53rd IEEE Conference on Decision and Control
December 15-17, 2014. Los Angeles, California, USA

A Bayesian Approach for Estimation of Linear-Regression LPV Models

Arash Golabi[1], Nader Meskin[1], Roland Tóth[2] and Javad Mohammadpour[3]

***Abstract*— In this paper, a Bayesian framework for identification of linear parameter-varying (LPV) models with finite impulse response (FIR) dynamic structure is introduced, in which the dependency structure of LPV system on the scheduling variables is identified based on a Gaussian Process (GP) formulation. Using this approach, a GP is employed to describe the distribution of the coefficient functions, that are dependent on the scheduling variables, in LPV linear-regression models. First, a prior distribution over the nonlinear functions representing the unknown coefficient dependencies of the model to be estimated is defined; then, a posterior distribution of these functions is obtained given measured data. The mean value of the posterior distribution is used to provide a model estimate. The approach is formulated with both static and dynamic dependency of the coefficient functions on the scheduling variables. The properties and performance of the proposed method are evaluated using illustrative examples.**

***Index Terms*— Linear parameter-varying systems; system identification; Bayesian method; Gaussian process; linear regression model**

I. Introduction

During the last decade, identification of *linear parameter-varying* (LPV) systems has attracted considerable attention (see, e.g., [1]–[7]). In general, different modeling frameworks such as input-output [8], state space [9] and orthonormal basis functions representations [10] can be considered for the identification of an LPV system. Estimation of LPV models in an input-output (I/O) setting has received a great deal of attention in the LPV identification literature [8], [11], [12]. Most of the methods for LPV I/O model identification have been developed in the discrete-time domain under a linear regression form with static dependency on the scheduling variables. The most basic model structure is an auto-regressive model with exogenous input (ARX) which is described as

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

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 $\mathbb{P} \subseteq \mathbb{R}^{n_p}$ and e is an *independent and identically distributed* (i.i.d) white stochastic noise process. The process (1) is fully characterized by the nonlinear coefficients $\{a_i\}_{i=1}^{n_\mathrm{a}}$ and $\{b_j\}_{j=0}^{n_\mathrm{b}}$. However, in practical cases, these coefficients are not fully known and often their estimation is required from measured data. The standard solution to overcome this problem is to parameterize these functions using an affine combination of *a priori* known basis functions. In this case, each function $\{a_i\}$ and $\{b_j\}$ is assumed to be represented as

$$a_i(p_k) = \alpha_{i0} + \alpha_{i1} f_{i1}(p(k)) + ... + \alpha_{iN} f_{iN}(p(k)), \tag{2}$$

where $f_{ij} : \mathbb{P} \to \mathbb{R}$ are prior fixed basis functions and $\alpha_{ij} \in \mathbb{R},\ j = 1, ..., N$ are the *unknown parameters* to be estimated. Inaccurate selection of the basis functions leads to structural bias while the over-parameterization may increase the variance of the estimated model. Moreover, LPV modeling of a nonlinear system requires that the nonlinear coefficients depend on the time-shifted version of scheduling variables p (dynamic dependence) while parameterization with such dependency significantly increases the dimension of the parameterization space. Hence, it is appealing to estimate and learn the underlying dependencies of the LPV model coefficients based on the measured data as well. Some recent works have used non-parametric approaches to obtain an efficient solution for this problem [5], [13]–[15].

In [13], a *Least Square Support Vector Machine* (LS-SVM) approach is used to reconstruct the dependency structure for linear regression-based LPV models. The SVM approaches represent a class of supervised learning methods for efficient reconstruction of underlying functional relationships and structures in data where the LS-SVM is a particular subclass formulated for regression problems [16]. In the present paper, a different learning algorithm called the *Bayesian method* is developed for structured identification of a *finite impulse response* (FIR) of an LPV system. Bayesian methods for system identification have been a subject of renewed interest in the last few years [17]–[21]. The Bayesian formulation for system identification problems has become popular mainly due to the introduction of a family of *a priori* descriptions (kernels) which encode structural properties of dynamical systems such as stability [22]–[24]. In this context, in [18] the impulse response is modeled for a linear system as the realization of a *Gaussian process* (GP) whose statistics include information not only on smoothness, but also on BIBO-stability. In this paper, we take advantage of Bayesian method to reconstruct the dependency structure of an FIR model of an LPV system. Unlike classical learning algorithms, Bayesian algorithms do not attempt to identify the best model

This publication was made possible by the NPRP grant (No. 5-574-2-233) from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

[1]Department of Electrical Engineering, College of Engineering, Qatar University, P.O. Box 2713, Doha, Qatar (e-mail: arash.golabi@qu.edu.qa, nader.meskin@qu.edu.qa).

