Model structure learning: A support vector machine approach for LPV linear-regression models

2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Model Structure Learning: A Support Vector Machine Approach for LPV Linear-Regression Models

Roland Tóth, Vincent Laurain, Wei Xing Zheng and Kameshwar Poolla

***Abstract*— Accurate parametric identification of Linear Parameter-Varying (LPV) systems requires an optimal prior selection of a set of functional dependencies for the parametrization of the model coefficients. Inaccurate selection leads to structural bias while over-parametrization results in a variance increase of the estimates. This corresponds to the classical bias-variance trade-off, but with a significantly larger degree of freedom and sensitivity in the LPV case. Hence, it is attractive to estimate the underlying model structure of LPV systems based on measured data, *i.e.,* to learn the underlying dependencies of the model coefficients together with model orders etc. In this paper a Least-Squares Support Vector Machine (LS-SVM) approach is introduced which is capable of reconstructing the dependency structure for linear regression based LPV models even in case of rational dynamic dependency. The properties of the approach are analyzed in the prediction error setting and its performance is evaluated on representative examples.**

***Index Terms*— Linear parameter-varying; support vector machines; linear regression; ARX; identification, model structure selection.**

I. 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., [1]–[7]). In discrete-time, the most basic model structure in this context is the so-called *auto-regressive model with exogenous input* (ARX) which is often defined in the *single-input single-output* (SISO) case as

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

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 $\mathbb{P} \subseteq \mathbb{R}^{n_\mathrm{p}}$ and e is a white stochastic noise process. Furthermore (to keep the notation simple), the coefficient functions $a_i, b_j : \mathbb{P} \to \mathbb{R}$ have *static dependence* on p, *i.e.,* they only depend on the instantaneous value of $p(k)$.

It is a general mark of LPV models that the signal relations are linear just as in the *linear time-invariant* (LTI) case, but the model parameters are functions of the measurable time-varying signal p. Using scheduling variables as changing operating conditions or endogenous/free signals of the plant, the class of LPV systems can describe both nonlinear and time-varying phenomena.

Identification of the LPV-ARX model (1) can be addressed by either following a *local* or a *global* approach. In the local case, LTI snapshots of the model are identified for constant levels of p followed by an interpolation over $\mathbb{P}$, while by using a global approach, the coefficient functions of (1) are estimated based on a data record $\mathcal{D}_N = \{(u(k), p(k), y(k))\}_{k=1}^N$ with varying p. These approaches have their pros and cons, but in the global setting a compact estimation of (1) becomes available by realizing that (1) can be written as a linear regression. This is possible under the assumption that each function a_i and b_j can be decomposed in terms of a priori selected basis set $\psi_{ij} : \mathbb{P} \to \mathbb{R}$ as

$$a_i(p(k)) = \theta_{i0} + \theta_{i1}\psi_{i1}(p(k)) + \ldots + \theta_{i1}\psi_{is_i}(p(k)), \quad (2)$$

where $\theta_{ij} \in \mathbb{R}$ are the *unknown parameters* to be identified. In the ARX case, this delivers identification of (1) in terms of a simple *least-squares* (LS) estimate in the classical *prediction-error* (PE) setting [1]. Furthermore, this parametrization allows to formulate an LPV extension of the PE framework, where stochastic analysis of the estimation and general noise models can be addressed [3], [5].

Besides the classical questions of model structure selection in terms of model order (like n_a and n_b, input delay, etc.), the adequate selection of the basis set $\{\psi_{ij}\}$ has a paramount importance in this setting. To capture the unknown dependence of a_i and b_j on p, which can range from simple polynomial to complicated rational or even discontinuous functions, often a large set of basis functions is applied from which only a few might actually be needed for an accurate approximation. This means that most commonly, the LS estimation is faced with a seriously over-parametrized model structure where the underlying true parameter vector θ_o can be rather sparse. Hence the variance of the estimates can be seriously large even if the order of the actual model is low. Moreover, if there is no prior information about the nonlinearities of the system, the basis set $\{\psi_{ij}\}$ can be particularly inadequate, leading to a potential structural bias. While regularization of the variance and accurate identification of the support of θ_o w.r.t. a given parametrization can be efficiently achieved via sparse estimators, like the *nonnegative garrote* (NNG) [8], adequate selection/estimation of $\{\psi_{ij}\}$ based on data poses a much harder problem.

Additionally, realization theory of LPV models and LPV

This work was supported by a Research Grant of the Australian Research Council and by the Netherlands Organization for Scientific Research (Grant No. 680-50-0927).

R. Tóth is with the Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands, email: r.toth@tudelft.nl.

V. Laurain and W. X. Zheng are with the School of Computing and Mathematics, University of Western Sydney, Penrith NSW 2751, Australia, email: vlaurain@scm.uws.edu.au; w.zheng@uws.edu.au.

