Automatica 123 (2021) 109296

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Linear parameter-varying subspace identification: A unified framework☆

Pepijn Bastiaan Cox [a,*], Roland Tóth [a,b]

[a] Control Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

[b] Systems and Control Laboratory, Institute for Computer Science and Control, Kende u. 13-17, H-1111 Budapest, Hungary

ARTICLE INFO

Article history:
Received 23 April 2020
Received in revised form 11 August 2020
Accepted 9 September 2020
Available online 18 November 2020

Keywords:
System identification
Linear parameter-varying systems
Subspace methods
State–space representations
Realization theory

ABSTRACT

In this paper, we establish a unified framework for subspace identification (SID) of linear parameter-varying (LPV) systems to estimate LPV state–space (SS) models in innovation form. This framework enables us to derive novel LPV SID schemes that are extensions of existing linear time-invariant (LTI) methods. More specifically, we derive the open-loop, closed-loop, and predictor-based data-equations (input–output surrogate forms of the SS representation) by systematically establishing an LPV subspace identification theory. We also show the additional challenges of the LPV setting compared to the LTI case. Based on the data-equations, several methods are proposed to estimate LPV-SS models based on a maximum-likelihood or a realization based argument. Furthermore, the established theoretical framework for the LPV subspace identification problem allows us to lower the number of to-be-estimated parameters and to overcome dimensionality problems of the involved matrices, leading to a decrease in the computational complexity of LPV SIDs in general. To the authors' knowledge, this paper is the first in-depth examination of the LPV subspace identification problem. The effectiveness of the proposed subspace identification methods are demonstrated and compared with existing methods in a Monte Carlo study of identifying a benchmark MIMO LPV system.

1. Introduction

Realization based state–space identification techniques, so-called *subspace identification* (SID) methods, have been successfully applied in practice to estimate time-varying and/or nonlinear dynamical systems using *linear parameter-varying* (LPV) *state–space* (SS) models. Successful application examples range from diesel engines (Schulz, Bussa, & Werner, 2016), wind-turbines (Felici, van Wingerden, & Verhaegen, 2007; van Wingerden, Houtzager, Felici, & Verhaegen, 2009), gas pipelines (Lopes Dos Santos, Azevedo-Perdicoulis, & Ramos, 2010), traffic flow models (Luspay, Kulcsár, Van Wingerden, & Verhaegen, 2009), and bioreactors (Verdult, Ljung, & Verhaegen, 2002) to nonlinear benchmark systems like the Lorenz attractor (Larimore, Cox, & Tóth, 2015). The existing techniques are based on *predictor based subspace identification* (PBSID) (van Wingerden & Verhaegen, 2009), *past-output multivariable output-error state–space* (PO-MOESP) (Felici et al., 2007), *canonical variate analysis* (CVA) (Larimore et al., 2015), or the successive approximation identification algorithm (Lopes Dos Santos et al., 2010). However, these methods lack a common unified theory to tackle the LPV SID problem.

The field of subspace identification applies realization theory to find SS model estimates based on surrogate *input–output* (IO) models (with appropriate noise models) estimated form data. These specialized IO models are estimated by using convex optimization and it can be shown that they correspond to a *maximum-likelihood* (ML) estimate under the considered assumptions. Then, an SS realization is obtained from the IO model by either a *direct realization* step or by an intermediate *projection* step. In the latter idea, a projection is found to estimate the unknown state-sequence via matrix decomposition methods, then the SS matrices are estimated in a least-squares fashion. Obtaining such a state-sequence is heavily based on realization theory, as the estimated state-basis should be consistent with the behavior of the underlying system. In the LTI setting, the IO model estimation and realization of the SS model under the presence of process and measurement noise is well understood (Katayama, 2005; Lindquist & Picci, 2015; Van Overschee & De Moor, 1996; Verhaegen & Verdult, 2007). In the LPV case, contrary to the LTI

☆ This paper has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 714663). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Juan I. Yuz under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: p.b.cox@tue.nl (P.B. Cox), r.toth@tue.nl (R. Tóth).

https://doi.org/10.1016/j.automatica.2020.109296

setting, the stochastic interpretation of the methods with the appropriate noise representation is not well understood neither is the connection between various methods ever been studied.

