y. \tag{10}$$

Step 3: Compute the estimated nonlinearities as

$$\hat{a}_i(\centerdot) = \hat{\rho}_i^\top\phi_i(\centerdot) = \sum_{k=1}^{N}\hat{\alpha}_k x_i(k)K^i(p(k),\centerdot), \tag{11a}$$

$$\hat{b}_j(\centerdot) = \hat{\rho}_{\tilde{j}}^\top\phi_{\tilde{j}}(\centerdot) = \sum_{k=1}^{N}\hat{\alpha}_k x_{\tilde{j}}(k)K^{\tilde{j}}(p(k),\centerdot). \tag{11b}$$

where $\tilde{j} = n_{\mathrm{a}} + 1 + j$ and $\hat{\alpha}_k$ is the k^{th} component of $\hat{\alpha}$. Note that the parameter vector ρ is never accessible in the SVM framework as only the combined estimates $\hat{a}_i(\centerdot) = \hat{\rho}_i{}^\top\phi_i(\centerdot)$ or $\hat{b}_j(\centerdot) = \hat{\rho}_{\tilde{j}}^\top\phi_{\tilde{j}}(\centerdot)$ can be computed using the kernel functions defined.

4. THE PITFALLS RELATED TO NOISE

So far, we have neither specified the noise v_{o} in the data-generating system (1) nor revealed its effect on the outcome of the function estimates. Nevertheless, for an experienced reader it might have been already clear that considering the model (3) and the optimization criterion (7), $e(k)$ (and $v_{\mathrm{o}}(k)$ respectively) must be a white stochastic noise process to achieve unbiased estimates by the presented LS-SVM scheme. Such an assumption is often found quite restrictive in practice. Thus, to extend this estimation approach to a broader class of colored noise processes – white noise filtered by an LTI filter –, some modified variants of the LS-SVM algorithm have been developed in the nonlinear identification setting (Espinoza et al. [2006], Falck et al. [2010]). Nevertheless, all theses techniques have been developed by implicitly assuming that the true noise process lies in the model set defined. In order to show the weaknesses of this assumption in the LPV case, we consider an example in which the noise is also correlated with p, *i.e.*, it has an LPV structure, a case which can often be found in LPV systems with process or actuator noise (Tóth [2010]). This means that an LTI assumption on the noise model is invalid.

Considered the following LPV data-generating system $\mathcal{S}_{\mathrm{o}}$:

$$\begin{aligned}y(k) =& \sum_{i=1}^{2} a_i(p(k),p(k-1))y(k-i)\\ &+ \sum_{j=1}^{2} b_j(p(k),p(k-1))u(k-i) + v_{\mathrm{o}}(k),\end{aligned} \tag{12a}$$

$$\begin{aligned}v_{\mathrm{o}}(k) =& \big[0.85\mathrm{sign}\big(p(k-1)-0.2\big)\big]v_{\mathrm{o}}(k-1)\\ &+ \big[1 + (0.2 + p(k))q^{-1}\big]e_{\mathrm{o}}(k),\end{aligned} \tag{12b}$$

with $\mathbb{P} = [-0.3, 0.7]$, e_{o} being a white noise process with zero-mean Gaussian distribution and

$$\begin{aligned}a_1(x,y) &= -0.7 + 0.5x - 0.3y,\\ a_2(x,y) &= -0.3 + 0.4x + 0.3y - 1.8xy - 0.8x^2 + 0.5y^2,\\ b_1(x,y) &= 0.5\left(\sin(10x) + \cos(4y)\right),\\ b_2(x,y) &= \mathrm{sinc}(5\sqrt{x^2+y^2}) - 0.7.\end{aligned}$$

The available measurement data consists of 50 estimation data sets and 50 validation data sets, each with data records $\{u(k), y(k), p(k)\}_{k=1}^{N}$ of length $N = 600$, where $u(k)$, $p(k)$ are white stochastic processes with $u(k)$ having a uniform distribution $\mathcal{U}(-1,1)$ and $p(k) \in \mathcal{U}(-0.3, 0.7)$. The robustness of the proposed algorithms is analyzed under a *signal-to-noise ratio* (SNR) of 8dB, computed as $\mathrm{SNR} = 20\log\frac{P_{\mathring{y}}}{P_{v_{\mathrm{o}}}}$ where $P_{\mathring{y}}$ and $P_{v_{\mathrm{o}}}$ are the average power of the noise-free output signal $\mathring{y}$ and the true noise process v_{o} respectively.