K. Poolla is with the Berkeley Center for Control and Identification, University of California at Berkeley, 5105 Etcheverry Hall, Berkeley, CA 94720-1740, USA, email: poolla@berkeley.edu.

modeling of nonlinear systems requires that the model coefficients not only depend on the instantaneous values of p (*static dependence*) but also on its time-shifted versions (*dynamic dependence*) [9]. Thus, estimating LPV-ARX models with dynamic dependence is often required for obtaining accurate models of the underlying system. Parametrization with such dependency seriously increases the dimension of the parametrization space and commonly renders the adequate selection of $\{\psi_{ij}\}$ based on heuristics to be infeasible in practice.

To obtain an efficient solution for this problem, a *support vector machine* (SVM) approach is introduced in this paper with the aim of providing "nonparametric" reconstruction of the dependency structure for linear regression based LPV models. SVM's have been originally developed as a class of *supervised learning* methods [10], [11] for efficient reconstruction of underlying functional relationships and structures in data. These approaches have also had a significant impact on *nonlinear block model* identification via various *least square-SVM* (LS-SVM) approaches [12]–[15]. In general, LS-SVM's are particular variations of the original SVM approach using an ℓ_2 loss function. Their main advantage is the uniqueness of the solution, obtained by solving a *linear problem* in a computationally efficient way. Hence our aim is to use the computational and reconstruction potential of LS-SVM's to provide an efficient solution in the LPV case.

It must be noted that in [4], a semi-parametric functional dependence estimation approach using *dispersion functions* was introduced. However, this approach, as we will see, is less effective in the bias-variance trade-off than the LS-SVM scheme to be proposed and hence it shows lower accuracy in the considered simulation study. Due to the underlying common *sum-of-norms* cost function and nonparametric nature of the estimation, the proposed LS-SVM scheme can be considered as a next step on this path towards an automated and effective model structure learning in the LPV setting.

The paper is organized as follows. In Section II, a short overview of the LPV-ARX model structure and its LS-based identification method is given, defining the problem setting. In Section III, such models are formulated in the LS-SVM setting and an algorithm is proposed for nonparametric identification. In Section IV, the properties of the proposed approach are analyzed and compared to sparse estimators and the dispersion approach of [4], while in Section V, the algorithm is validated on a Monte Carlo study and compared in performance to [4]. Finally, conclusions and future directions of the research are given in Section VI.

II. LPV IDENTIFICATION VIA ARX MODELS

In this paper, we focus on LPV systems with ARX type of IO representation defined in (1). Introduce

$$\left[\begin{array}{ccc}\phi_1 & \dots & \phi_{n_\mathrm{g}}\end{array}\right]^\top \triangleq \left[\begin{array}{ccccccc}a_1 & \dots & a_{n_\mathrm{a}} & b_0 & \dots & b_{n_\mathrm{b}}\end{array}\right]^\top,$$

with $n_\mathrm{g} = n_\mathrm{a} + n_\mathrm{b} + 1$, where each $\phi_i(\centerdot)$ is a real function with static dependence on p. It is assumed that these functions are non-singular on $\mathbb{P}$, thus the solutions of the system given by (1) are well-defined and the process part is completely characterized by $\{\phi_i(\centerdot)\}_{i=1}^{n_\mathrm{g}}$. As it is almost exclusively done in the LPV identification literature (see, *e.g.,* [1]–[3], [5], [16], [17]), assume that each ϕ_i is linearly parameterized as

$$\phi_i(\centerdot) = \theta_{i0} + \sum_{j=1}^{s_i} \theta_{ij}\psi_{ij}(\centerdot), \tag{3}$$

where $\{\theta_{ij}\}_{i=1,j=1}^{n_\mathrm{g},s_i}$ are unknown parameters and $\{\psi_{ij}\}_{i=1,j=1}^{n_\mathrm{g},s_i}$, with $s_i \in \mathbb{N}$, are functions chosen by the user. Denote $p(k)$ as p_k. In this case, (1) can be written as

$$y(k) = \theta^\top \varphi(k) + e(k), \tag{4}$$

where $\theta = \left[\begin{array}{cccccc}\theta_{1,0} & \dots & \theta_{1,s_1} & \theta_{2,0} & \dots \theta_{n_\mathrm{g},s_{n_\mathrm{g}}}\end{array}\right]^\top$ and

$$\begin{aligned}\varphi(k) = \Big[\; & -y(k-1) \quad -\psi_{11}(p_k)y(k-1) \quad \dots \\ & -\psi_{1s_1}(p_k)y(k-1) \quad \dots \quad -\psi_{n_\mathrm{a}s_{n_\mathrm{a}}}(p_k)y(k-n_\mathrm{a}) \\ & u(k) \quad \dots \quad \psi_{n_\mathrm{g}s_{n_\mathrm{g}}}(p_k)u(k-n_\mathrm{b}) \;\Big]^\top.\end{aligned}$$

