

Alternative Form of Predictor Based Identification of LPV-SS Models with Innovation Noise

Pepijn Cox[†] and Roland Tóth[†]

Abstract—In this paper, we present an approach to identify linear parameter-varying (LPV) systems with a state-space (SS) model structure in an innovation form where the coefficient functions have static and affine dependency on the scheduling signal. With this scheme, the curse of dimensionality problem is reduced, compared to existing predictor based LPV subspace identification schemes. The investigated LPV-SS model is reformulated into an equivalent impulse response form, which turns out to be a moving average with exogenous inputs (MAX) system. The Markov coefficient functions of the LPV-MAX representation are multi-linear in the scheduling signal and its time-shifts, contrary to the predictor based schemes where the corresponding LPV auto-regressive with exogenous inputs system is multi-quadratic in the scheduling signal and its time-shifts. In this paper, we will prove that under certain conditions on the input and scheduling signals, the ℓ_2 loss function of the one-step-ahead prediction error for the LPV-MAX model has only one unique minimum, corresponding to the original underlying system. Hence, identifying the LPV-MAX model in the prediction error minimization framework will be consistent and unbiased. The LPV-SS model is realized by applying an efficient basis reduced Ho-Kalman realization on the identified LPV-MAX model. The performance of the proposed scheme is assessed on a Monte Carlo simulation study.

I. INTRODUCTION

In recent years, identification of *linear parameter-varying state-space* (LPV-SS) models has received considerable attention, e.g., [1]–[4], resulting in a successful extension of *subspace identification* (SID) and *prediction error minimization* (PEM) methods to the LPV case. The LPV-SS identification framework has been applied to data-driven modelling of wind turbines [5], forced Lorenz attractors [6], power management of Web service systems [7], or to capture traffic flow [8], to mention a few. A popular choice is to capture the underlying system as a discrete-time LPV-SS representation with an innovation noise model:

$$qx = \mathcal{A}(p)x + \mathcal{B}(p)u + \mathcal{K}(p)e, \quad (1a)$$

$$y = \mathcal{C}(p)x + \mathcal{D}(p)u + e, \quad (1b)$$

where $x : \mathbb{Z} \rightarrow \mathbb{X} = \mathbb{R}^{n_x}$ is the state variable, $y : \mathbb{Z} \rightarrow \mathbb{Y} = \mathbb{R}^{n_y}$ is the measured output signal, $u : \mathbb{Z} \rightarrow \mathbb{U} = \mathbb{R}^{n_u}$ denotes the input signal, $p : \mathbb{Z} \rightarrow \mathbb{P} \subseteq \mathbb{R}^{n_p}$ is the scheduling variable, q is the forward time-shift operator, e.g., $qx(t) = x(t+1)$ where $t \in \mathbb{Z}$ is the discrete time, $e : \mathbb{Z} \rightarrow \mathbb{R}^{n_y}$ is a sampled path of a zero-mean i.i.d. stationary noise processes with Gaussian distribution, i.e., $e(t) \sim \mathcal{N}(0, \Sigma_e)$ with a nonsingular covariance $\Sigma_e \in \mathbb{R}^{n_y \times n_y}$. The matrix functions $\mathcal{A}(\cdot), \dots, \mathcal{K}(\cdot)$ defining the SS representation (1)

are usually taken to be affine combinations of p , to coincide with the majority of LPV control synthesis methods, e.g., [9]. Hence, these matrix functions are defined as

$$\mathcal{A}(p) = A_0 + \sum_{i=1}^{n_p} A_i p_i, \quad (2)$$

where $\mathcal{B}(p), \dots, \mathcal{K}(p)$ are equivalently parametrized as (2) with $\{A_i, B_i, C_i, D_i, K_i\}_{i=0}^{n_p}$ being constant matrices of appropriate dimensions.

Identification of LPV-SS models can be formulated in the PEM framework, where the PEM methods can roughly be categorized as: i) methodologies which assume that full state measurements are available, e.g., [10]; ii) grey box schemes, where only a small subset of all parameters is estimated, e.g., [11]; iii) set-membership approaches, e.g., [12]; or iv) direct PEM methods using gradient based methodologies, e.g., [2], [3]. However, assuming that a full measurement of the state or detailed model of the system is available, is unrealistic in many applications. Therefore, only set-membership and direct PEM methods classify as ‘black-box’ identification schemes. In general, the set-membership approaches have a significant higher computational load compared to the direct PEM methods, as the set-membership schemes estimate a feasible model set for which only heuristic techniques exist. In addition, these schemes have often unrealistic noise assumptions and consistency of the overall scheme is difficult to show. For the direct PEM methods, the nonlinear and nonunique optimization problem is usually solved by applying gradient-based search strategies or an expectation maximization strategy. Regardless of the strategy, direct PEM methods are solved in an iterative way, are prone to local minima, and their convergence depends heavily on a proper initial guess.

On the contrary, SID methods use convex optimization, to identify a specific LPV *input-output* (IO) structure, from which an LPV-SS model is realized by using matrix decomposition methods. Subspace schemes have extensively been applied in the *linear time invariant* (LTI) case and extensions to the LPV framework can be found, e.g., [1], [4], [13], [14]. Unfortunately, SID methods usually depend on approximations to get a convex problem, and, as a consequence, in the LPV case, these approaches suffer heavily from the curse of dimensionality and/or result in ill-conditioned estimation problems with high parameter variances.

Based on these considerations, efficient estimation of LPV-SS models remains a central problem to be solved.

In this paper, we would like to reduce the dimensionality problems associated with the predictor based SID methods

[†] Pepijn Cox and Roland Tóth are with the Control Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, {p.b.cox, r.toth}@tue.nl

by restating the LPV-IO identification setting to avoid multi-quadratic dependency on the scheduling signal and its time-shifts. First, the LPV-SS representation (1) is formulated in terms of its *impulse response representation* (IIR). As Section II highlights, the corresponding IIR turns out to be a *moving average with exogenous inputs* (MAX) representation, where the Markov coefficient functions are multi-linear in the scheduling signal and its time-shifts. In Section III, the identification of the multivariable MAX model is given. We will prove that minimizing the ℓ_2 loss function of the one-step-ahead prediction error has only one unique minimum under some mild conditions on the input and scheduling signals, which is the main contribution of this paper. To this end, we extend the uniqueness proof of the LTI multivariable moving average model [15] to the LPV case, which is essential in showing uniqueness of the overall LPV-MAX PEM identification method. Hence, applying a pseudo linear regression to identify the LPV-MAX model results in a consistent and unbiased estimate. After identifying the LPV-MAX model, the LPV-SS model is estimated by utilizing the bases reduced Ho-Kalman realization scheme of [16], which, in this paper, is modified to also realize the noise model. In Section V, the performance of the identification method is assessed by a Monte-Carlo study, followed by some conclusions in Section VI.