[2]Control Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, P.O. Box 543, 5600 MB Eindhoven, The Netherlands (e-mail: r.toth@tue.nl).

[3]Complex Systems Controls Lab, College of Engineering, The University of Georgia, Athens, GA 30602, USA (e-mail: javadm@uga.edu).

from data or make the best guess in terms of predictions for new test inputs. Instead, they compute a posterior distribution over models or compute posterior predictive distributions for new test inputs [25]. These distributions provide a useful way to quantify the uncertainties of the estimated model, which can be exploited to make more robust predictions on new test points.

This paper is organized as follows. In Section II, the problem of LPV model identification based on the Bayesian approach is considered and the main derivations for reconstructing the dependency structure of LPV systems are given. In Section III, properties and performance of the proposed approach are studied through numerical simulations. The concluding remarks are given in Section IV.

II. LPV model Identification Based on a Bayesian Approach

In this section, the Bayesian formulation is used to reconstruct the dependency structure of a given LPV system. Here, we consider an LPV FIR model (a special case of ARX model (1)), i.e., $a_i = 0, i = 1, ..., n_\mathrm{a}$. The model of the system can now be represented as

$$y(k) = \sum_{i=1}^{n_\mathrm{g}} g_i(p(k))\phi_i(k) + e(k), \tag{3}$$

where $n_\mathrm{g} = n_\mathrm{b} + 1$ and

$$\begin{bmatrix} g_1 & ... & g_{n_\mathrm{g}} \end{bmatrix} = \begin{bmatrix} b_0 & ... & b_{n_\mathrm{b}} \end{bmatrix},$$
$$\phi_{j+1}(k) = u(k-j), \quad j = 0, ..., n_\mathrm{b}. \tag{4}$$

In the context of the Bayesian approach, GP is used to describe a distribution over nonlinear functions representing the coefficients of an LPV model to be estimated and the posterior distributions of these functions are obtained, given the observations. Formally, a GP is a stochastic process such that any finite subcollection of random variables has a multivariate Gaussian distribution [25]. Consider a standard Gaussian process regression model as [25]

$$y(k) = \mathcal{F}(x(k)) + e(k), \tag{5}$$

where $x(k)$ is input vector, $\mathcal{F}$ is the function value, $y(k)$ is the observed target value and $e(k)$ is an i.i.d. noise process with $e(k) \sim \mathcal{N}(0, \sigma^2)$. It is also assumed that the function $\mathcal{F}(.)$ is a particular realization of a multivariate process with a zero-mean Gaussian distribution as

$$\mathcal{F}(.) \sim \mathcal{GP}(0, \mathcal{K}(.,.)), \tag{6}$$

with a symmetric positive semidefinite covariance function $\mathcal{K}(.,.)$. Note that $\mathcal{GP}(.,.)$ denotes distribution as a Gaussian process. The joint distribution of the observed target values and the function value at a test point x^* under the assumed prior can be written as

$$\begin{bmatrix} y \\ \mathcal{F}^* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \mathcal{K}(X,X) + \sigma^2 I & \mathcal{K}(X,x^*) \\ \mathcal{K}(x^*,X) & \mathcal{K}(x^*,x^*) \end{bmatrix}\right), \tag{7}$$

where X is a vector of input data. Note that if there are n training points and one test point x^*, then $\mathcal{K}(X, x^*)$ denotes an $n \times 1$ matrix. The posterior distribution for Gaussian process regression is as follows [25]

$$\mathbb{E}[\mathcal{F}^*] = \bar{\mathcal{F}}^* = {\kappa^*}^\top (\mathcal{K}(X,X) + \sigma^2 I)^{-1} y, \tag{8}$$
$$\mathrm{cov}[\mathcal{F}^*] = \mathcal{K}(x^*,x^*) - {\kappa^*}^\top (\mathcal{K}(X,X) + \sigma^2 I)^{-1} \kappa^*, \tag{9}$$

where $\kappa^* = \mathcal{K}(x^*, X)$.