Given a data set $\mathcal{D}_N = \{(u(k), p(k), y(k))\}_{k=1}^N$, the LS estimate for the linear regression model (4) is given by

$$\hat{\theta} = \arg\min_{\theta\in\mathbb{R}^n} \mathcal{V}(\theta, e), \tag{5}$$

where $n = \sum_{i=1}^{n_\mathrm{g}} 1 + s_i$ (according to (3)), and

$$\mathcal{V}(\theta, e) \triangleq \frac{1}{N}\|e(k)\|_{\ell_2}^2, \quad e(k) \triangleq y(k) - \theta^\top\varphi(k). \tag{6}$$

To guarantee a unique solution of (5), it is assumed that $\{\psi_{ij}\}_{i=1,j=1}^{n_\mathrm{g},s_i}$ are chosen such that (1) is globally identifiable (there exist no θ and θ', such that the 1-step ahead predictor resulting from (1) is not distinguishable for θ and θ') and that $\mathcal{D}_N$ provides a *persistently exciting* regressor in (4) (see [5], [18]). Note that identifiability in particular holds for (4) with $e \neq 0$ iff for each $i \in \{1, \dots, n_\mathrm{g}\}$, $\{\psi_{ij}\}_{j=1}^{s_i}$ corresponds to a set of linearly independent functions on $\mathbb{P}$. By organizing the data as

$$Y = \left[\begin{array}{cccc}y(1) & y(2) & \dots & y(N)\end{array}\right]^\top, \tag{7a}$$

$$\Phi = \left[\begin{array}{cccc}\varphi(1) & \varphi(2) & \dots & \varphi(N)\end{array}\right]^\top, \tag{7b}$$

the optimal solution to (5) can be expressed as

$$\hat{\theta}_N = \left(\Phi^\top\Phi\right)^{-1}\Phi^\top Y. \tag{8}$$

III. LS-SVM FOR LPV SYSTEMS

In this section, we show how the SVM approach can be formulated with respect to the estimation of (1) without specifying the underlying dependencies required to derive a simple LS solution. As we will see, without such a prior specification, the SVM is still capable of conducting consistent estimation of the underlying system and preserving the low computational need of the LS estimation.

A. LPV modeling in the SVM setting

In contrast with the standard LS setting introduced in Section II, the structural dependence of the coefficients ϕ_i on p is now assumed to be *a priori* unknown. Consequently, the parametrized model of system (1) is introduced as

$$\mathcal{M}_{\omega,\varphi}: \qquad y(k) = \sum_{i=1}^{n_\mathrm{g}} \omega_i^\top \phi_i(p_k)x_i(k) + e(k), \tag{9}$$

where each $\phi_i : \mathbb{R} \to \mathbb{R}^{n_{\mathrm{H}}}$ denotes an undefined, potentially infinite ($n_{\mathrm{H}} = \infty$) dimensional *feature map*, $\omega_i \in \mathbb{R}^{n_{\mathrm{H}}}$ is the i^{th} parameter vector and

$$x_i(k) = y(k-i), \quad i = 1, \ldots, n_{\mathrm{a}}, \tag{10a}$$

$$x_{n_{\mathrm{a}}+1+j}(k) = u(k-j), \quad j = 0, \ldots, n_{\mathrm{b}}. \tag{10b}$$

Additionally, introduce $\omega = \begin{bmatrix} \omega_1^\top & \ldots \omega_{n_{\mathrm{g}}}^\top \end{bmatrix}^\top \in \mathbb{R}^{n_{\mathrm{g}} n_{\mathrm{H}}}$ and

$$\varphi(k) = \begin{bmatrix} \phi_1^\top(p_k)x_1(k) & \ldots & \phi_{n_{\mathrm{g}}}^\top(p_k)x_{n_{\mathrm{g}}}(k) \end{bmatrix}^\top, \tag{11}$$

such that (9) can be rewritten in the regression form as

$$y(k) = \omega^\top \varphi(k) + e(k). \tag{12}$$

B. Ridge regression

To simplify the notation, let $\phi_i(k) = \phi_i(p_k)$. The LS-SVM approach aims at minimizing the cost function

$$\mathcal{J}(\omega, e) = \frac{1}{2}\sum_{i=1}^{n_{\mathrm{g}}} \omega_i^\top \omega_i + \frac{\gamma}{2}\sum_{k=1}^{N} e^2(t_k) = \frac{1}{2}\|\omega\|_{\ell_2}^2 + \frac{\gamma}{2}\|e(t_k)\|_{\ell_2}^2 \tag{13}$$

where the scalar $\gamma \in \mathbb{R}_0^+$ is the *regularization parameter*. Note that (13) is a so-called *sum-of-norms* criterion as it contains both the equation error term from (9) and a regularization term: the ℓ_2 cost of ω scaled by γ. This added regularization term, as we will see later, is used to address the estimation (learning) of the unknown dependencies.