II. IMPULSE RESPONSE REPRESENTATION

To be able to solve our identification problem, we use an alternative representation of (1):

Lemma 1 (IIR representation [17]): Any asymptotically stable¹ LPV system given in terms of representation (1) has a convergent series expansion in terms of the pulse-basis $\{q^{-i}\}_{i=0}^{\infty}$ given by

$$y = \underbrace{\sum_{i=0}^{\infty} (g_i \diamond p) q^{-i} u}_{\text{process model}} + \underbrace{\sum_{j=0}^{\infty} (h_j \diamond p) q^{-j} e}_{\text{noise model}}, \quad (3)$$

where $g_i \in \mathcal{R}^{n_y \times n_u}$ and $h_j \in \mathcal{R}^{n_y \times n_e}$ are the expansion coefficient functions, i.e., *Markov coefficients*, of the process and noise dynamics, respectively. The ring of all real polynomial functions with finite dimensional domain is defined by \mathcal{R} , the operator $\diamond : (\mathcal{R}, \mathbb{P}^{\mathbb{Z}}) \rightarrow (\mathbb{R}^{n_y \times n_u})^{\mathbb{Z}}$ denotes $(g_i \diamond p) = g_i(p(t), \dots, p(t-\tau))$ with $\tau \in \mathbb{Z}$. \square

The IIR coefficients $\{g_i, h_i\}_{i=0}^{\infty}$ of (3) are given by

$$y = \underbrace{\mathcal{D}(p)}_{g_0 \diamond p} u + \underbrace{\mathcal{C}(p)\mathcal{B}(q^{-1}p)}_{g_1 \diamond p} q^{-1} u + \underbrace{\mathcal{C}(p)\mathcal{A}(q^{-1}p)\mathcal{B}(q^{-2}p)}_{g_2 \diamond p} q^{-2} u + \dots + \underbrace{I}_{h_0 \diamond p} e + \underbrace{\mathcal{C}(p)\mathcal{K}(q^{-1}p)}_{h_1 \diamond p} q^{-1} e + \underbrace{\mathcal{C}(p)\mathcal{A}(q^{-1}p)\mathcal{K}(q^{-2}p)}_{h_2 \diamond p} q^{-2} e + \dots \quad (4)$$

where g_i, h_i converges to the zero function in de ℓ_{∞} sense as $i \rightarrow \infty$. The Markov coefficients of the process part can be written as

¹An LPV system, represented in terms of (1), is called asymptotically stable in the deterministic sense, if, for all trajectories of $(u(t), e(t), p(t), y(t))$ satisfying (1), with $u(t) = 0, e(t) = 0$ for $t \geq 0$ and $p(t) \in \mathbb{P}$, it holds that $\lim_{t \rightarrow \infty} |y(t)| = 0$.

$$g_m \diamond p = \mathcal{C}(p)\mathcal{A}(q^{-1}p) \cdots \mathcal{A}(q^{-(m-1)}p)\mathcal{B}(q^{-m}p) = \sum_{i=0}^{n_p} \sum_{j=0}^{n_p} \cdots \sum_{k=0}^{n_p} \sum_{l=0}^{n_p} C_i A_j \cdots A_k B_l p_i (q^{-1}p_j) \cdots (q^{-m}p_l), \quad (5)$$

and the Markov coefficients of the noise part h_m are similar to (5), however, $\mathcal{B}(q^{-m}p)$ is exchanged for $\mathcal{K}(q^{-m}p)$. The individual products $C_i A_j \cdots A_k B_l$ or $C_i A_j \cdots A_k K_l$ are called the *sub-Markov parameters* of the process model and noise model, respectively. Due to the convergence of g_i and h_j , it is often sufficient to truncate (3):

$$y \approx \sum_{i=0}^{n_b} (g_i \diamond p) q^{-i} u + \sum_{j=0}^{n_c} (h_j \diamond p) q^{-j} e, \quad (6)$$

where $n_b > 0$ and $n_c > 0$ are the order of the resulting *finite impulse response* (FIR) models for the process and noise part.

The sub-Markov parameters in (5) have a multi-linear dependency on the elements of p and its time-shifts, contrary to the LPV *auto-regressive with exogenous inputs* (ARX) formulation of other predictor based subspace schemes, e.g. [1], which have multi-quadratic dependency. The LPV-ARX model² uses $\tilde{\mathcal{A}}(p) = \mathcal{A}(p) - \mathcal{K}(p)\mathcal{C}(p)$, therefore, increasing the complexity of the LPV-IO model to be identified. To see this, substitute $\mathcal{A}(p) = \sum A_i p_i$ by $\tilde{\mathcal{A}}(p) = \sum \sum A_i p_i - K_i C_j p_j$ (with $p_0 = 1$) in (5) and, similarly, \mathcal{B} by $\tilde{\mathcal{B}}$. To compare the additional parameters, truncate the LPV-ARX model (7) by orders n_a and n_d for the u and y polynomial, respectively; then the LPV-ARX model (7) has $n_y(n_u \sum_{i=0}^{n_a} (1+n_p)^{2i+1} + n_y \sum_{j=1}^{n_d} (1+n_p)^{2j})$ parameters and the LPV model (6) has $n_y(n_u \sum_{i=1}^{n_b+1} (1+n_p)^i + n_y \sum_{j=2}^{n_c+1} (1+n_p)^j)$. For example, if $n_u = n_y = n_p = n_a = n_b = n_c = n_d = 2$ then the LPV-ARX model has 1452 parameters and the proposed LPV model has 300, hence, a significant reduction is achieved. Therefore, a common assumption by other predictor based subspace schemes is to take either \mathcal{K} or \mathcal{C}, \mathcal{D} to be constant matrices, avoiding the additional modelling complexity. However, by identifying (4), we can keep \mathcal{K}, \mathcal{C} , and \mathcal{D} to be parameter dependent.