LPV subspace schemes also suffer heavily from the curse of dimensionality, e.g., see van Wingerden and Verhaegen (2009, Table 1), resulting in ill-conditioning of the estimation problem and high parameter variance. Consequently, two common assumptions are taken to reduce the dimensionality: (i) the excitation, in terms of the variation of the scheduling variable p, is periodic or white (Felici et al., 2007; van Wingerden et al., 2009), and/or (ii) the output-equation of the SS representation is assumed to be p-independent (Luspay et al., 2009; Schulz et al., 2016; van Wingerden & Verhaegen, 2009). However, such assumptions restrict practical applicability of the methods. To tackle ill-conditioning and to reduce estimation variance, kernel based regularization techniques have been proposed (van Wingerden & Verhaegen, 2009; Verdult & Verhaegen, 2005). However, computational complexity of the involved kernels grows polynomially or exponentially w.r.t. the design parameters, which significantly compromises the effectiveness of these schemes. Alternatively, SS models can directly be estimated by minimization of the ℓ_2-loss in terms of the prediction-error associated with the model. These so-called *prediction-error methods* (PEM) minimize the ℓ_2-loss directly using gradient-based methodologies (Lee & Poolla, 1997; Verdult, Bergboer, & Verhaegen, 2003; Wills & Ninness, 2008, 2011) or by the expectation–maximization scheme (Wills & Ninness, 2011). However, minimization of the ℓ_2-loss w.r.t. to the LPV-SS model parameters is a nonlinear and nonunique optimization problem, requiring an initial estimate close to the global optimum.

The goal of this paper is to obtain a unified formulation to treat the LPV subspace identification problem and derive its associated stochastic properties by systematically establishing an LPV SID theory. This unified framework enables us to (i) understand relations and performance of LPV SIDs, (ii) extend most of the successful LTI subspace schemes to the LPV setting, (iii) decrease the dimensionality problems, and (iv) relax assumptions on the scheduling signal. In addition, we establish stochastic LPV realization theory which provides state estimation with maximum likelihood efficiency. To the authors' knowledge, this paper is the first in-depth treatment of the subspace theory in the LPV case. In this paper, we focus on projection based schemes, but the direct realization schemes can easily be abstracted from the developed results, i.e., see Cox (2018). We follow well-known concepts from the LTI literature, e.g., Van Overschee and De Moor (1996), Verhaegen and Verdult (2007), and our theoretic results are also based on preliminary studies in the LPV setting (van Wingerden & Verhaegen, 2009; Verdult, 2002; Verdult & Verhaegen, 2005). The main contributions of this paper are:

(i) Formulating the state estimation problem by a maximum-likelihood approach based on canonical correlation analysis and by a realization based approach.
(ii) Stochastic interpretation of state estimation with maximum likelihood efficiency under the presence of noise.
(iii) Computationally efficient formulation of SIDs to decrease the effects of the curse of dimensionality.

The unified subspace theory is tackled in the global identification setting, i.e., under general trajectories of the scheduling signal, contrary to some results in the literature (Felici et al., 2007; Lopes dos Santos, Ramos, & Martins de Carvalho, 2008; van Wingerden, Felici, & Verhaegen, 2007).

The proposed schemes based on a realization argument could also be applied in a setting where the scheduling signal is affected by a noise that might be correlated to the measured input and output signals. This often occurs if the scheduling p is a function of the measured input or output signals. In such case, the IO estimation step could be performed by using an instrumental variable approach (Tóth, Laurain, Gilson and Garnier, 2012). However, investigation of such formulation is outside of the scope of this paper.