Next, the Bayesian formulation is used for the identification of LPV FIR models. Equation (3) can be rewritten as

$$y(k) = \sum_{i=1}^{n_\mathrm{g}} f_i(p_k, \phi_i(k)) + e(k), \tag{10}$$

where $f_i(p_k, \phi_i(k)) = g_i(p_k)\phi_i(k)$ and for notation simplicity, $p(k)$ is denoted as p_k. For each $g_i(.)$, the mean and covariance for all $p_j, p_k \in \mathbb{P}$ are defined as

$$\mathbb{E}[g_i(p_j)] = \bar{g}_i(p_j), \tag{11}$$
$$\mathrm{cov}[g_i(p_k), g_i(p_j)] = \mathbb{E}[g_i(p_k) g_i(p_j)] = \mathcal{K}^i(p_k, p_j), \tag{12}$$

where it is assumed that a data set is given, called the estimation data set $\mathcal{D}_{set} = \{p(k), u(k), y(k)\}_{k=1}^N$. Here $\mathcal{K}^i$ is a positive semidefinite kernel function. The Gaussian kernel is a commonly used one in non-parametric estimation and is defined as [26]

$$\mathcal{K}^i(p_j, p_k) = \lambda_i \exp\left(-\frac{||p_j - p_k||^2}{\sigma_i^2}\right), \tag{13}$$

which is also referred to as *Radial Basis Function* (RBF) or Gaussian kernel.

Note that the kernel function in (13) is written as a function of the scheduling variable p. For this particular Gaussian function, the kernel value is almost λ_i between nonlinear coefficients whose corresponding scheduling parameters are very close ($\mathcal{K}^i(p_j, p_k) \approx \lambda_i$ where p_j is very close to p_k) and decreases as the distance in the scheduling variable space increases. Let us define p^* as a test point and $\mathcal{P} = \{p(1), ..., p(N)\}$, $\mathcal{Y} = \{y(1), ..., y(N)\}$ and $\mathcal{U} = \{u(1), ..., u(N)\}$. Then, from equations (8) and (9) one can obtain the posterior distribution for each g_i as follows

$$\mathbb{E}\left[g_i(p^*)|p^*, \mathcal{P}, \mathcal{Y}, \mathcal{U}\right] = \sum_{k=1}^{N} \alpha_k \phi_i(k) \mathcal{K}^i(p(k), p^*), \tag{14}$$
$$\mathrm{cov}\left[g_i(p^*)|p^*, \mathcal{P}, \mathcal{Y}, \mathcal{U}\right] = \mathcal{K}^i(p^*, p^*) - \sum_{k=1}^{N} \beta_k^i(p^*) \phi_i(k) \mathcal{K}^i(p(k), p^*), \tag{15}$$

where α_k and β_k^i are the elements of N-dimensional vectors α and β^i respectively, defined by

$$\alpha = (K + \sigma^2 I)^{-1} Y, \tag{16}$$

and

$$Y = \begin{bmatrix} y(1) & ... & y(N) \end{bmatrix}^\top$$
$$\beta^i(p^*) = (K + \sigma^2 I)^{-1} k_i^*(p^*),$$

$$k_i^*(p^*) = \begin{bmatrix} \phi_i(k)\mathcal{K}^i(p_1, p^*) \\ . \\ . \\ . \\ \phi_i(k)\mathcal{K}^i(p_N, p^*) \end{bmatrix},$$

$$[K]_{j,k} = \sum_{i=1}^{n_g} [K^i]_{j,k},$$

$$[K^i]_{j,k} = \phi_i(j)\left(\mathcal{K}^i(p_j, p_k)\right)\phi_i(k).$$

Finally, the one-step-ahead prediction of the output of time instant k w.r.t data set $\hat{\mathcal{D}}_N = \{(\hat{u}(k), \hat{p}(k))\}_{k=1}^N$ (different from the estimator data reference) can be obtained as

$$\hat{y}(k) = \sum_{i=1}^{n_g} g_i(\hat{p}(k))\hat{u}(k+1-i). \tag{17}$$

It should be noted that in principle, LS-SVM approach [13] and the proposed Bayesian method are similar. The main difference is the availability of the estimation of the *covariance* function (15). The Bayesian method computes a posterior distribution over models, which is the main difference between the Bayesian method and the LS-SVM approach. Computation of the covariance function provides a useful way to quantify the uncertainties of the estimated model of the underlying dependency structure in the identification of an LPV model.