III. IDENTIFICATION OF A MAX MODEL

A. Problem Setting

The Markov coefficients (4) of the impulse response (6) representing the LPV-SS system (1) are functions in the scheduling signal and its time shifts, i.e., (5). Hence (6) is a *moving average with exogenous inputs* (MAX) system:

$$y = \mathcal{B}(p, q^{-1})u + \mathcal{C}(p, q^{-1})e, \quad (8)$$

with the process and noise filter given by

$$\mathcal{B}(p, q^{-1}) = \sum_{i=0}^{n_b} \mathcal{B}_i(p) q^{-i}, \quad \mathcal{C}(p, q^{-1}) = \sum_{j=0}^{n_c} \mathcal{C}_j(p) q^{-j}, \quad (9)$$

²The LPV-ARX representation is found by substituting e of (1b) into (1a) and, by using this modified state equation, writing out (1b) as:

$$y = \mathcal{D}(p)u + \sum_{i=1}^{\infty} \mathcal{C}(p) \left[\prod_{j=1}^{i-1} \tilde{\mathcal{A}}(q^{-j}p) \right] \tilde{\mathcal{B}}(q^{-i}p) q^{-i} u + \sum_{i=1}^{\infty} \mathcal{C}(p) \left[\prod_{j=1}^{i-1} \tilde{\mathcal{A}}(q^{-j}p) \right] \mathcal{K}(q^{-i}p) q^{-i} y + e \quad (7)$$

where $\tilde{\mathcal{A}}(p) = \mathcal{A}(p) - \mathcal{K}(p)\mathcal{C}(p)$, $\tilde{\mathcal{B}}(p) = \mathcal{B}(p) - \mathcal{K}(p)\mathcal{D}(p)$, and $\prod_{j=1}^0 \tilde{\mathcal{A}} = I$.

where $\mathcal{B}_i(p) = (g_i \diamond p)$ and $\mathcal{C}_j(p) = (h_j \diamond p)$. Hence, $\mathcal{B}_i(p)$ and $\mathcal{C}_j(p)$ are multi-linear matrix functions in the scheduling signal and its times-shifts from $t, \dots, t-i$, similar to the Markov coefficients (5). In this paper, the LPV-MAX model (9) is identified by using the PEM setting. To avoid identifiability issues, (1) with functional dependencies (2) is assumed to be structurally state observable and state reachable w.r.t. both u and e in the deterministic sense³, which implies joint minimality of (1).

In order to apply the PEM framework, we select the *model structure* similar to (9):

$$y = \hat{\mathcal{B}}(\theta, p, q^{-1})u + \hat{\mathcal{C}}(\theta, p, q^{-1})\varepsilon(\theta), \quad (10)$$

where $\varepsilon(\theta) : \mathbb{Z} \rightarrow \mathbb{Y}$ is the one-step-ahead prediction error and the process $\hat{\mathcal{B}}$ and noise $\hat{\mathcal{C}}$ models are considered to be polynomials of q^{-1} with p -dependent coefficients similar to \mathcal{B} and \mathcal{C} in (9) with orders \hat{n}_b, \hat{n}_c , respectively. These polynomials are parametrized in terms of the sub-Markov parameters resulting in an overall parameter vector θ .

The data-generating system (1) is aimed to be identified using a given identification dataset $\mathcal{D}_N = \{u(t), p(t), y(t)\}_{t=1}^N$ generated by (1) and minimizing the following ℓ_2 loss function

$$V_N(\theta) = \text{Tr} \left(\frac{1}{N} \sum_{t=1}^N \varepsilon(\theta, t) \varepsilon^\top(\theta, t) \right) \quad (11)$$

where N is the number of data points collected in \mathcal{D}_N . Under weak regularity conditions it is well known that

$$V_N(\theta) \rightarrow V_\infty(\theta) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \text{Tr}(\mathbb{E} \{ \varepsilon(\theta, t) \varepsilon^\top(\theta, t) \}), \quad (12)$$

with probability one and uniformly in θ [18], which removes the effect of sample based realizations of u, p , and e in the analysis. Due to the uniform convergence of (12), the results of Section III-C and III-D give information about the shape of $V_N(\theta)$ for large N .

First, let us define a notation to indicate which sub-Markov parameters (5) of the filters \mathcal{B}, \mathcal{C} are selected. Denote with \mathbb{I}_s^v the set $\{s, s+1, \dots, v\}$. Then, $[\mathbb{I}_s^v]^n$ defines the set of all sequences of the form (i_1, \dots, i_n) with $i_1, \dots, i_n \in \mathbb{I}_s^v$. The elements of \mathbb{I}_s^v will be viewed as characters and the finite sequences of elements of \mathbb{I}_s^v will be referred to as strings. Then $[\mathbb{I}_s^v]^n$ is the set of all strings containing exactly n characters. Then a selection with $n \geq 0$ is constructed from $\eta \in [\mathbb{I}_0^{n_p}]_0^n$ with $[\mathbb{I}_0^{n_p}]_0^n = \{\epsilon\} \cup \mathbb{I}_0^{n_p} \cup \dots \cup [\mathbb{I}_0^{n_p}]_0^n$ and ϵ denoting the empty string. As an example, $[\mathbb{I}_0^1]_0^2 = \{\epsilon, 0, 1, 00, 01, 10, 11\}$. Define by $\#(\eta)$ the amount of characters of a single string in the set. With this notation, we will simplify the notation of the sub-Markov parameters as

$$C_{[\eta_b]_1} A_{[\eta_b]_2} \dots A_{[\eta_b]_{i-1}} B_{[\eta_b]_i} = \mathbf{g}_{\eta_b}, \quad (13)$$

where $i = \#(\eta_b)$, $[\eta]_j$ denotes the j -th character of the string η and $\eta_b \in [\mathbb{I}_0^{n_p}]_2^{n_b+1}$. Hence, $[\mathbb{I}_0^{n_p}]_2^{n_b+1}$ is the set of all possible combinations of the sub-Markov parameters for the process filter. Similarly, we indicate a sub-Markov parameter of the noise filter \mathcal{C} , process model $\hat{\mathcal{B}}$, and noise model $\hat{\mathcal{C}}$

³There exists at least one $p \in \mathbb{P}^Z$ such that n_x -step observability and reachability holds for all time moments on the support of the signals [17].