Even knowing the correct model order, this estimation problem is challenging due to the dynamic dependency on p, the different nature of the nonlinearities involved, and the relatively low number of data with respect to the considerable power of the noise present. No *a priori* knowledge is supposed to be available concerning the nonlinearities. In order to avoid the selection problem of the kernels and all possible optimality considerations involved, the kernels used in the following estimation are all considered to be 2-dimensional RBF kernels with the same hyper parameter $\sigma > 0$:

$$K_i(\mathbf{p}_j, \mathbf{p}_k) = \exp\left(-\frac{\|\mathbf{p}_j - \mathbf{p}_k\|_2^2}{\sigma^2}\right), \quad i = 1, 2, 3, 4 \qquad (13)$$

where $\mathbf{p}_i = [\, p(i) \;\; p(i-1) \,]^\top$. The aim of choosing the same kernel for all nonlinearities is to limit the number of hyper-parameters to two: the regularization parameter γ and the width of the RBF's σ in (13) and hence the complexity of their adequate choice (or optimization). Note that such hyper-parameters represent a structural freedom in the LS-SVM estimation scheme, which can be commonly found in all regularization based approaches. One might argue that selection of these parameters corresponds to a model structure selection problem, which is indeed true, but with a seriously reduced degree of freedom and much better understood consequences of the choice (see Vapnik [1998]). For example, increasing γ reduces the bias, but increases the variance while σ is related to the inter-sample distance and therefore the search for its adequate value can be restrained to a certain range.

In this example, the hyper-parameters γ and σ are optimized by maximizing the *best fit rate* (BFR) of the estimated model w.r.t. the validation data set where:

$$\text{BFR} = 100\% \cdot \max\left(1 - \frac{\|y(k) - \hat{y}(k)\|_{\ell_2}}{\|y(k) - \bar{y}\|_{\ell_2}}, 0\right), \qquad (14)$$

with (y, u, p) being the validation data, $\bar{y}$ is the mean of y and $\hat{y}$ is the simulated output of the model (3) for (u, p), estimated by the LS-SVM on the estimation data set. This optimization has resulted in $\sigma = 0.3$ and $\gamma = 4200$ for this particular example.

In order to clearly demonstrate the pitfalls linked to the noise assumption in the LS-SVM, the "nonlinearity" in $a_1^{\text{o}}(\centerdot)$ is chosen as a simple plane. In Figure 1, the mean of the function estimates $\hat{a}_1(\centerdot)$ provided by the LS-SVM over the 50 Monte-Carlo runs using the estimation data sets is depicted together with the true $a_1^{\text{o}}(\centerdot)$. The switching dynamical nature of the underlying noise process centered at $p(k-1) = 0.2$ (see (12b)) results in a clear deterioration of the function estimates, which resemble to a saturation type of nonlinearity. This bias is due to the fact that $v_{\text{o}}(k) \neq e_{\text{o}}(k)$ which is assumed in the least-squares based estimation problem (7). Furthermore, the bias of the function estimates is different for the space $p(k-1) < 0.2$ and $p(k-1) > 0.2$. This is important because a filtering operation on the data such as proposed in Espinoza et al. [2006] might suppress the bias in one of the regions, but can not efficiently suppress it on the whole scheduling domain $\mathbb{P}$. This shows that in the realistic case of an unknown noise structure, even if a large number of data is available, it is not possible to separate the true underlying nature of the nonlinearity from the undesirable noise effect. This highlights a clear need for the improvement of the LPV LS-SVM method to be able to apply it under general, p-correlated noise conditions such as in the given example. Otherwise, violation of the assumed noise conditions can lead to seriously different models of the system preventing the synthesis of high-performance LPV controllers.