A. Bayesian Identification for LPV Systems with Dynamic Dependency

In this section, the identification problem of LPV FIR models where the nonlinear coefficient functions have dynamic dependency on the scheduling variables p is considered. In this case, Equation (3) with dynamic dependency can be substituted by

$$y(k) \;=\; \sum_{i=1}^{n_g} g_i(p_k, p_{k-1}, ..., p_{k-n})\phi_i(k) \;+\; e(k). \tag{18}$$

Here, it is assumed that the value of n (for physical systems, n cannot be greater than n_b) to learn the underlying dependencies of the model coefficients is available. Following a similar approach to the one proposed in the previous section, the main result can be written as

$$\mathbb{E}\left[g_i(\mathbf{p}^*)|\mathbf{p}^*, \mathcal{P}, \mathcal{Y}, \mathcal{U}\right] = \sum_{k=1}^{N} \alpha_k \phi_i(k)\mathcal{K}^i(\mathbf{p_{k,n}}, \mathbf{p}^*), \tag{19}$$

$$\begin{aligned}\mathrm{cov}\left[g_i(\mathbf{p}^*)|\mathbf{p}^*, \mathcal{P}, \mathcal{Y}, \mathcal{U}\right] &= \mathcal{K}^i(\mathbf{p}^*, \mathbf{p}^*) - \\ &\sum_{k=1}^{N} \beta_k^i(\mathbf{p}^*)\phi_i(k)\mathcal{K}^i(\mathbf{p_{k,n}}, \mathbf{p}^*)\end{aligned} \tag{20}$$

where

$$\mathcal{K}^i(\mathbf{p_{j,n}}, \mathbf{p_{k,n}}) = \lambda_i \exp\left(-\frac{||\mathbf{p_{j,n}} - \mathbf{p_{k,n}}||_2^2}{\sigma_i^2}\right).$$

and $\mathbf{p_{k,n}} = [p_k, p_{k-1}, ..., p_{k-n}]^\top$ and $||.||_2$ denotes the vector 2-norm such that $||v||_2 = \sqrt{v^\top v}$.

B. Experimental Design

In this section, we explain one approach to choose the input, scheduling variables and size of the training data for identification of a given LPV system for the presented approach. Consider a system defined in an LPV FIR form with order of $n_b + 1$ as

$$y(k) = \sum_{j=0}^{n_b} b_j(p_k)u(k-j) + e(k). \tag{21}$$

In this case, one choice for the input signal is a periodic function with $T = n_b + 1$ as

$$u(k) = \begin{cases} 1 & k = 1 \\ 0 & k = 2, ..., n_b + 1 \end{cases} \tag{22}$$

so that one can separate the effect of each coefficient $b_j(p_k)$ on the output y and hence the dependency structure can be obtained easily as

$$y(1) = b_0(p_1),\; y(2) = b_1(p_2), ..., y(n_b + 1) = b_{n_b}(p_{n_b+1}). \tag{23}$$

However, it should be noted that with the choice of the input signal as in (22), value of $b_0(p_k)$ for $k = 2, ..., n_b + 1$ cannot be identified. To overcome this problem, a periodic scheduling parameter with period of $T = n'$ should be considered such that remainder of division of n' by $(n_b + 1)$ is 1 and the training data with size of $N > n' \times (n_b + 1)$ is needed. Therefore, all values of $b_j(p_k)$ for $k = 1, ..., n'$ can be kept and observed in the output signal y.

III. Simulation study

To illustrate the performance of the presented method, we consider three examples. The first example is used to estimate nonlinear coefficient functions of a given LPV system. In the second example, we consider a case, where the order of the estimated model is less than that of the true system. In the last example, the effectiveness of the proposed algorithm is examined in the identification of an LPV system with dynamic dependency.

A. Example 1

The data-generating system is defined by the FIR structure [13]

$$y(k) = \sum_{i=0}^{2} b_i(p_{k-i})u(k-i) + e_0(k), \tag{24}$$

with $\mathbb{P} = [-1, 1]$ and

$$\begin{aligned} &b_0(p_k) = -\exp(-p_k), \quad b_1(p_{k-1}) = 1 + p_{k-1}, \\ &b_2(p_{k-2}) = \tan^{-1}(p_{k-2}). \end{aligned} \tag{25}$$

This I/O representation has a nonlinear dynamic dependence on p and is in the FIR form. A data set $\mathcal{D}_N$ with $N = 200$ samples is generated by (24) using a periodic $u(k)$ with $T = 3$ as

$$u(k) = \begin{cases} 1 & k = 1 \\ 0 & k = 2, 3 \end{cases} \tag{26}$$

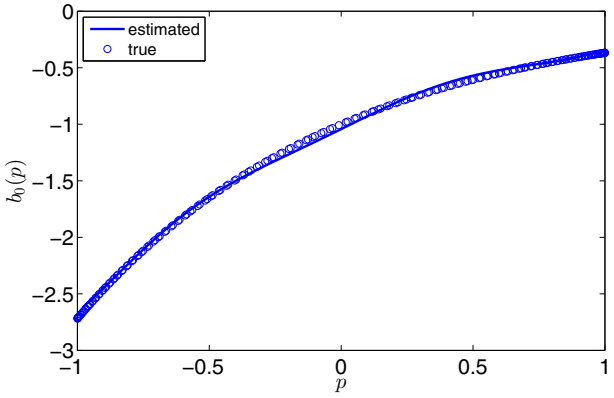

Fig. 1: Estimation results of the coefficient function b_0.