by $\eta_c \in [\mathbb{I}_0^{n_p}]_2^{n_c+1}$, $\eta_b \in [\mathbb{I}_0^{n_p}]_2^{\hat{n}_b+1}$, and $\eta_e \in [\mathbb{I}_0^{n_p}]_2^{\hat{n}_c+1}$, respectively. Then define⁴

$$Q(\theta) = \mathbb{E} \{ \varepsilon(\theta, t) \varepsilon(\theta, t)^\top \} \quad (14a)$$

$$\begin{aligned} \hat{\Gamma}(\theta, p, q^{-1}) &= \sum_{i=0}^{\infty} \hat{\Gamma}_i(\theta, p, t) q^{-i} \\ &= \hat{\mathcal{C}}^{-1}(\theta, p, q^{-1}), \quad (\Gamma_0 = I) \end{aligned} \quad (14b)$$

$$R_k(\theta, t) = \varepsilon(\theta, t-k) \varepsilon^\top(\theta, t), \quad k \geq 0, \quad (14c)$$

where

$$\mathbb{E} \{ \cdot \} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \mathbb{E} \{ \cdot \}.$$

Remark, we use a simplified notation $\hat{\Gamma}_i(\theta, p, t) = (\hat{\Gamma}_i(\theta) \diamond p)(t)$ where each $\hat{\Gamma}_i(\theta, p, t)$ depends on $p(t), \dots, p(t-i)$.

B. Pseudo linear regression

The advantage of the given problem setting of Section III-A is that it can easily be solved by applying global pseudo linear regression methods, e.g., [19, Algorithm 3]. Pseudo linear regression has a relatively low computational load. Due to space limitations, the algorithm is not presented here. If we take \hat{n}_b and \hat{n}_c to be finite, the corresponding LPV-MAX model can be seen as an extended LTI model. In the LTI case, pseudo linear regression has been studied extensively and it is known that it will converge if no auto-regressive part is present [20]. However, a key part of the convergence is to have only one unique solution of the minimization problem (12), which is analysed in Sections III-C and III-D.

C. Global minima

To show that a global minimum exists, the following assumption is taken:

A1 The input signal u is uncorrelated to the noise signal e . With the aforementioned identification setting, we get:

Theorem 2: Given the data generating system (8) with functional dependencies (9), model structure (10), $\hat{n}_b \geq n_b$, $\hat{n}_c \geq n_c$, and u satisfying A1; then $V_N(\theta)$ has a global minimum

$$\min_{\theta} V_\infty(\theta) = \text{Tr}(\Sigma_e), \quad (15)$$

which can be obtained if the elements of θ satisfy

$$\hat{\mathbf{g}}_{\eta_b} = \mathbf{g}_{\eta_b}, \quad \hat{\mathbf{h}}_{\eta_c} = \mathbf{h}_{\eta_c}, \quad (16a)$$

for all indices $\eta_b \in [\mathbb{I}_0^{n_p}]_2^{n_b+1}$ and $\eta_c \in [\mathbb{I}_0^{n_p}]_2^{n_c+1}$. In addition, if $\hat{n}_b > n_b, \hat{n}_c > n_c$,

$$\hat{\mathbf{g}}_{\eta_b} = 0, \quad \hat{\mathbf{h}}_{\eta_c} = 0, \quad (16b)$$

for all $\eta_b \in [\mathbb{I}_0^{n_p}]_{n_b+2}^{\hat{n}_b+1}$ and $\eta_c \in [\mathbb{I}_0^{n_p}]_{n_c+2}^{\hat{n}_c+1}$. \square

Proof: For notational ease, denote $\mathcal{B}(p, q^{-1}), \hat{\mathcal{B}}(\theta, p, q^{-1}), \mathcal{C}(p, q^{-1}), \hat{\Gamma}(\theta, p, q^{-1})$, and $\varepsilon(\theta)$ as $\mathcal{B}, \hat{\mathcal{B}}, \mathcal{C}, \hat{\Gamma}$, and ε , respectively. Then, using (8) and (10), let us rewrite ε in terms of u and e as

$$\varepsilon = \hat{\Gamma} \left[(\mathcal{B} - \hat{\mathcal{B}}) u + \mathcal{C} e \right].$$

Hence, as e and u are uncorrelated (A1), the loss function (12) separates as

$$V_\infty(\theta) = V_{\infty,1}(\theta) + V_{\infty,2}(\theta), \quad (17)$$

⁴The inverse filter $\hat{\Gamma}(\theta, p, q^{-1})$ of $\hat{\mathcal{C}}(\theta, p, q^{-1})$ in (14b) is generated by $\hat{\Gamma}(\theta, p, q^{-1}) = \sum_{i=0}^{\infty} (I - \hat{\mathcal{C}}(\theta, p, q^{-1}))^i$ [19].

with

$$V_{\infty,1}(\theta) = \text{Tr} \left[\mathbb{E} \left\{ \hat{\Gamma} (\mathcal{B} - \hat{\mathcal{B}}) u(t) u^\top(t) (\mathcal{B}^\top - \hat{\mathcal{B}}^\top) \hat{\Gamma}^\top \right\} \right], \quad (18a)$$

$$V_{\infty,2}(\theta) = \text{Tr} \left[\mathbb{E} \left\{ \hat{\Gamma} \mathcal{C} e(t) e^\top(t) \mathcal{C}^\top \hat{\Gamma}^\top \right\} \right]. \quad (18b)$$

Obviously, $V_{\infty,1}(\theta) \geq 0$ with equivalence if $\mathcal{B} \equiv \hat{\mathcal{B}}$. For $V_{\infty,2}(\theta)$, remark that the filters \mathcal{C} and $\hat{\mathcal{C}}$ are monic,

$$\varepsilon = \hat{\Gamma} \mathcal{C} e = e + v, \quad (19)$$

where the random signal $v(t)$ depends linearly on the past samples of e , i.e., $\{e(\tau)\}_{\tau=-\infty}^{t-1}$, but is independent of $e(t)$. Hence, it follows that

$$\begin{aligned} V_{\infty,2}(\theta) &= \text{Tr} \left(\mathbb{E} \left\{ (e(t) + v(t))(e(t) + v(t))^\top \right\} \right) \\ &= \text{Tr} \left(\mathbb{E} \{ e(t) e^\top(t) \} \right) + \text{Tr} \left(\mathbb{E} \{ v(t) v^\top(t) \} \right) \\ &\geq \text{Tr} \left(\mathbb{E} \{ e(t) e^\top(t) \} \right) = \text{Tr}(\Sigma_e). \end{aligned} \quad (20)$$

Eq. (20) holds with equality if $v \equiv 0$, which follows if $\hat{\Gamma} \mathcal{C} = I$ implying $\mathcal{C} \equiv \hat{\mathcal{C}}$. Hence, $V_{\infty}(\theta) = \text{Tr}(\Sigma_e)$ if (16) holds, which is a global minimum of $\min_{\theta} V_{\infty}(\theta)$. ■

Theorem 2 does not imply uniqueness of the solution, only that a global minimum exists if the sub-Markov parameters of the model (10) are equal to those of the original system (9). Uniqueness of the solution is proven next.