5. AN IV APPROACH FOR LS-SVM'S

To cope with general unknown noise conditions, one way is to increase the complexity of the noise model. However, this solution increases the complexity of the estimation problem, *e.g.*, resulting in a non-convex optimization, and

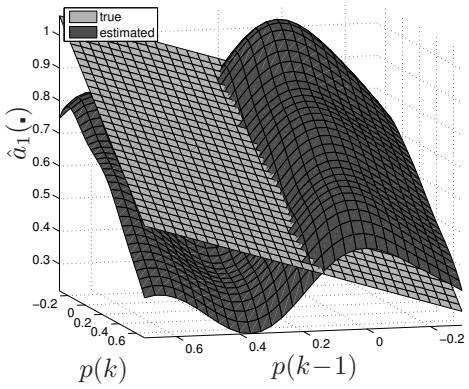

Fig. 1. Mean estimate of a_1^{o} by the LS-SVM approach.

it does not ensure that the model is rich enough to describe the dynamics of the underlying noise process. The latter unavoidably leads to an approximation trade-off in many practical scenarios.

Another possibility is to develop an estimation method robust w.r.t. modeling error on the noise. This is the core idea behind IV approaches in the classical prediction error setting. The same concept can also be harnessed in the LS-SVM case to provide, under certain conditions, unbiased function estimates independently of the true nature of the noise (Laurain et al. [2011]).

Consider a linear-regression model which can be expressed as $y(k) = \rho^\top \varphi(k) + e(k)$ and a true data-generating system which can be expressed as $y(k) = \rho_{\text{o}}^\top \varphi(k) + v_{\text{o}}(k)$. The idea behind the IV approach is to introduce a so-called instrument $\zeta(k) \in \mathbb{R}^{n_\rho}$ in the linear-regression problem which guarantees the unbiasedness (w.r.t. the noise) of the estimate by satisfying that $\mathbb{E}\{\zeta(k) v_{\text{o}}(k)\} = 0$ for all $k \in \mathbb{Z}$. In other words, an IV-based method produces an unbiased estimate under the following conditions:

- X1 The chosen instrument $\zeta(k)$ is uncorrelated to the true noise $v_{\text{o}}(k)$.
- X2 The true noise process has a zero mean and therefore $\mathbb{E}\{v_{\text{o}}(k)\} = 0$.

In the LS-SVM context, the instrument can be introduced into (6), by modifying the identification criterion (7) as

$$\hat{\rho}^{\text{IV}} = \text{sol}\left\{\frac{1}{N}\sum_{k=1}^{N} \rho + \gamma \zeta(k)\left[y(k) - \varphi^\top \rho\right] = 0\right\}. \qquad (15)$$

It can be shown (see Laurain et al. [2011]), that this criterion leads to unbiased function estimates by following three steps similar to the LS-SVM solution:

Step 1: Construct the matrix $\Xi = \Phi Z^\top \in \mathbb{R}^{N \times N}$ where $Z = [\, \zeta(1) \;\; \ldots \;\; \zeta(N) \,]^\top$ (in the LS-SVM case $\Omega = \Phi\Phi^\top$).

Step 2: Use the matrix Ξ to compute

$$\hat{\alpha} = \left(\Xi + \gamma^{-1} I_N\right)^{-1} Y. \qquad (16)$$

Step 3: The nonlinear functions $f_i(\centerdot) = (\hat{\rho}_i^{\text{IV}})^\top \varphi_i(\centerdot)$ are computed using the expression:

$$\hat{\rho}^{\text{IV}} = \sum_{k=1}^{N} \hat{\alpha}_k \zeta(k) \qquad (17)$$

A remarkable property of this modified scheme is that it requires nearly the same computational load as the original LS-SVM approach.