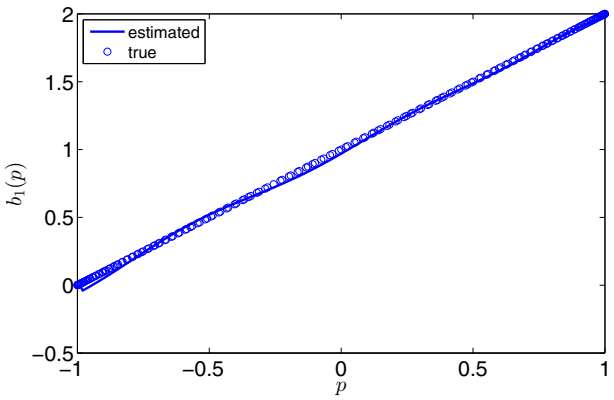

Fig. 2: Estimation results of the coefficient function b_1.

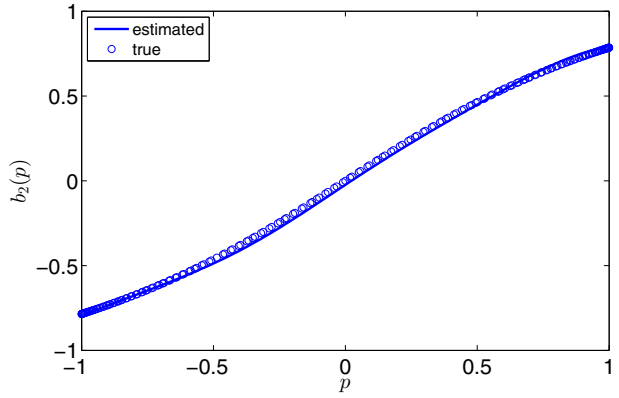

Fig. 3: Estimation results of the coefficient function b_2.

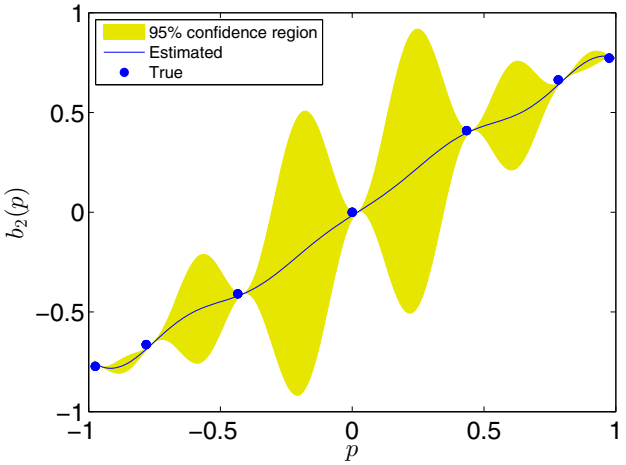

Fig. 4: 95% confidence region for predictions of the coefficient function $b_2(p_k)$.

and a scheduling trajectory $p_k = \sin(\frac{2\pi k}{61}) = \sin(0.1030k)$. The noise $e_0(k)$ is assumed to be i.i.d. and have a Gaussian distribution $\mathcal{N}(0, \sigma^2)$ with $\sigma > 0$. To investigate the performance at the presence of a noise e_0, the *signal-to-noise ratio* (SNR) is set as $\text{SNR} = 10 \log \frac{P_{\mathcal{X}}}{P_{e_0}} = 10\text{dB}$, where $P_{\mathcal{X}}$ is the average power of signal $\mathcal{X}$, which is deterministic component of y in (24). To characterize the nonlinearities in this system, RBF kernels have been used with $\sigma_1 = \sigma_3 = \sigma_2 = 0.7$ and $\lambda_1 = 1.5, \lambda_2 = \lambda_3 = 1$ based on maximizing the fitness score in (27). The obtained results with respect to the estimation of the underlying coefficient functions are shown in Figures 1 to 3. To quantify the model quality, the *fitness score* or *Best Fit Rate* (BFR) is used [13]:

$$\text{BFR}^i = 100\% \times \max\left(1 - \frac{||b_i(k) - g_i(k)||_{\ell_2}}{||b_i(k) - \bar{b}_i||_{\ell_2}}, 0\right). \quad (27)$$

The obtained fitness score for functions $b_0(p), b_1(p)$ and $b_2(p)$ are $96.85\%, 98.63\%$ and 97.69%, respectively.