This paper is organized as follows: first, the assumed data-generating system with LPV-SS representation and general innovation noise structure are presented and the open-loop, closed-loop, and predictor-based data-equations are derived (Section 2). Then, the considered parametric LPV-SS identification problem is introduced (Section 3). Next, the state realization problem is tackled based both on a maximum-likelihood and a realization argument. This is accomplished first for the open-loop (Section 4) and then for the closed-loop identification setting (Section 5), leading to the LPV formulation of various well-known LTI subspace methods. The efficiency of the unified framework is demonstrated by a Monte Carlo study on an LPV-SS identification benchmark (Section 6).

2. The LPV data-equations

In this section, surrogate input–output representations of SS models are formulated which are key in solving the subspace identification problem. Namely, we derive the LPV open-loop data-equation (Section 2.2), closed-loop data-equation (Section 2.3), and the predictor-based data-equation (Section 2.4) for LPV data-generating systems in an SS form (Section 2.1).

2.1. The data-generating system

The goal is to obtain an SS model estimate of the data-generating system $\mathcal{S}_o$ represented in the following LPV-SS *innovation form*[1]

$$x_{t+1} = \mathcal{A}(p_t)x_t + \mathcal{B}(p_t)u_t + \mathcal{K}(p_t)\xi_t, \tag{1a}$$

$$y_t = \mathcal{C}(p_t)x_t + \mathcal{D}(p_t)u_t + \xi_t, \tag{1b}$$

where $x : \mathbb{Z} \to \mathbb{X} = \mathbb{R}^{n_x}$ is the state variable, $y : \mathbb{Z} \to \mathbb{Y} = \mathbb{R}^{n_y}$ is the measured output signal, $u : \mathbb{Z} \to \mathbb{U} = \mathbb{R}^{n_u}$ is the input signal, $p : \mathbb{Z} \to \mathbb{P} \subset \mathbb{R}^{n_p}$ is the scheduling signal, and $\xi : \mathbb{Z} \to \mathbb{R}^{n_y}$ is the sample path realization of the zero-mean stationary process:

$$\boldsymbol{\xi}_t \sim \mathcal{N}(0, \Xi^2), \tag{2}$$

where $\boldsymbol{\xi}_t : \Omega \to \mathbb{R}^{n_y}$ is a white noise process with sample space Ω (set of possible outcomes) and $\Xi^2 \in \mathbb{R}^{n_x \times n_x}$ is a positive definite covariance matrix. Furthermore, we will assume u, p, y, ξ to have left compact support to avoid technicalities with initial conditions. The matrix functions $\mathcal{A}(\cdot), \ldots, \mathcal{K}(\cdot)$ defining the SS representation (1) are affine combinations of bounded scalar

[1] In the majority of the subspace literature (Van Overschee & De Moor, 1996; van Wingerden et al., 2009; Verhaegen & Verdult, 2007), the data-generating system is assumed to be in the innovation form as given in (1). However, in Cox (2018), it is shown that the noise description in (1) is not equivalent to a state–space form with general noise representation, i.e., a representation with different noise processes on the state and output equation. Cox (2018) also shows that a static, affine $\mathcal{K}(p_t)$ can approximate the general setting if the state dimension is increased. In practice, we often need to restrict parameterization of $\mathcal{K}$, e.g., to the static, affine parameterization in (1), to reduce complexity of the estimation method and variance of the model estimates. Hence, despite the possible increase of the state order of the equivalent innovation form, the usage of this affine form has been found adequate in practical applications (Felici et al., 2007; Luspay et al., 2009; Schulz et al., 2016; van Wingerden et al., 2009).