Consider the model $\mathcal{M}_{\omega,\varphi}$ as in (9) whose estimation corresponds to the following optimization problem

$$\min_{w,b,e} \quad \mathcal{J}(\omega, e) = \frac{1}{2}\sum_{i=1}^{n_{\mathrm{g}}} \omega_i^\top \omega_i + \frac{\gamma}{2}\sum_{t=1}^{N} e(k)^2, \tag{14a}$$

$$\text{s.t.} \quad e(k) = y(k) - \sum_{i=1}^{n_{\mathrm{g}}} \omega_i^\top \phi_i(k) x_i(k). \tag{14b}$$

This constrained optimization problem is solved by constructing the *Lagrangian*:

$$\mathcal{L}(\omega, e, \alpha) = \mathcal{J}(\omega, e) - \sum_{k=1}^{N} \alpha_k \left(\sum_{i=1}^{n_{\mathrm{g}}} \omega_i^\top \phi_i(k) x_i(k) + e(k) - y(k) \right) \tag{15}$$

with $\alpha_k \in \mathbb{R}$ being the *Lagrangian multipliers*. The global optimum is obtained when

$$\frac{\partial \mathcal{L}}{\partial e} = 0 \quad \to \quad \alpha_k = \gamma e(k), \tag{16a}$$

$$\frac{\partial \mathcal{L}}{\partial \omega_i} = 0 \quad \to \quad \omega_i = \sum_{k=1}^{N} \alpha_k \phi_i(k) x_i(k), \tag{16b}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_k} = 0 \quad \to \quad e(k) = y(k) - \sum_{i=1}^{n_{\mathrm{g}}} \omega_i^\top \phi_i(k) x_i(k). \tag{16c}$$

Substituting (16a) and (16b) into (16c) leads to the following set of equations

$$y(k) = \sum_{i=1}^{n_{\mathrm{g}}} \underbrace{\left(\sum_{k=1}^{N} \alpha_k x_i(k) \phi_i^\top(k) \right)}_{\omega_i^\top} \phi_i(k) x_i(k) + \underbrace{\gamma^{-1}\alpha_k}_{e(k)} \tag{17}$$

for $k \in \{1, \ldots, N\}$. This is equivalent to

$$Y = \left(\Omega + \gamma^{-1} I_N\right) \alpha, \tag{18}$$

where $\alpha = [\alpha_1 \ \ldots \ \alpha_N]^\top \in \mathbb{R}^N$, and Ω is the so-called *Kernel matrix*, which is defined in this case as

$$[\Omega]_{j,k} = \sum_{i=1}^{n_{\mathrm{g}}} [\Omega^i]_{j,k} \tag{19}$$

with

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

Here K^i is a positive definite kernel function defining the inner products of $\phi_i^\top(j)\phi_i(k)$. Consequently, K^i defines Ω and hence characterizes the feature maps $\{\phi_i\}_{i=1}^{n_{\mathrm{g}}}$ in an efficient fashion. This allows to characterize a wide range of nonlinear dependencies as a linear combination of infinitely many functions ($n_{\mathrm{H}} = \infty$) defined through the choice of the particular inner product and a relatively low dimensional parameter α. Called the *kernel trick* [10], [11], this approach allows the identification of the coefficient functions a_i and b_j without explicitly defining the feature maps involved. Note that the kernel trick can be applied as a regularization approach in a wide range of optimization problems, like in [16], where it was used to regularize dimension explosion in LPV subspace approaches.

A typical type of kernel is, for example, the *Radial Basis Function* (RBF) kernel:

$$K^i(p_j, p_k) = \exp\left(-\frac{\|p_j - p_k\|_{\ell_2}^2}{\sigma_i^2} \right), \tag{20}$$

but other kernels, like *polynomial* kernels, can also be used. The choice of the kernel defines the class of dependencies that can be represented. By using a particular kernel, *i.e.,* defining Ω, the solution of (18) is given by

$$\alpha = \left(\Omega + \gamma^{-1} I_N\right)^{-1} Y. \tag{21}$$

Using α, *i.e.*, the minimizer of (14a-b), the model estimate is computed according to (16b). This gives that the estimated coefficient functions are obtained as

$$a_i(\centerdot) = \omega_i^\top \phi_i(\centerdot) = \sum_{k=1}^{N} \alpha_k x_i(k) K^i(p(k), \centerdot), \tag{22a}$$

$$b_j(\centerdot) = \omega_{\tilde{j}}^\top \phi_{\tilde{j}}(\centerdot) = \sum_{k=1}^{N} \alpha_k x_{\tilde{j}}(k) K^{\tilde{j}}(p(k), \centerdot), \tag{22b}$$

where $\tilde{j} = n_{\mathrm{a}} + 1 + j$. Note that the parameter vector ω is never accessible in the SVM framework, and only the combined estimation ${\omega_i}^\top \phi_i(\centerdot) = a_i(\centerdot)$ or $b_j(\centerdot) = \omega_{\tilde{j}}^\top \phi_{\tilde{j}}(\centerdot)$ is computable using the defined kernel functions.