D. Stationary point

Next, to show uniqueness of the of minimum in Theorem 2, we prove that there exists only one stationary point of (15), hence, the global minimum of (15) is unique and it can always be obtained by optimization methods. To prove this, the following additional assumptions are taken:

- A2** Each signal p_i is assumed to be a zero-mean white noise process with finite variance and independent of e . The processes p_i are mutually independent and the higher order moments are bounded, given by $\mathbb{E}\{p_i^k p_j^l\} = c_{i,j}^{k+l} < \infty$ for $i, j \in \mathbb{I}_1^{n_p}$ and $2 \leq k + j \leq 6 + \hat{n}_c$.

- A3** The given input u is chosen such that the signal

$$\tilde{u}(t) = \begin{bmatrix} 1 \\ p(t - \hat{n}_b) \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ p(t) \end{bmatrix} \otimes u(t) \quad (21)$$

is persistently exciting of order $n_y \sum_{i=1}^{\hat{n}_b} (1 + n_p)^i n_u$ [18].

The first assumption, the choice of the scheduling signal, can be made less restrictive, however, it will make the analysis more involved. In this case, we restrict the scheduling signal to be stationary, hence, $\text{Tr}(\mathbb{E}\{\cdot\}) = \text{Tr}(\mathbb{E}\{\cdot\})$ for $V_{\infty,2}$ in (18b). The second assumption is needed to be able to uniquely distinguish all parameters in $\hat{\mathcal{B}}$, which is discussed in the proof of Theorem 5.

As the loss function V_{∞} can be separated as given in (17), we will first focus on the analysis of the loss function w.r.t. the noise model, i.e., $V_{\infty,2}$ (18b). The stationary points of (18b) are solutions of the following equations

$$\frac{\partial V_{\infty,2}(\theta)}{\partial \theta_{ij, \eta_e}} = 0, \quad 1 \leq i, j \leq n_y, \quad \forall \eta_{\hat{c}} \in [\mathbb{I}_0^{n_p}]_2^{\hat{n}_c+1}, \quad (22)$$

where θ_{ij, η_e} only contains the parameters of the noise model. Using the notation of (14), the partial derivative (22) is

$$0 = \text{Tr} \left\{ \frac{\partial \text{Tr} Q(\theta)}{\partial Q(\theta)} \frac{\partial Q(\theta)}{\partial \theta_{ij, \eta_e}} \right\} = 2 \mathbb{E} \left\{ \varepsilon^\top(\theta, t) \frac{\partial \varepsilon(\theta, t)}{\partial \theta_{ij, \eta_e}} \right\}. \quad (23)$$

Taking the partial derivative of (10) w.r.t. the model parameters gives

$$0 = \hat{\mathcal{C}}(\theta, p, q^{-1}) \frac{\partial \varepsilon(\theta, t)}{\partial \theta_{ij, \eta_e}} + s_i s_j^\top p_{\eta_e}(t) \varepsilon(\theta, t - k) \quad (24)$$

for each $t \in \mathbb{Z}$, where $k = \#(\eta_{\hat{c}}) - 1$, s_i is a selector vector for which only the i -th element is non-zero, and $p_{\eta_e}(t)$ defines the product of different scheduling signals and its time-shifts $p_{\eta_e}(t) = \prod_{i=0}^{\#(\eta_{\hat{c}})-1} p_{[\eta_{\hat{c}}]_i}(t-i)$ and, for notational simplicity, $p_0(t) = 1, \forall t \in \mathbb{Z}$. See that (24) is equivalent to

$$\frac{\partial \varepsilon(\theta, t)}{\partial \theta_{ij, \eta_e}} = -\hat{\Gamma}(\theta, p, q^{-1}) s_i s_j^\top p_{\eta_e}(t) \varepsilon(\theta, t - k), \quad (25)$$

and substituting (25) in (23) gives

$$\begin{aligned} 0 &= \mathbb{E} \left\{ \varepsilon^\top(\theta, t) \sum_{r=0}^{\infty} \hat{\Gamma}_r(\theta, p, t) s_i s_j^\top p_{\eta_e}(t-r) \varepsilon(\theta, t-k-r) \right\} \\ &= s_j^\top \mathbb{E} \left\{ \varepsilon(\theta, t-k) \sum_{r=0}^{\infty} \varepsilon^\top(\theta, t+r) \hat{\Gamma}_r(\theta, p, t) p_{\eta_e}(t-r) \right\} s_i, \end{aligned}$$

Hence, we find that (22) is equivalent to:

$$\mathbb{E} \left\{ \sum_{r=0}^{\infty} R_{k+r}(\theta, t) \hat{\Gamma}_r(\theta, p, t) p_{\eta_e}(t-r) \right\} = 0, \quad \forall \eta_{\hat{c}} \in [\mathbb{I}_0^{n_p}]_2^{\hat{n}_c+1}, \quad (26)$$

with $k = \#(\eta_{\hat{c}}) - 1$. We would like to highlight that the analysis given next differs from the LTI case [15] as ε is not independent of p , e.g., $\hat{\Gamma}_r$ and R_{k+r} are not independent in (26). Hence, this dependency makes the analysis more involved. Note that the time-shift operator q is non-communicative if applied on a coefficient function, e.g., $q^{-1} \mathcal{C}_i(p) = \mathcal{C}_i(q^{-1} p) q^{-1}$ [21]. Hence, for the following analysis, the time-instance will be indicated as $\mathcal{C}_j(p, t-i)$ for $\mathcal{C}_j(p(t-i), \dots, p(t-i-j))$. First, decompose R_{k+r} as:

Lemma 3: If A2 holds, then are the matrices R_r for $r \geq 1$ defined in (14c) given by

$$\begin{aligned} \bar{\mathbb{E}} \{ R_r(\theta, t) \} &= \mathbb{E} \left\{ e(t-r) e^\top(t-r) \Omega^\top(\theta, p, t-r) \bar{\Gamma}_r^\top(\theta, p, t) \right\} \\ &+ \mathbb{E} \left\{ \sum_{i=1}^{\infty} \bar{\Gamma}_i(\theta, p, t-r) \Omega(\theta, p, t-i-r) e(t-r-i) \right. \\ &\quad \left. e^\top(t-r-i) \Omega^\top(\theta, p, t-i-r) \bar{\Gamma}_{i+r}^\top(\theta, p, t) \right\} \end{aligned} \quad (27)$$

with

$$\bar{\Gamma}_k(\theta, p, t) = \begin{bmatrix} \hat{\Gamma}_{k-1}(\theta, p, t) & \dots & \hat{\Gamma}_{k-\hat{n}_c}(\theta, p, t) \end{bmatrix}, \quad (28a)$$

$$\Omega(\theta, p, t) = \begin{bmatrix} \mathcal{C}_1(p, t+1) - \hat{\mathcal{C}}_1(\theta, p, t+1) \\ \vdots \\ \mathcal{C}_{n_c}(p, t+n_c) - \hat{\mathcal{C}}_{n_c}(\theta, p, t+n_c) \\ -\hat{\mathcal{C}}_{n_c+1}(\theta, p, t+n_c+1) \\ \vdots \\ -\hat{\mathcal{C}}_{\hat{n}_c}(\theta, p, t+\hat{n}_c) \end{bmatrix}, \quad (28b)$$

where $\Gamma_l(\cdot) = 0$ for $l < 0$. □

Proof: See [22]. ■

Now we return to the uniqueness of $V_{\infty,2}$:

Lemma 4: Given the data generating system (8) with functional dependencies (9), model structure (10) with $\hat{n}_c \geq n_c$, $\mathcal{B}(p, q^{-1}) = \hat{\mathcal{B}}(\theta, p, q^{-1}) \equiv 0$, then, under assumption A2 the global minimum for

$$\min_{\theta} V_{\infty,2}(\theta) = \text{Tr}(\Sigma_e), \quad (29)$$

is the unique stationary point of $V_{\infty,2}$. \square

Proof: See that (27) can be substituted in (26) as e is independent from p . Hence,

$$\begin{aligned} & \mathbb{E} \left\{ \sum_{r=0}^{\infty} \left[e^{(t-r-k)} e^{\top} (t-r-k) \Omega^{\top}(\theta, p, t-r-k) \bar{\Gamma}_{r-k}^{\top}(\theta, p, t) \right. \right. \\ & + \sum_{i=1}^{\infty} \bar{\Gamma}_i(\theta, p, t-r-k) \Omega(\theta, p, t-i-r-k) e^{(t-r-i)} e^{\top} (t-i-r-k) \\ & \left. \left. \Omega^{\top}(\theta, p, t-i-r-k) \bar{\Gamma}_{i+r-k}^{\top}(\theta, p, t) \hat{\Gamma}_r(\theta, p, t) p \eta_c(t-r) \right] \right\} = 0, \end{aligned} \quad (30)$$

for all $\eta_c \in [\mathbb{I}_0^{n_p}]_2^{\hat{n}_c+1}$ and $k = \#(\eta_c) - 1$. Remark that $\hat{h}_{\eta_c} = 0$ cannot be a solution of (30) as $\hat{\Gamma}_0 = I$. Also see that the maximal amount of products of p with equivalent time-shift is $6 + \hat{n}_c$ in (30). Therefore, in A2, we have that all moments of p up to the $6 + \hat{n}_c$ moment are needed to be bounded. The technical details of the remainder of the proof are omitted here, because of the page limitation and can be found at [22]. Broadly, it shows that all elements in (30) are full column rank in the functional sense, only Ω can be rank deficient, with $\Omega = 0$ for $\mathcal{C} \equiv \mathcal{C}$, see (28b). Concluding, $\theta_h = \hat{\theta}_h$ is the only solution for which the set of equations (26) is satisfied. Therefore, it is the only unique stationary point of $V_{\infty,2}$. \blacksquare

Lemma 4 is the LPV extension of the LTI result of [15]. Using this preliminary analysis the following result holds:

Theorem 5: Given the data generating system (8) with functional dependencies (9), model structure (10), $\hat{n}_b \geq n_b$, $\hat{n}_c \geq n_c$, then under assumptions A1-A3 the global minimum

$$\min_{\theta} V_{\infty}(\theta) = \text{Tr}(\Sigma_e), \quad (31)$$

is the only stationary point of V_{∞} and is found if (16) is satisfied. \square

Proof: Recall from (17) that $V_{\infty} = V_{\infty,1} + V_{\infty,2}$. Lemma 4 proofs that for $V_{\infty,2}$ there is only one stationary point. Hence, $\mathcal{C}(p, q^{-1}) = \hat{\mathcal{C}}(\theta, p, q^{-1})$ at this stationary point. The underlying filter $\hat{\mathcal{C}}$ is a polynomial in q^{-1} with parameter-varying coefficients and it is monic; hence it is full rank in the functional sense and, therefore, its inverse $\hat{\Gamma}$ is also full rank. Next, we need to prove that $V_{\infty,1}$ has only one stationary point. If we consider \tilde{u} as the input signal, then the filters \mathcal{B} and $\hat{\mathcal{B}}$ can be written as a multi-input LTI filter. The extended input signal \tilde{u} is persistent of excitation of order $n_y \sum_{i=1}^{\hat{n}_b} (1+n_p)^i n_u$, thus $[\mathcal{B} - \hat{\mathcal{B}}]u$ cannot be zero for $\mathcal{B} \neq \hat{\mathcal{B}}$. Recall, $\hat{\Gamma}(p, q^{-1})$ also cannot lose rank, hence, there exists only one stationary point of $V_{\infty,1}$ satisfying (16), e.g., see [18]. Therefore, the only stationary point of V_{∞} is found for (16). \blacksquare

Theorem 5 proofs that under the conditions A1-A3 there can only exist one stationary point of (12). Hence, identification of the LPV-MAX model (10) is consistent and unbiased.