functions $\psi^{[i]}(\cdot) : \mathbb{P} \to \mathbb{R}$:

$$\begin{aligned}\mathcal{A}(p_t) &= \sum_{i=0}^{n_\psi} A_i \psi^{[i]}(p_t), \quad \mathcal{B}(p_t) = \sum_{i=0}^{n_\psi} B_i \psi^{[i]}(p_t),\\ \mathcal{C}(p_t) &= \sum_{i=0}^{n_\psi} C_i \psi^{[i]}(p_t), \quad \mathcal{D}(p_t) = \sum_{i=0}^{n_\psi} D_i \psi^{[i]}(p_t),\\ \mathcal{K}(p_t) &= \sum_{i=0}^{n_\psi} K_i \psi^{[i]}(p_t),\end{aligned} \tag{3}$$

where $\{A_i, B_i, C_i, D_i, K_i\}_{i=0}^{n_\psi}$ are constant, real matrices with appropriate dimensions and $\psi^{[0]}(\cdot) = 1$ is assumed to be constant. Additionally, for well-posedness, it is assumed that $\{\psi^{[i]}\}_{i=1}^{n_\psi}$ are linearly independent over an appropriate function space and are normalized w.r.t. an appropriate norm or inner product (Tóth, Abbas and Werner, 2012). Due to the freedom to consider arbitrary functions $\psi^{[i]}$, (3) can capture a wide class of static nonlinearities and time-varying behaviors. For notational simplicity, we define $\psi_t = [\psi^{[0]}(p_t) \quad \dots \quad \psi^{[n_\psi]}(p_t)]$.

2.2. The open-loop data-equation

The first step in tackling the subspace identification problem is to represent the dynamics of the data-generating system (1) as an equivalent IO representation, the so-called *data-equation*. The unknowns in these data-equations are estimated by convex optimization and the final SS model is obtained from these data-equations using matrix decomposition techniques (see Section 3–5 for more details). Hence, the data-equations are key in formulating the subspace problem.

Open-loop data-equations are rarely used in the literature, as the innovation noise ξ_t is unknown. In light of the MAX identification setting in Cox (2018) and Cox and Tóth (2016), the innovation noise ξ_t can be uniquely obtained by convex optimization, which renders the open-loop equations attractive for further investigation, similar to Mercère, Markovsky, and Ramos (2016) in the LTI setting. Using (1b), the output w.r.t. a *future window* $f \in \mathbb{N}_+$, where $\mathbb{N}_+ = \{i \in \mathbb{Z} \mid i > 0\}$, starting from time-instance t can be written as

$$\mathrm{y}_t^{t+f} = (\mathcal{O}_f \diamond p)_t x_t + (\breve{\mathcal{L}}_f \diamond p)_t \breve{\mathrm{z}}_t^{t+f} + \xi_t^{t+f}, \tag{4}$$

where $\breve{z}_t = [u_t^\top \ \xi_t^\top]$ is the extended "input" signal and y_t^{t+f}, ξ_t^{t+f}, and $\breve{\mathrm{z}}_t^{t+f}$ are sequences according to the notation

$$\mathrm{q}_l^s = \begin{cases} \begin{bmatrix} q_l^\top & q_{l+1}^\top & \cdots & q_{s-1}^\top \end{bmatrix}^\top & \text{if } s > l, \\ \begin{bmatrix} q_{l-1}^\top & \cdots & q_{s+1}^\top & q_s^\top \end{bmatrix}^\top & \text{if } s < l. \end{cases}$$

Furthermore, the matrix functions in (4) are given as

$$(\mathcal{O}_f \diamond p)_t = \begin{bmatrix} \mathcal{C}^\top(p_t) & \cdots & \left(\mathcal{C}(p_{t+f}) \prod_{i=1}^{f} \mathcal{A}(p_{t+f-i})\right)^\top \end{bmatrix}^\top, \tag{5a}$$

$$\breve{\mathcal{B}}(p_t) = \begin{bmatrix} \mathcal{B}(p_t) & \mathcal{K}(p_t) \end{bmatrix}, \tag{5b}$$

$$\breve{\mathcal{D}}(p_t) = \begin{bmatrix} \mathcal{D}(p_t) & 0_{n_y \times n_y} \end{bmatrix}, \tag{5c}$$

and $\breve{\mathcal{L}}_f$ is as given in (5d) inside Box I where $\mathcal{A}_t, \dots, \breve{\mathcal{D}}_t$ is a shorthand notation for $\mathcal{A}(p_t), \dots, \breve{\mathcal{D}}(p_t)$. Here, $\prod_{i=1}^{f}$ is considered with left multiplication. In (4) and (5a)–(5d), the $\diamond$ operator is a shorthand notation for dynamic dependency on the scheduling signal, i.e., $(\mathcal{O}_f \diamond p)_t = \mathcal{O}_f(p_t, p_{t-1}, p_{t-2}, \dots)$.