IV. Properties and comparison

In this section, a brief overview of the major properties and advantages of the proposed LPV LS-SVM scheme is given and the approach is compared to the NNG method [8] and the dispersion function method [4].

A. Advantages of the semi-parametric formulation

The SVM scheme does not require explicit declaration of the feature maps ϕ_i or estimation of the high dimensional parameter vectors ω_i. Instead, it requires the declaration of the kernel functions K^i, $i = 1, \ldots, n_{\mathrm{g}}$, which explicitly define the set of nonlinear functions where the optimal estimate of the dependencies is searched for. For exact recovery of these dependencies, $\Pi_i = \mathrm{Span}(K^i(\mathtt{p}, \centerdot))$ for $\mathtt{p} \in \mathbb{P}$ should contain the corresponding dependency, *e.g.,* $a_i(\centerdot)$. In case of polynomial dependencies with maximum degree of d, it is sufficient to choose d^{th}-order polynomial kernels, however choosing an RBF kernel gives the representation possibility of a wide range of nonlinearities even including rational functions. Independent definitions of K^i also allow any *a priori* knowledge to be easily included in the model structure. We refer to [11] for further discussions on the adequate selection of kernels.

Regarding the over-parametrization based LS solution in Section II, which also involves a choice of the functions $\psi_{i,j}$, the LS-SVM scheme has a significant advantage: it only requires the estimation of α, whose size is equal to N, and this is independent of the set of nonlinearities chosen (even for infinite dimensional feature maps). Consequently, it avoids using ultra-large scale over-parametrization by simply adopting an RBF kernel. Therefore, it represents a highly attractive bias-variance trade-off. Regarding consistency properties of the LS-SVM estimation, we refer to the classical results in [10], [11].

Recently an instrumental variable extension of the discussed LS-SVM scheme has been derived in [19], which allows the consistent estimation of nonlinear regression models in case the noise involved is not white. This approach can be extended to LPV-IO models with general noise structure like *Box-Jenkins* (see [3], [5]), which gives a wide applicability of the proposed LPV LS-SVM approach in practice. Furthermore, extension of the method to *multiple-input multiple-output* (MIMO) models follows easily just like in the LS case and due to the flexibility of the definitions of the kernel functions, recovery of complex dynamic dependencies over p can also be ensured.

B. Comparison to sparse estimators

The NNG method, proposed in the LPV case in [8], is based on the over-parametrized regression form (4), thus it requires the priori selection of $\{\psi_{i,j}\}$. However, besides the minimization of $\|e(k)\|_{\ell_2}^2$ it also aims to minimize $\|\theta\|_{\ell_0}$. The latter means that it tries to shrink the support of $\hat{\theta}$ to the most necessary basis functions associated parameters. This is done by using weighting of a given LS solution $\hat{\theta}$ of (4) regularized through a parameter $\lambda > 0$. As λ increases, the weights of the less important parameters shrink, and finally end up exactly at zero. This results in less complex model estimates, as long as the overall fit of the model estimate on the available (validation) data is still acceptable. An efficient way to implement this strategy is to use a path following parametric estimation, which calculates a piecewise affine solution path for λ [20]. However, such approaches are computationally more demanding than the LS-SVM scheme. Other sparse estimators, like the Lasso approach, implement the above described strategy by minimizing the combined objectives of $\|e(k)\|_{\ell_2}^2$ and $\|\theta\|_{\ell_0}$. It is immediate that due to the fact that these estimators must operate on a relatively large parametrization space, the proposed LS-SVM scheme can be considered more attractive. Furthermore, if $N \to \infty$, sparse estimators can consistently estimate the support of the true θ_{o}, which corresponds to the data-generating system in the model set, but are inconsistent in the ℓ_2 sense (see [21]). The latter is a particular drawback in comparison to the LS-SVM.

C. Comparison to the dispersion approach

The dispersion approach proposed in [4] can be considered to be similar to the LS-SVM method as it is also a semi-parametric approach with a required number of estimated parameters being $n_{\mathrm{g}} \times N$, it uses a sum-of-norms type of cost function with a regularization parameter and it does not require priori knowledge of the underlying dependencies (not even a choice of kernels). However, the larger number of parameters to be estimated ($n_{\mathrm{g}} \times N > N$) implies that the achievable bias-variance trade-off by the dispersion approach is inherently worse than in the SVM case. This claim is also supported by an experimental study in Section V. Furthermore, the dispersion method corresponds to a quadratic optimization problem which can be considered to be computationally more demanding than the analytical SVM solution. Nonetheless, due to the availability of efficient interior point methods, the difference between them in this aspect is often negligible in practice. Additionally, the flexible definition of the kernels in the SVM case allows to include prior information about the expected dependencies (if available) for more accurate results, unlike in the dispersion case.