Now consider the case that we have $p_k = \sin(\frac{2\pi k}{7})$ which is a periodic function with $T = 7$. Unlike the previous scenario, we have only 7 points of scheduling parameters. As the number of training examples decreases, the size of the confidence region for prediction of the model grows to reflect the increasing uncertainty in the model estimates. Here, the confidence region is defined based on the *covariance* function in Equation (20). Our aim is to find the 95% confidence region for predictions of the model coefficient $b_2(p_k)$. As shown in Figure 5, one of the advantages of using GP is the fact that one can compute an actual variance of the estimated function and designate where more samples are required to improve the estimation.

B. Example 2

In this example, we consider a case, in which an LPV model with order less than the order of the original system is identified using the proposed Bayesian estimation approach. The actual LPV system is represented by

$$y(k) = \sum_{i=0}^{4} b_i(p_{k-i})u(k-i) + e_0(k), \quad (28)$$

with $\mathbb{P} = [-1, 1]$ and

$$\begin{aligned} &b_0(p_k) = -\exp(-p_k), \quad b_1(p_{k-1}) = 1 + p_{k-1}, \\ &b_2(p_{k-2}) = \tan^{-1}(p_{k-2}), \quad b_3(p_{k-3}) = -p_{k-3}, \\ &b_4(p_{k-4}) = -\sin(p_{k-4}). \end{aligned} \quad (29)$$

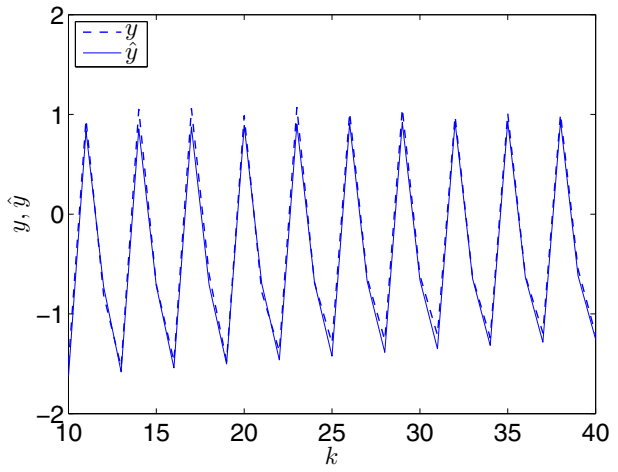

Fig. 5: True output and the predicted output by the estimated model for k=10:40.

.

signals $u(k)$ and p_k are similarly generated as in Example 1 and $e_0(k) = 0$ (noise-free measurement) is considered to estimate an LPV-FIR model with order of 3. The true output and the output generated by the identified third-order LPV model $\hat{y}$ are shown in Figure 5.

To quantify the model quality in this example, we consider the BFR w.r.t the output prediction

$$\text{BFR} = 100\% \times \max\left(1 - \frac{||y - \hat{y}||_{\ell_2}}{||y - \bar{y}||_{\ell_2}}, 0\right), \qquad (30)$$

where $\bar{y}$ is the *mean* of y. The BFR in this example is calculated to be 93.7%.

C. Example 3

In this example, the following LPV system with a dynamic dependency is considered as

$$y(k) = \sum_{i=0}^{2} b_i(\mathbf{p}(\mathbf{k}-\mathbf{i},\mathbf{1}))u(k-i) + e_0(k), \qquad (31)$$

with $\mathbb{P} = [-1, 1]$ and

$$\begin{aligned} b_0(\mathbf{p}(\mathbf{k},\mathbf{1})) &= -\exp(-p_k) + p_{k-1}^2, \\ b_1((\mathbf{p}(\mathbf{k}-\mathbf{1},\mathbf{1}))) &= 1 + p_{k-1} + p_{k-2}^3, \\ b_2((\mathbf{p}(\mathbf{k}-\mathbf{2},\mathbf{1}))) &= \tan^{-1}(p_{k-2}) + \sin(p_{k-3}). \end{aligned} \qquad (32)$$