Next, the state can be decomposed by using the past values of the input and noise signals:

$$x_t = (\breve{\mathcal{R}}_p \diamond p)_t \breve{\mathrm{z}}_t^{t-p} + \mathcal{X}_p, \tag{6}$$

with *past window* $p \in \mathbb{N}_+$, past data $\breve{\mathrm{z}}_t^{t-p}$, and

$$(\breve{\mathcal{R}}_p \diamond p)_t = \Bigg[\breve{\mathcal{B}}(p_{t-1}) \quad \mathcal{A}(p_{t-1})\breve{\mathcal{B}}(p_{t-2}) \quad \cdots \quad \left[\prod_{i=1}^{p-1} \mathcal{A}(p_{t-i})\right] \breve{\mathcal{B}}(p_{t-p})\Bigg], \tag{7a}$$

$$\mathcal{X}_p = \left[\prod_{i=1}^{p} \mathcal{A}(p_{t-i})\right] x_{t-p}. \tag{7b}$$

Combining the output-equation based on the future values (4) with the state-equation based on the past values (6) results in the *open-loop data-equation*

$$\mathrm{y}_t^{t+f} = (\mathcal{O}_f \breve{\mathcal{R}}_p \diamond p)_t \breve{\mathrm{z}}_t^{t-p} + (\breve{\mathcal{L}}_f \diamond p)_t \breve{\mathrm{z}}_t^{t+f} + \xi_t^{t+f} + (\mathcal{O}_f \diamond p)_t \mathcal{X}_p, \tag{8}$$

which has the form of a MIMO LPV-IO model. Estimating the underlying IO relationship of (8) requires the input-scheduling pair (u, p) and the innovation noise ξ to be uncorrelated in order to obtain an unbiased estimate of the relationship (8) under PEM, e.g., see Chiuso (2007), Jansson (2005) and Verhaegen and Verdult (2007). The case when (u, p) and ξ are uncorrelated is usually referred to as the open-loop identification setting (Eykhoff, 1974; Ljung, 1999; Verhaegen & Verdult, 2007), characterized by the following two assumptions:

- **A.1** The input signal u is quasi-stationary and uncorrelated with ξ, i.e., $\bar{\mathbb{E}}\{u_t(\xi_{t+\tau})^\top\} = \bar{\mathbb{E}}\{u_t(\xi_{t-\tau})^\top\} = 0$ for all $\tau \in \mathrm{N}_0$.[2]
- **A.2** The scheduling signal p is quasi-stationary and uncorrelated with ξ.

Assumptions A.1 and A.2 are not restricting, for example, when considering the LPV modeling problem of a thermal loop in a wafer scanner. The thermal distribution of the wafer varies with the position, but it does not influence the measurement noise of the position sensor and, therefore, the position as scheduling signal fulfills Assumption A.2. On the other hand, an inverted pendulum setup with stabilizing controller where the angle of the pendulum is the scheduling signal (and output) will not satisfy Assumptions A.1–A.2. In such a case, p is correlated with past values of ξ due to the closed-loop interconnection between plant and controller.