V. Illustrative Examples

To assess the performance of the presented algorithm, two examples are presented in this section. The first example is borrowed from [4], which is used to compare the proposed LPV LS-SVM method to the dispersion approach of [4]. The second example is used to demonstrate the statistical properties and reconstruction capabilities of the LS-SVM algorithm on a representative Monte Carlo simulation.

A. Example 1: LPV-SVM vs. the dispersion approach

In the example of [4], the considered LPV data-generating system is given as:

$$y(k) = \sum_{i=0}^{2} b_i(p_{k-i})u(k-i) + e_{\mathrm{o}}(k), \tag{23}$$

with $\mathbb{P} = [-1, 1]$, e_{o} being a zero mean stochastic noise process and

$$b_0(p_k) = -\exp(-p_k), \qquad b_1(p_{k-1}) = 1 + p_{k-1},$$
$$b_2(p_{k-2}) = \tan^{-1}(p_{k-2}).$$

Note that this IO representation has a nonlinear dynamic dependence on p and it is in a so-called *finite impulse response* (FIR) form (a special case of ARX). A model of (23) can be formulated in the proposed SVM setting as

$$y(k) = \sum_{i=0}^{2} \omega_i^\top \phi_i(p_{k-i}) u(k-i) + e(k). \qquad (24)$$

Note that dynamic dependence of the feature maps in (24) does not impose any difficulty in formulating the ridge regression. To be able to compare the results, the same conditions of excitation and measurements as in [4] are used during simulations of (23). A data set $\mathcal{D}_N$ with $N = 400$ is generated by (23) using $u(k) = \sin(\frac{\pi}{2}k)$, $p(k) = \sin(0.25k)$ and $e_o(k) \equiv 0$ (noise-free measurement). Note that in $\mathcal{D}_N$, $u(k) = -u(k-2)$ for all k.

On the gathered data set $\mathcal{D}_N$, the proposed LS-SVM approach has been applied and the obtained results with respect to the estimation of the underlying coefficient functions are shown in Figure 1. To characterize the nonlinearities in this system, RBF kernels have been used for K, K^2 and K^3 with $\sigma_1 = \sigma_2 = \sigma_3 = 0.7$. Based on trial-and-error, the regularization parameter has been tuned to $\gamma = 500$. It is important to mention that in this noise-free case, the choice of these parameters is not very critical. Their tuning is consequently not necessary to demonstrate the advantageous properties of the proposed method in this example. As in the data set, $u(k) = -u(k-2)$ for all k, b_0 and b_2 are not uniquely identifiable: any pair of functions $\{b_0(p_k)+f, b_2(p_{k-2})+f\}$ produces the same output response for any arbitrary $f \in \mathbb{R}$ under the given excitation. This fact results in a pure constant bias for b_0 and b_2 over $\mathbb{P}$ (with $f = 0.81$ using the given parameter settings), and the bias is clearly visible in the results of Figure 1 given by the dashed lines. By adding an extra constraint to (14b) for centering the estimated coefficient functions, this bias can be effectively eliminated and the estimated coefficient functions (dashed-doted lines in Figure 1) show a perfect fit over $\mathbb{P}$.

Comparing the performance of the LS-SVM to the results of [4], the dispersion approach results in a much larger error for b_0 and b_2 for low values of p (see Fig. 1). This can be the effect of the better bias-variance trade-off when using the LS-SVM approach for data sets of such size. As stated before, a particular advantage of the LS-SVM over the dispersion approach is the smaller number of estimated parameters.

B. Example 2: LPV-ARX model

To assess the stochastic performance, as a next example, the proposed LPV LS-SVM approach is tested on an LPV data-generating system in an ARX form under rather severe noise conditions. In this case, the considered data-generating system is described by

$$y(k) + a_1(p_k) y(k-1) = \sum_{i=0}^{1} b_i(p_k) u(k-i) + e_o(k), \quad (25)$$

with $\mathbb{P} = [-1, 1]$, e_o being a zero mean stochastic noise process and

$$a_1(p_k) = 0.1 \cdot \frac{\sin(\pi^2 p_k)}{\pi^2 p_k},$$

$$b_0(p_k) = \begin{cases} +0.5 & \text{if } p_k > 0.5 \\ p_k & \text{if } -0.5 \le p_k \le 0.5 \\ -0.5 & \text{if} p_k < -0.5 \end{cases}$$