A data set $\mathcal{D}_N$ with $N = 200$ is generated by (31) with $u(k)$ being periodic with $T = 3$, $u(1) = 1, u(2) = u(3) = 0$, $p_k = \sin(0.1030k)$ and the noise $e_0(k)$ is assumed to be i.i.d. which as Gaussian distribution $\mathcal{N}(0, \sigma^2)$ with $\sigma > 0$. The SNR is set as 10dB. The obtained results with respect to the estimation of the coefficient functions b_0, b_1 and b_2 are shown in Figures 6 to 8. The obtained BFR for the functions b_0, b_1 and b_2 are calculated to be $95.60\%, 97.20\%$ and 92.84%, respectively.

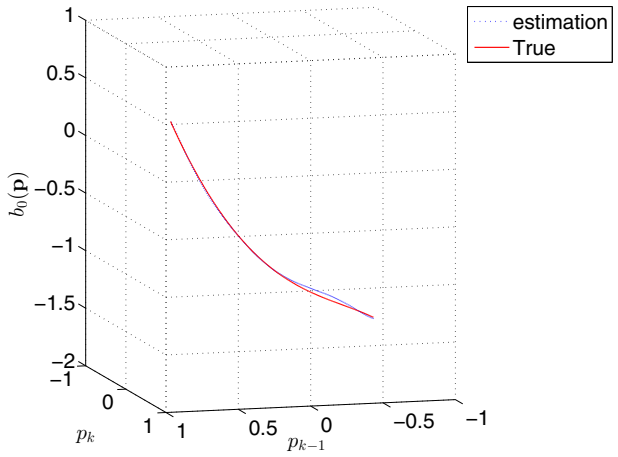

Fig. 6: Estimation results of the coefficient function b_0.

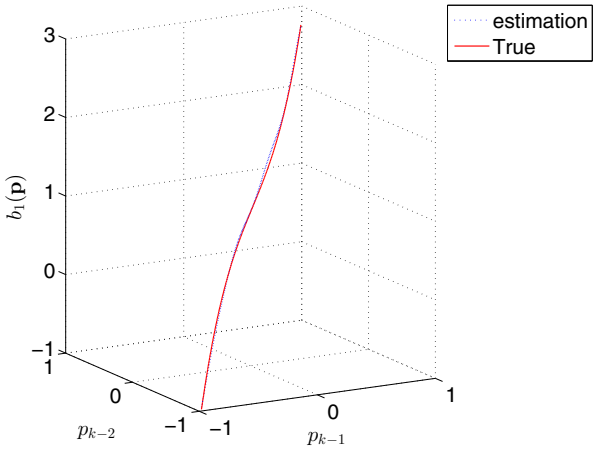

Fig. 7: Estimation results of the coefficient function b_1.

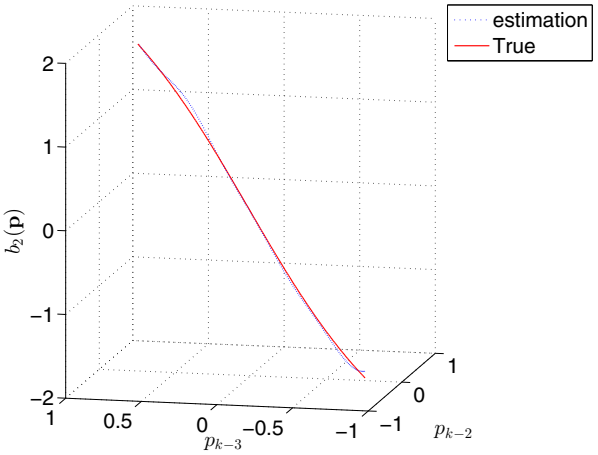

Fig. 8: Estimation results of the coefficient function b_2.

IV. Conclusion

In this paper, a Bayesian formulation for identification of LPV FIR models has been introduced. A Gaussian process (GP) is used to describe the distribution over nonlinear coefficient functions. Without a prior information about parametrization of the underlying coefficient functions, the proposed approach is capable of reconstructing the dependency structure of the LPV model and providing confidence bounds based on the posterior distribution. The posterior distribution is the main difference between Bayesian method and the LS-SVM approach, which provides a useful way to quantify the uncertainties of the estimated model and the coefficient functions.

References

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

[2] P. Gebraad, J. van Wingerden, P. Fleming, and A. Wright, "LPV identification of wind turbine rotor vibrational dynamics using periodic disturbance basis functions," *Control Systems Technology, IEEE Transactions on*, vol. 21, no. 4, pp. 1183–1190, July 2013.