6. THE IV-LS-SVM IN THE LPV CASE

6.1 The choice of an instrument

In LPV systems, most p-dependent polynomial noise structures (Box-Jenkins, Output Error, ARMAX, ARX, etc.) fulfill condition X2. Therefore, any instrument uncorrelated to the noise will directly imply unbiased estimates with respect to the noise. The instrument proposed below is motivated by empirical experience originating from the LTI context, but due to space restriction, its statistical properties are not studied here. The instrument proposed is defined as:

$$\zeta(k) = \left[\, \phi_1^\top(p_k)\xi_1(k) \;\ldots\; \phi_{n_\mathrm{g}}^\top(p_k)\xi_{n_\mathrm{g}}(k) \,\right]^\top, \tag{18}$$

with

$$\xi_i(k) = \hat{y}(k-i), \qquad i = 1, \ldots, n_\mathrm{a}, \tag{19a}$$
$$\xi_{n_\mathrm{a}+1+j}(k) = u(k-j), \qquad j = 0, \ldots, n_\mathrm{b}, \tag{19b}$$

where $\hat{y}$ is either the noise-free output of the system or it denotes its estimate. As long as $\hat{y}$ is uncorrelated to the noise, condition X1 is fulfilled and therefore the obtained estimate will be unbiased. The computation of $\hat{y}$ will be discussed later.

Considering the instrument (18), Ξ is reads as

$$[\Xi]_{j,k} = \sum_{i=1}^{n_\mathrm{g}} [\Xi^i]_{j,k} \tag{20}$$

with $[\Xi^{(i)}]_{j,k} = x_i(j)\phi_i^\top(j)\phi_i(k)\xi_i(k)$. Just as in the LPV LS-SVM case, a kernel function K_i is used to define the inner product $\phi_i^\top(j), \phi_i(k)$. The kernel trick is applied in the same way and therefore

$$[\Xi^{(i)}]_{j,k} = x_i(j)K_i\big(p(j), p(k)\big)\xi_i(k). \tag{21}$$

Using expressions (18) and (17), the estimated nonlinearities can be expressed as

$$\hat{a}_i(\centerdot) = (\hat{\rho}_i^{\mathrm{IV}})^\top \phi_i(\centerdot) = \sum_{k=1}^{N} \hat{\alpha}_k \xi_i(k) K^i(p(k), \centerdot), \tag{22a}$$

$$\hat{b}_j(\centerdot) = (\hat{\rho}_{\tilde{j}}^{\mathrm{IV}})^\top \phi_{\tilde{j}}(\centerdot) = \sum_{k=1}^{N} \hat{\alpha}_k \xi_{\tilde{j}}(k) K^{\tilde{j}}(p(k), \centerdot). \tag{22b}$$

where $\tilde{j} = n_\mathrm{a} + 1 + j$ and $\hat{\alpha}_k$ is the k^{th} component of $\hat{\alpha}$.

6.2 Computation of the instrument

In practice, the estimate $\hat{y}$ of the noise-free output trajectory $\hat{y}$ needs to be computed such that it is really uncorrelated to the noise $v_\mathrm{o}(k)$. In case several realizations y_i of the output are available under the same excitation conditions, it is possible to use y_i as the output and y_j as $\hat{y}_i$ in the instrument for example. In the next example, a refined scheme such as in Young [1984] is chosen where $\hat{y}$ is computed using an auxiliary model which is iteratively refined. This choice is empirically justified here and can lead to an optimal parametric estimation of LPV models as shown in Laurain et al. [2010]. The algorithm used for the Refined IV-LS-SVM method can be expressed as in Algorithm 1:

Algorithm 1 Refined IV-LS-SVM

1: Initialization: Set $\tau = 0$ and $\hat{y} = y$.
2: **repeat**
3: Compute $\Xi^{(\tau)}$ as in (21) and $\hat{\alpha}^{(\tau)}$ using (16).
4: Compute the estimated nonlinear functions $\hat{a}_i^{(\tau)}(\centerdot)$ and $\hat{b}_j^{(\tau)}(\centerdot)$ using (22a-b).
5: Simulate the output of the obtained model $\hat{y}^{(\tau)}$ and use it to generate the instrument. Increment τ by 1.
6: **until** $||\alpha^{(\tau)} - \alpha^{(\tau-1)}||_2$ is under a certain threshold.

7. COMPARISON OF THE LS AND THE IV BASED APPROACHES

To show the benefit of the proposed refined IV-LS-SVM approach in the LPV setting, the example of Section 4 is reconsidered. Using the same data sets and the same value of the hyper-parameters $\gamma = 4200$ and $\sigma = 0.3$, the estimates have been recomputed by the refined IV-LS-SVM in the 50 Monte-Carlo runs. Note that using the same γ and σ gives an advantage to the LS-SVM as these parameters were optimized for the best performance of that approach.