IV. LPV-SS REALIZATION

The next step in the proposed identification scheme is to realize the LPV-SS model from the identified LPV-MAX

model. By treating the noise as an additional input, i.e., extending $\hat{B}_i = [B_i \ K_i]$ for $i = 0, \dots, n_p$, an isomorphic to the original LPV-SS representation (1)⁵ is obtained by employing the bases reduced Ho-Kalman realization scheme of [16]. This bases reduced realization can considerably decrease the size of the Hankel matrix by selecting only its non-repetitive elements and, therefore, reducing the computational load, compared to realization on the full Hankel matrix [4, Eq. (48)]. In the basis reduced realization, the SVD is only applied on a $n_o \times n_r$ matrix with $n_o, n_r \geq n_x$ instead of a matrix with size $n_y \sum_{l=1}^i (1+n_p)^l \times (n_u+n_y) \sum_{l=1}^j (1+n_p)^l$, for $i, j \geq 2$ in the full realization case. The realization scheme allows to reconstruct the state bases of the LPV-SS model based upon only the process \mathcal{B} or noise model \mathcal{C} , simply by only selecting their corresponding sub-Markov to fill the $\mathcal{H}_{\nu, \varsigma}$ matrix in [16, Eq. (20)]. This can be useful, if, for example, one of both models is poorly estimated then the other model can be used to more accurately reconstruct the reachability and observability matrix.

V. SIMULATION EXAMPLE

The proposed identification scheme is tested on the benchmark model used in [24]. The benchmark contains an MIMO LPV-SS model with input dimension $n_u = 2$, scheduling dimension $n_p = 2$, state dimension $n_x = 2$, and output dimension $n_y = 2$. We added the following parameter independent function

$$\mathcal{K} = \begin{bmatrix} 0.32 & 0.16 \\ 0.64 & 0.24 \end{bmatrix}.$$

The simulation output or one-step-ahead predicted output \hat{y} of the estimated model is compared to measured output or the one-step-ahead predicted output y of the oracle, respectively, by means of the *best fit rate* (BFR)⁶

$$\text{BFR} = \max \left\{ 1 - \frac{\frac{1}{N} \sum_{t=1}^N \|y_t - \hat{y}_t\|_2}{\frac{1}{N} \sum_{t=1}^N \|y_t - \bar{y}\|_2}, 0 \right\} \cdot 100\%, \quad (32)$$

using a validation dataset \mathcal{D}_{val} as in [24]. In (32), \bar{y} defines the mean of the simulation output or one-step-ahead predicted output y of the oracle. In the realization step, the basis reduced scheme uses $n_r = 10$, $n_o = 8$ bases, where the controllability matrix is spanned by $\varsigma = \{(\epsilon, 0, 1), (\epsilon, 0, 2), (\epsilon, 1, 1), (0, 0, 1), (0, 0, 2), (1, 0, 1), (1, 0, 2), (2, 0, 1), (2, 0, 2), (1, 1, 1)\}$ and the observability is spanned by $\nu = \{(1, 0, \epsilon), (2, 0, \epsilon), (1, 0, 0), (2, 0, 0), (1, 0, 1), (2, 0, 1), (1, 0, 2), (2, 0, 2)\}$. The truncation order is $\hat{n}_b = 4$ and $\hat{n}_c = 2$. To evaluate the statistical properties of the identification scheme, $N_{\text{MC}} = 100$ Monte Carlo runs are carried out. In each run, a new realization of the input and scheduling signal is used. We use two different methods of finding the search direction in each of the pseudo linear regression (PLR) iterations: i) ℓ_2 regularized least squares estimate, or ii) enhanced Gaus-Newton optimization method [23]. The extension is trivial and, therefore, is not given here. For the ℓ_2 regularized least squares estimate, we optimized the

⁵There exists a constant, nonsingular transformation matrix $T \in \mathbb{R}^{n_x \times n_x}$ such that: $T\hat{A}_i = A_i T$, $T[\hat{B}_i \ \hat{K}_i] = [B_i \ K_i]$, and $\hat{C}_i = C_i T$, $\forall i \in \mathbb{I}_0^{n_p}$.

⁶Usually the BFR are defined per channel. Eq. (32) are the average performance criteria over all channels.

TABLE I: Mean and standard deviation (std) of the BFR of the identification algorithm per Monte-Carlo run for different $\text{SNR}_y = \{\infty, 40, 10\}$ dB. The BFR is based on the simulated output and one-step-ahead predicted output of the estimated model on the \mathcal{D}_{val} for $N_{\text{MC}} = 100$ Monte-Carlo simulations. GN indicates the enhanced Gauss-Newton optimization method [23]. The PBSID has a past and future window equal to 3.

	LPV-MAX and Regularized LS		LPV-MAX and GN		[1]	
	Simulation	Prediction	Simulation	Prediction	Simulation	Prediction
∞ dB	96.46 (0.5843)	96.46 (0.5840)	96.69 (0.5511)	97.21 (0.4354)	99.50 (0.1164)	99.64 (0.0834)
40 dB	96.52 (0.5954)	97.08 (0.4496)	96.87 (0.5429)	97.17 (0.4163)	99.48 (0.1306)	99.04 (0.0553)
10 dB	90.78 (1.570)	84.98 (1.780)	90.76 (1.662)	84.29 (1.804)	93.06 (0.8757)	85.25 (0.5100)

regularization parameters by a line-search and found $\lambda_1 = 1$, $\lambda_2 = 100$ for $\text{SNR} = \infty$ dB and $\lambda_1 = 0.1$, $\lambda_2 = 1$ for $\text{SNR} = \{40, 10\}$ dB (λ_1 is w.r.t. the process parameters and λ_2 w.r.t. the noise parameters).

Table I displays the mean and standard deviation of the BFR of the identification algorithm for different $\text{SNR}_y = \{\infty, 40, 10\}$ dB. The table indicates that all approaches are capable of identifying the underlying dynamics. However, our approach is mildly outperformed by [1]. The approach of [1] has numerically efficient implementation, in terms of a kernel based approach, which can also be used in our case. In addition, we noticed that the iterations of the PLR are not very robust, as reflected by the increased standard deviation. On the other hand, with this paper, we would like to show how to conceptually reduce the amount of parameters used to identify the input-output model if \mathcal{K} , \mathcal{C} , and \mathcal{D} are parameter varying. Hence, evolving to an numerical efficient implementation is for future research.