[3] D. Vizer, G. Mercere, O. Prot, and J. Ramos, "A local approach framework for black-box and gray-box LPV system identification," in *Control Conference (ECC), 2013 European*, July 2013, pp. 1916–1921.

[4] W. Larimore and H. Javaherian, "Identification of linear parameter-varying engine models," in *American Control Conference (ACC), 2013*, June 2013, pp. 2274–2279.

[5] K. Hsu, T. L. Vincent, and K. Poola, "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, 2008, pp. 846–851.

[6] S. Wattamwar, S. Weiland, and T. Backx, "Identification of low-order parameter-varying models for large-scale systems," *Journal of Process Control*, vol. 20, no. 1, pp. 158–172, 2010.

[7] R. Tóth, C. Lyzell, M.Enqvist, P. S. C. Heuberger, and P. M. J. V. 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, 2009, pp. 7406–7411.

[8] B. Bamieh and L. Giarre, "Identification of linear parameter varying models," *Int. Journal of Robust and Nonlinear Control*, vol. 12, no. 1, pp. 841–853, Apr. 2002.

[9] J. W. van Wingerden and M. Verhaegen, "Subspace identification of Bilinear and LPV systems for open- and closed-loop data," *Automatica*, vol. 45, pp. 372–381, 2009.

[10] 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.

[11] H. Abbas and H. Werner, "An Instrumental Variable Technique for Open-Loop and Closed-Loop Identfication of Input-Output LPV Models," in *Proc. of the European Control Conf.*, Budapest, Hungary, 2009, pp. 2646–2651.

[12] S. Boonto and H. Werner, "Closed-loop system identification of LPV input-output models - application to an arm-driven inverted pendulum," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, Dec 2008, pp. 2606–2611.

[13] R. Tóth, V. Laurain, W.-X. Zheng, and K. Poolla, "Model structure learning: A support vector machine approach for LPV linear-regression models," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, Dec 2011, pp. 3192–3197.

[14] V. Laurain, R. Tóth, W.-X. Zheng, and M. Gilson, "Nonparametric identification of LPV models under general noise conditions: An LS-SVM based approach," in *Proceedings of the 16th IFAC Symposium on System Identification*, July 2012, pp. 1761–1766.

[15] D. Piga and R. Tóth, "LPV model order selection in an ls-svm setting," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, Dec 2013, pp. 4128–4133.

[16] B. Scholkopf and A. Smola, *Learning with kernels*, ser. Lecture Notes in Control and Information Sciences. Berlin, Germany: Cambridge MA:MIT press, 1998.

[17] A. Chiuso and G. Pillonetto, "A Bayesian approach to sparse dynamic network identification," *Automatica*, vol. 48, no. 8, pp. 1553 – 1565, 2012.

[18] G. Pillonetto and G. D. Nicolao, "A new kernel-based approach for linear system identification," *Automatica*, vol. 46, no. 1, pp. 81 – 93, 2010.

[19] A. Chiuso, G. Pillonetto, and G. De Nicolao, "Subspace identification using predictor estimation via Gaussian regression," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, Dec 2008, pp. 3299–3304.

[20] T. Chen, H. Ohlsson, and L. Ljung, "On the estimation of transfer functions, regularizations and gaussian processes-revisited," *Automatica*, vol. 48, no. 8, pp. 1525 – 1535, 2012.

[21] V. Peterka, "Bayesian approach to system identification," in *Trends and Progress in System Identification*. Pergamon Press, 1981, pp. 239–304.

[22] G. Pillonetto and A. Chiuso, "Tuning complexity in kernel-based linear system identification: The robustness of the marginal likelihood estimator," in *Control Conference (ECC), 2014 European*, June 2014, pp. 2386–2391.

[23] A. Chiuso and G. Pillonetto, "Bayesian and nonparametric methods for system identification and model selection," in *Control Conference (ECC), 2014 European*, June 2014, pp. 2376–2381.

[24] G. Pillonetto, F. Dinuzzo, T. Chen, G. D. Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657 – 682, 2014.

[25] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. London, England: Massachusetts Institute of Technology, 2006, vol. 403.

[26] G. Pillonetto, M. H. Quang, and A. Chiuso, "A new kernel-based approach for nonlinear system identification," *IEEE Transactions on Automatic Control*, vol. 56, no. 12, pp. 2825–2840, December 2011.