$$b_1(p_k) = -0.2 \cdot p(t_k)^2.$$

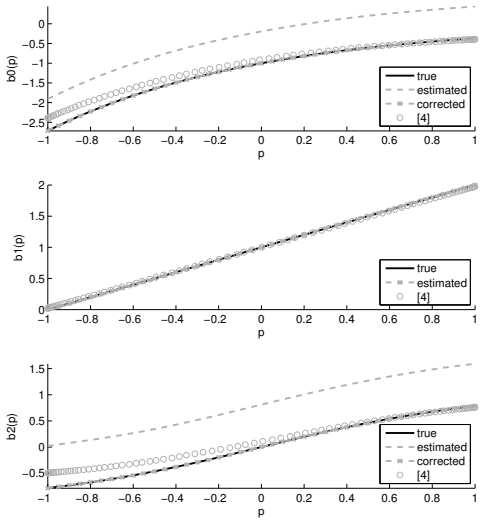

Fig. 1. Estimation results of the coefficient functions in Example V-A by the proposed LPV-SVM approach, including the results before and after offset correction along with the results from [4].

TABLE I

MEAN AND STANDARD DEVIATION OF THE BFR ON VALIDATION DATA.

	Mean	STD
LS-SVM	95.22	0.005

In this case, the IO representation of the system has only static dependence on p, but the nonlinearities involved in a_1 and b_0 are difficult to approximate (especially if using a polynomial parametrization as it is normally presented in the literature of regression based methods).

The identification problem of (25) is formulated in the proposed LS-SVM setting by considering the following model structure

$$y(k) = \sum_{i=0}^{3} \omega_i^\top \phi_i(p_{k-i}) x_i(k) + e(k), \qquad (26)$$

where $x_1(k) = y(k-1)$ and $x_2(k) = u(k)$, $x_3(k) = u(k-1)$. To provide an informative data set $\mathcal{D}_N$ for identification, u is taken as a zero-mean white noise process with a uniform distribution $\mathcal{U}(-1, 1)$ and with length $N = 1500$. Furthermore, $e_o(k)$ is assumed to have a Gaussian distribution $\mathcal{N}(0, \sigma_{e_o}^2)$ with $\sigma_{e_o} > 0$. To investigate the performance under fairly severe noise conditions, the *signal-to-noise ratio* (SNR) is set as $\text{SNR} = 10 \log \frac{P_\chi}{P_{\chi - y}} = 10\text{dB}$, where P_χ is the average power of signal χ, which is the deterministic component of y (noise-free output of (25)).

For numerical illustration, a Monte-Carlo simulation of $N_{\text{MC}} = 100$ runs has been accomplished with new realization of the noise and input in each run. Using the gathered data sets, the LPV LS-SVM approach has been applied to estimate (26) with RBF kernels for K^1, K^2 and K^3 having $\sigma_1 = \sigma_2 = \sigma_3 = 0.7$. The regularization parameter, based on trial-and-error, has been tuned to $\gamma = 10^4$. The estimation results for a_1, b_0 and b_1 are displayed in Figure 2 in terms of mean and standard deviation of the estimates over the 100 Monte Carlo runs. It is remarkable that without using any prior information about the system (except the continuity of the dependencies on $\mathbb{P}$ implied by the RBF kernels),

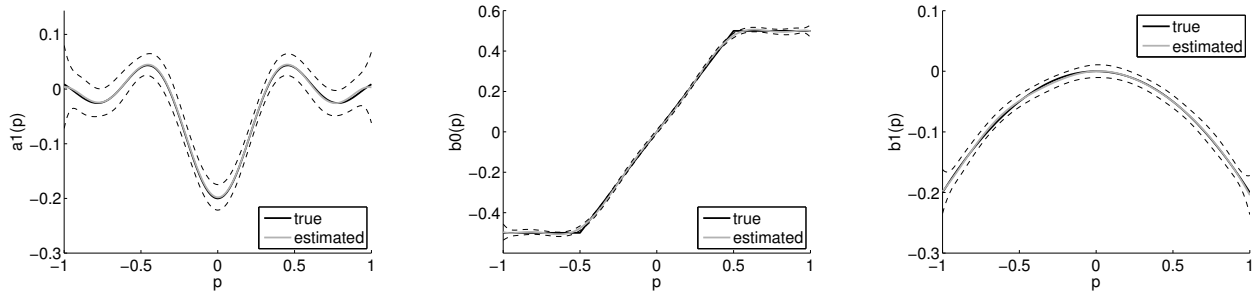

Fig. 2. Estimation results of the coefficient functions in Example V-B by the proposed LPV-SVM approach. The true nonlinear functions are given with solid black together with the mean estimate (solid grey) and $+/-$ standard deviation (dashed black) computed over 100 Monte-Carlo runs.

the considered nonlinear functions are estimated without a significant bias and the variance is also small compared to the heavy noise conditions. To quantify the model quality, let us consider the *fitness score* or *Best Fit Rate* (BFR):

$$\mathrm{BFR} = 100\% \cdot \max\left(1 - \frac{\|\chi(k) - \hat{\chi}(k)\|_{\ell_2}}{\|\chi(k) - \bar{\chi}\|_{\ell_2}}, 0\right), \quad (27)$$

where $\bar{\chi}$ is the mean of χ. The mean and the standard deviation of the BFR with respect to the model estimates are computed on a validation data set and displayed in Table I. From these measures, it follows that the proposed approach provides a quite accurate estimation approach for this non-trivial LPV model. Furthermore its computational load is relatively low and the method does not need any prior structural information about the dependencies.