Similarly to the previous case, the mean of the function estimates $\hat{a}_1(\centerdot)$ provided by the refined IV-LS-SVM over the 50 Monte-Carlo runs using the estimation data sets is depicted together with the true $a_1^\mathrm{o}(\centerdot)$ in Figure 2. It clearly appears that according to the theoretical results, the bias w.r.t. the noise has completely disappeared in comparison to Figure 1.

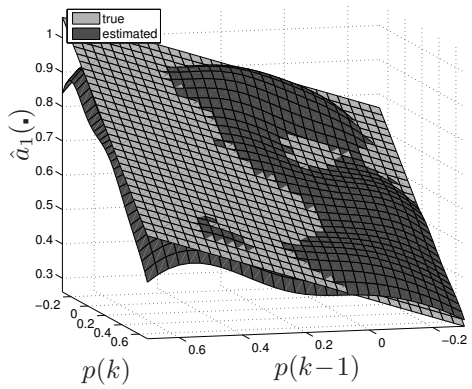

Fig. 2. Mean estimate of a_1^o by the refined IV-LS-SVM approach.

This result is not limited to the nonlinearity a_1^o. To show this, the statistical properties of the estimators are analyzed through the average bias B and the standard deviation S_td of the coefficient functions on $\mathbb{P}$. Using the notation $\bar{a}_i(\centerdot\centerdot) = \mathrm{Mean}_\mathrm{MC}\big(\hat{a}_i(\centerdot\centerdot)\big)$ for the mean and $\breve{a}_i(\centerdot\centerdot) = \mathrm{Std}_\mathrm{MC}\big(\hat{a}_i(\centerdot\centerdot)\big)$ for the standard deviation of the estimated functions over the Monte-Carlo runs, B and S_td are defined as:

$$B = \mathrm{Mean}_\mathbb{P}\big(\mathrm{abs}(a_i^\mathrm{o}(\centerdot\centerdot) - \bar{a}_i(\centerdot\centerdot))\big), \tag{23a}$$
$$S_\mathrm{td} = \mathrm{Mean}_\mathbb{P}\big(\breve{a}_i(\centerdot\centerdot)\big). \tag{23b}$$

Table 1 shows these values for both methods. It can be seen that the LS-SVM method has got the most bias for the nonlinearities a_1 and a_2 linked to y which is corrupted

Table 1. Mean and standard deviation of the estimated functions for both methods.

Coefficient		a_1	a_2	b_1	b_2
LS-SVM	B	0.1471	0.0838	0.0300	0.0433
	S_{td}	0.0896	0.0935	0.0941	0.0979
IV-LS-SVM	B	0.0291	0.0231	0.0302	0.0403
	S_{td}	0.1062	0.1026	0.1006	0.1043

Table 2. Average BFR on 600 data points.

Method	$\overline{\text{BFR}}$
LS-SVM	$73.76\% \pm 2.78\%$
RIV-LS-SVM	$76.48\% \pm 2.25\%$

by the noise. Regarding b_1 and b_2, both methods propose equivalent estimates. It can be further seen that, like in most other frameworks, the unbiasedness of the IV-based method is counterbalanced by an increased variance of the estimates. Nevertheless, the proposed refined scheme achieves a standard deviation close to the one obtained by the LS-SVM method.

Accuracy of the estimated models in terms of the average BFR ($\overline{\text{BFR}}$) on the validation data along the Monte-Carlo simulation is presented in Table 2. The IV-LS-SVM algorithm performs equivalently well compared to the LS-SVM method. Naturally, a performance drop occurs when the data set is further decreased as the impact of the increased variance becomes more dominant in such a scenario, reaching a critical data length, under which estimation by the IV scheme provides worse results than the LS-SVM scheme.