2.3. The closed-loop data-equation

To overcome the limitations of the open-loop setting, the data-equation (8) can be written in an alternative form. Analogously to LTI identification (Chiuso, 2007; Jansson, 2005; Verhaegen & Verdult, 2007), the output-equation (1b) is substituted in the state-equation (1a), resulting in

$$x_{t+1} = \tilde{\mathcal{A}}(p_t)x_t + \tilde{\mathcal{B}}(p_t)\tilde{z}_t, \tag{9}$$

where $\tilde{z}_t = [u_t^\top \ y_t^\top]^\top$ and the corresponding matrix functions are

$$\tilde{\mathcal{A}}(p_t) = \mathcal{A}(p_t) - \mathcal{K}(p_t)\mathcal{C}(p_t), \tag{10a}$$

$$\tilde{\mathcal{B}}(p_t) = [\mathcal{B}(p_t) - \mathcal{K}(p_t)\mathcal{D}(p_t) \quad \mathcal{K}(p_t)]. \tag{10b}$$

It is important to note that (9) does not depend explicitly on the stochastic process ξ. Hence, the state-equation (9) can be

[2] The generalized expectation operation $\bar{\mathbb{E}}$ of a process u is defined as $\bar{\mathbb{E}}\{u_t\} = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathbb{E}\{u_t\}$. A process u is said to be *quasi-stationary* if there exists finite $c_1, c_2 \in \mathbb{R}$ such that (i) $\|\mathbb{E}\{u_t\}\|_2 < c_1$ for all t, and (ii) $\left\|\mathrm{Tr}\left(\bar{\mathbb{E}}\{u_t u_{t-\tau}^\top\}\right)\right\|_2 < c_2$ for all τ, e.g., see Ljung (1999).

$$(\check{\mathcal{L}}_f \diamond p)_t = \begin{bmatrix} \check{\mathcal{D}}_t & 0 & 0 & \cdots & 0 \\ \mathcal{C}_{t+1}\check{\mathcal{B}}_t & \check{\mathcal{D}}_{t+1} & 0 & \cdots & 0 \\ \mathcal{C}_{t+2}\mathcal{A}_{t+1}\check{\mathcal{B}}_t & \mathcal{C}_{t+2}\check{\mathcal{B}}_{t+1} & \check{\mathcal{D}}_{t+2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathcal{C}_{t+f-1}\left[\prod_{i=2}^{f-1}\mathcal{A}_{t+f-i}\right]\check{\mathcal{B}}_t & \mathcal{C}_{t+f-1}\left[\prod_{i=2}^{f-2}\mathcal{A}_{t+f-i}\right]\check{\mathcal{B}}_{t+1} & \mathcal{C}_{t+f-1}\left[\prod_{i=2}^{f-3}\mathcal{A}_{t+f-i}\right]\check{\mathcal{B}}_{t+2} & \cdots & \check{\mathcal{D}}_{t+f-1} \end{bmatrix}. \tag{5d}$$

Box I.

treated in a deterministic setting. However, moving from the open-loop to the closed-loop dynamics comes at the cost of polynomial dependency of the $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$ matrix functions. Opposed to the LTI setting where $\mathcal{K} = K \in \mathbb{R}^{n_x \times n_y}$, applying (9) instead of (1a) increases the model complexity. Using (9), the stacked output-equation (4) can equivalently be represented as

$$\mathrm{y}_t^{t+f} = (\tilde{\mathcal{O}}_f \diamond p)_t x_t + (\tilde{\mathcal{L}}_f \diamond p)_t \tilde{\mathrm{z}}_t^{t+f} + \xi_t^{t+f}. \tag{11}$$

In (11), $(\tilde{\mathcal{O}}_f \diamond p)_t$ denotes the observability matrix with $\tilde{\mathcal{A}}$ instead of $\mathcal{A}$, $(\tilde{\mathcal{L}}_f \diamond p)_t$ is constructed with $\tilde{\mathcal{A}}$, $\tilde{\mathcal{B}}$, and the future values $\tilde{\mathrm{z}}_t^{t+f}$ are similarly stacked as $\check{\mathrm{z}}_t^{t+f}$ in (4). Note that $\check{\mathrm{z}}_t^{t+f}$ is dependent on the pair (u_t, ξ_t) and $\tilde{\mathrm{z}}_t^{t+f}$ on (u_t, y_t). The state can be written as a combination of past signals (similar to (6))

$$x_t = (\tilde{\mathcal{R}}_p \diamond p)_t \tilde{\mathrm{z}}_t^{t-p} + \tilde{\mathcal{X}}_p, \tag{12}$$

where $p \in \mathbb{N}_+$ is the past window, $(\tilde{\mathcal{R}}_f \diamond p)_t$ denotes the reachability matrix (7a) with $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$ instead of $\mathcal{A}$ and $\mathcal{B}$, and $\tilde{\mathcal{X}}_p$ is the initial condition (7b) with $\tilde{\mathcal{A}}$ instead of $\mathcal{A}$.