VI. Conclusion

In this a paper a semi-parametric identification approach based on *least-squares support vector machines* (LS-SVM), has been introduced for LPV regression models. In contrast to the currently used over-parametrization based techniques for *least squares* (LS) estimation of such models, the proposed approach is capable of providing consistent estimates without prior information on the parametrization of the underlying coefficient dependencies. This is not only favorable in the common practical situation of unknown structural dependency of the model, but also lowers the variance of the estimates due to the efficient dual optimization scheme involved. Furthermore, the computational load of the method is relatively low. It has also been shown on a relevant example that the LS-SVM approach achieves a better performance with respect to the dispersion function approach proposed in the literature for LPV semi-parametric identification. Besides, the proposed approach is capable of capturing difficult nonlinear dependencies. An interesting topic for future research is to test the performance of the LS-SVM approach on real applications and also to capture hysteresis type of dependencies using multidimensional kernels with dynamic dependency.

References

[1] B. Bamieh and L. Giarré, "Identification of linear parameter varying models," *Int. Journal of Robust and Nonlinear Control*, vol. 12, pp. 841–853, 2002.

[2] X. Wei, "Advanced LPV techniques for diesel engines," Ph.D. dissertation, Johannes Kepler University, Linz, 2006.

[3] V. Laurain, M. Gilson, R. Tóth, and H. Garnier, "Refined instrumental variable methods for identification of LPV Box-Jenkins models," *Automatica*, vol. 46, no. 6, pp. 959–967, 2010.

[4] 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*, San Antonio, Texas, USA, Sept. 2008, pp. 846–851.

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

[6] S. K. Wattamwar, S. Weiland, and T. Backx, "Identification of low-order parameter-varying models for large-scale systems," *Journal of Process Control*, vol. 20, no. 2, pp. 158–172, 2010.

[7] X. Jin, B. Huang, and D. S. Shook, "Multiple model LPV approach to nonlinear process identification with EM algorithm," *Journal of Process Control*, vol. 21, no. 1, pp. 182–193, 2011.

[8] 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*, Shanghai, China, Dec. 2009, pp. 7406–7411.

[9] R. Tóth, H. Abbas, and W. Werner, "On the state-space realization of LPV input-output models: Practical approaches," *in print, IEEE Trans. on Control Systems Technology*, 2011.

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

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

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

[13] T. Falck, K. Pelckmans, J. Suykens, and B. De Moor, "Identification of wiener-hammerstein systems using LS-SVMs," in *15th IFAC symposium on System Identification*, Saint Malo, France, July 2009.

[14] I. Goethals, K. Pelckmans, J. Suykens, and B. De Moor, "Identification of MIMO Hammerstein models using least squares support vector machines," *Automatica*, vol. 41, no. 7, pp. 1263–1272, 2005.

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

[16] V. Verdult and M. Verhaegen, "Kernel methods for subspace identification of multivariable LPV and bilinear systems," *Automatica*, vol. 41, no. 9, pp. 1557–1565, 2005.

[17] R. Tóth, P. S. C. Heuberger, and P. M. J. Van den Hof, "Asymptotically optimal orthonormal basis functions for LPV system identification," *Automatica*, vol. 45, no. 6, pp. 1359–1370, 2009.

[18] A. G. Dankers, R. Tóth, P. S. C. Heuberger, X. Bombois, and P. M. J. Van den Hof, "Identifiability and the informativity of data sets for LPV-ARX identification," in *Proc. of the 50th IEEE Conf. on Decision and Control*, Orlando, Florida, USA, Dec. 2011.

[19] V. Laurain, W. 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*, Orlando, Florida, USA, Dec. 2011.

[20] C. Lyzell, J. Roll, and L. Ljung, "The use of nonnegative garrote for order selection of ARX models," in *Proc. of the 45th IEEE Conf. on Decision and Control*, Cancun, Mexico, Dec. 2008, pp. 1974–1979.

[21] H. Leeb and B. Pötscher, "Sparse estimators and the oracle property, or the return of hodges' estimator," *Journal of Econometrics*, vol. 142, no. 1, pp. 201–211, 2008.