However in case the available data-record exceeds a critical length, the LS-SVM method cannot be directly applied on the whole data set. As the dimension N increases, the inversion of $\Omega \in N \times N$ becomes numerically intractable. Therefore, a commonly used approach is to split the whole data set into several reasonable sized data sets and to use the mean estimated model as the final result. Therefore Figure 3 displays the evolution of the fitting score of the averaged model as the number of data points increases. It can be seen that as expected, when the data length increases, the effect of the variance becomes negligible in comparison to the benefits brought by removing the bias.

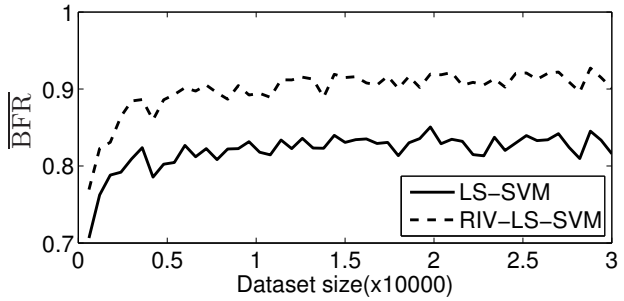

Fig. 3. Average BFR vs data length.

8. CONCLUSION

In this a paper, two recently introduced LS-SVM based methods have been compared when they are applied to the identification of polynomial LPV models in a structural learning context. It has been shown using a representative example that, for general noise structures, the bias caused by the use of an LS-based optimization scheme can lead to serious estimation error which can mislead controller synthesis. In order to cope with the noise modeling error, a recently introduced IV-based LS-SVM method has been applied to the LPV identification problem. It has been shown that the IV-based scheme guarantees unbiased estimates at the cost of an increased variance. Nevertheless, it has been demonstrated that the proposed solution results in a relatively small variance increase even for short data records. Furthermore, the IV-LS-SVM scheme preserves the computational efficiency of the original LS-SVM scheme, which makes it an attractive identification tool for the structural exploration of LPV models.

REFERENCES

B. Bamieh and L. Giarré. Identification of linear parameter varying models. *Int. J. Robust and Nonlinear Control*, 12:841–853, 2002.

V. Cerone and D. Regruto. Set-membership identification of LPV models with uncertain measurements of the time-varying parameter. In *Proc. of the 47th IEEE Conf. on Decision and Control*, pages 4491–4496, Cancun, Mexico, Dec. 2008.

M. Espinoza, J.A.K Suykens, and B De Moor. LS-SVM regression with autocorrelated errors. In *Proc. of the 14th IFAC Symposium on System Identification*, pages 582–587, Newcastle, Australia, Mar. 2006.

T. Falck, J. A. K Suykens, and B. De Moor. Linear parametric noise models for least squares support vector machines. In *Proc. of the 49th IEEE Conf. on Decision and Control*, pages 6389–6394, Atlanta, USA, Dec. 2010.

F. Giri and E.-W. Bai. *Block-oriented Nonlinear System Identification*. Lecture Notes in Control and Information Sciences. Berlin: Springer-Germany, 2010.

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.

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.

V. Laurain, W-X. Zheng, and R. Tóth. Introducing instrumental variables in the LS-SVM based identification framework. In *Proc. of the 50th IEEE Conf. on Decision and Control*, pages 3198–3203, Orlando, Florida, USA, Dec. 2011.

L. Ljung. *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 2nd edition, 1999.

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

T. Söderström and P. Stoica. *Instrumental Variable Methods for System Identification*. New York: Springer-Verlag, 1983.

J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. Singapore: World Scientific, 2002.

R. Tóth. *Modeling and Identification of Linear Parameter-Varying Systems*. Lecture Notes in Control and Information Sciences, Vol. 403. Berlin: Springer, 2010.

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, Dec. 2009.

R. Tóth, V. Laurain, W-X. Zheng, and K. Poolla. A support vector machine approach for LPV linear-regression models. In *Proc. of the 50th IEEE Conf. on Decision and Control*, pages 3192–3197, Orlando, Florida, USA, Dec. 2011.

V. Vapnik. *Statistical Learning Theory*. New York: Wiley-Interscience, 1998.

P. C. Young. *Recursive Estimation and Time-Series Analysis*. Berlin: Springer-Verlag, 1984.