VI. CONCLUSIONS

In this paper, an identification scheme to estimate LPV-SS models in innovation form with static and affine dependence on the scheduling signal has been presented. The LPV-SS representation has been reformulated into its corresponding LPV-MAX system, which significantly reduced the amount of parameters of the LPV-IO model compared to other predictor based schemes. We had proven that the LPV-MAX model could consistently be identified using the prediction error minimization framework, under some mild conditions on the input and scheduling signals. The LPV-SS model has been realized from the LPV-MAX model by using the basis reduced Ho-Kalman realization scheme. This realization scheme significantly reduces the amount of parameters needed to realize the LPV-SS model. Hence, the overall scheme decrease the computational load significantly, especially with moderate to large LPV-SS models where \mathcal{K} , \mathcal{C} , and \mathcal{D} are parameter varying. The proposed identification scheme has been illustrated with a simulation example.

REFERENCES

- [1] J. W. van Wingerden and M. Verhaegen, "Subspace identification of bilinear and LPV systems for open- and closed-loop data," *Automatica*, vol. 45, no. 2, pp. 372–381, 2009.
- [2] A. Wills and B. Ninness, "System identification of linear parameter varying state-space models," in *Linear parameter-varying system identification: new developments and trends*, P. Lopes dos Santos, T. P. Azevedo Perdicoulis, C. Novara, J. A. Ramos, and D. E. Rivera, Eds. World Scientific, 2011, ch. 11, pp. 295–316.
- [3] V. Verdult, N. Bergboer, and M. Verhaegen, "Identification of fully parameterized linear and nonlinear state-space systems by projected gradient search," in *Proc. of the 13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands, Aug. 2003, pp. 737–742.
- [4] R. Tóth, H. S. Abbas, and H. Werner, "On the state-space realization of LPV input-output models: practical approaches," *IEEE Trans. on Control Systems Technology*, vol. 20, no. 1, pp. 139–153, Jan. 2012.
- [5] J. W. van Wingerden, I. Houtzager, F. Felici, and M. Verhaegen, "Closed-loop identification of the time-varying dynamics of variable-speed wind turbines," *Int. J. of Robust and Nonlinear Control*, vol. 19, no. 1, pp. 4–21, 2009.
- [6] W. E. Larimore, P. B. Cox, and R. Tóth, "CVA identification of nonlinear systems with LPV state-space models of affine dependence," in *Proc. of the American Control Conf.*, Chicago, IL, USA, Jul. 2015, pp. 831–837.
- [7] M. Tanelli, D. Ardagna, and M. Lovera, "On-and off-line model identification for power management of Web service systems," in *Proc. of the 47th IEEE Conf. on Decision and Control*, Cancun, Mexico, 2008, pp. 4497–4502.
- [8] T. Luspay, B. Kulcsár, J. W. Van Wingerden, and M. Verhaegen, "On the identification of LPV traffic flow model," in *Proc. of the European Control Conf.*, Budapest, Hungary, Aug. 2009, pp. 1752–1757.
- [9] J. Mohammadpour and C. Scherer, Eds., *Control of linear parameter varying systems with applications*. Springer, 2012.
- [10] M. Nemani, R. Ravikanth, and B. A. Bamieh, "Identification of linear parametrically varying systems," in *Proc. of the 34th IEEE Conf. on Decision and Control*, New Orleans, LA, USA, 1995, pp. 2990–2995.
- [11] P. Gáspár, Z. Szabó, and J. Bokor, "A grey-box identification of an LPV vehicle model for observer-based side-slip angle estimation," in *Proc. of the American Control Conf.*, New York City, USA, Jul. 2007, pp. 2961–2965.
- [12] C. Novara, "SM Identification of State-Space LPV systems," in *Linear parameter-varying system identification: new developments and trends*, P. Lopes dos Santos, T. P. Azevedo Perdicoulis, C. Novara, J. A. Ramos, and D. E. Rivera, Eds. World Scientific, 2011, ch. 11, pp. 65–93.
- [13] W. E. Larimore, "Identification of nonlinear parameter-varying systems via canonical variate analysis," in *Proc. of the American Control Conf.*, Washington, DC, USA, Jun 2013, pp. 2247–2262.
- [14] P. Lopes dos Santos, J. A. Ramos, and J. L. M. de Carvalho, "Identification of linear parameter varying systems using an iterative deterministic-stochastic subspace approach," in *Proc. of the European Control Conf.*, Kos, Greece, Jul 2007, pp. 4867–4873.
- [15] P. Stoica and T. Söderström, "Uniqueness of prediction error estimates of multivariable moving average models," *Automatica*, vol. 18, no. 5, pp. 617–620, 1982.
- [16] P. B. Cox, R. Tóth, and M. Petreczky, "Estimation of LPV-SS models with static dependency using correlation analysis," in *Proc. of the 1st IFAC Workshop on Linear Parameter Varying Systems*, Grenoble, France, Oct. 2015, pp. 91–96.
- [17] R. Tóth, *Modeling and identification of linear parameter-varying systems*. Springer, 2010.
- [18] L. Ljung, *System identification: theory for the user*. Springer, 1989.
- [19] R. Tóth, P. S. C. Heuberger, and P. M. J. Van den Hof, "Prediction-error identification of LPV systems: present and beyond," in *Control of Linear Parameter Varying Systems with Applications*, J. Mohammadpour and C. W. Scherer, Eds. Springer, 2012, ch. 2, pp. 27–58.
- [20] L. Ljung, T. Söderström, and I. Gustavsson, "Counterexamples to general convergence of a commonly used recursive identification method," *IEEE Trans. on Automatic Control*, vol. 20, no. 5, pp. 643–652, 1975.
- [21] R. Tóth, "Identification problems in the context of compressive sensing," Tech. Rep., 2010.
- [22] P. B. Cox and R. Tóth, "Alternative form of predictor based identification of LPV-SS models with innovation noise," arXiv: 1674538v1.
- [23] A. Wills and B. Ninness, "On gradient-based search for multivariable system estimates," *IEEE Trans. on Automatic Control*, vol. 53, no. 1, pp. 298–306, Feb. 2008.
- [24] V. Verdult and M. Verhaegen, "Kernel methods for subspace identification of multivariable LPV and bilinear systems," *Automatica*, vol. 41, pp. 1557–1565, 2005.