Combining (11) and (12) results in the *closed-loop data-equation*:

$$\mathrm{y}_t^{t+f} = (\tilde{\mathcal{O}}_f \tilde{\mathcal{R}}_p \diamond p)_t \tilde{\mathrm{z}}_t^{t-p} + (\tilde{\mathcal{L}}_f \diamond p)_t \tilde{\mathrm{z}}_t^{t+f} + \xi_t^{t+f} + (\tilde{\mathcal{O}}_f \diamond p)_t \tilde{\mathcal{X}}_p. \tag{13}$$

To formulate our identification problem in the closed-loop case, we take the following assumptions:

- **A.3** The input signal u is quasi-stationary and uncorrelated with future values of ξ, i.e., $\bar{\mathbb{E}}\{u_t(\xi_{t+\tau})^\top\} = 0$ for all $\tau \in \mathbb{N}_0$.
- **A.4** The scheduling signal p is quasi-stationary and uncorrelated with future values of ξ.

Assumptions A.3 and A.4 allow to identify systems under general feedback structures, e.g., see Eykhoff (1974), Ljung (1999) and Verhaegen and Verdult (2007).

2.4. Derivation of the predictor

A commonly applied data-equation for subspace identification is the predictor form, e.g., see Chiuso (2007), Chiuso and Picci (2005) and van Wingerden and Verhaegen (2009). To formulate the one-step-ahead predictor for the output, the closed-loop state (9) is substituted into the output-equation (1b) and we take the conditional expectation, resulting in:

$$\hat{y}_{t|t-1} = \mathcal{C}(p_t)\tilde{\mathcal{X}}_p + \mathcal{D}(p_t)u_t + \sum_{i=1}^{p} \mathcal{C}(p_t)\left[\prod_{j=1}^{i-1}(\tilde{\mathcal{A}} \diamond p)_{t-j}\right](\tilde{\mathcal{B}} \diamond p)_{t-i}\tilde{z}_{t-i}. \tag{14}$$

Note that (14) is the minimal variance estimator of y_t and that (14) represents an LPV-ARX model where $p \to \infty$ will diminish the influence of the initial condition $\tilde{\mathcal{X}}_p$ under the assumption that $\tilde{\mathcal{A}}$ is stable. The one-step-ahead predictor of the output can be similarly stacked as the closed-loop data-equation (13) leading to the *predictor-based data-equation*:

$$\hat{\mathrm{y}}_{t|t-1}^{t+f|t+f-1} = (\tilde{\mathcal{O}}_f \tilde{\mathcal{R}}_p \diamond p)_t \tilde{\mathrm{z}}_t^{t-p} + (\tilde{\mathcal{L}}_f \diamond p)_t \tilde{\mathrm{z}}_t^{t+f} + (\tilde{\mathcal{O}}_f \diamond p)_t \tilde{\mathcal{X}}_p. \tag{15}$$

Note that (15) is the one-step-ahead predictor of (13). Hence, the SS representation of $\mathcal{S}_o$ can be captured by the predictor (14) from which (15) can be constructed (Chiuso, 2007; Chiuso & Picci, 2005; van Wingerden et al., 2009; van Wingerden & Verhaegen, 2009).

3. Parametric subspace identification setting

Known LTI and LPV subspace schemes are based on the aforementioned data-equations or their simplifications. The subspace schemes rely on matrix decomposition techniques applied on $\mathcal{O}_f\mathcal{R}_p$ to obtain a realization of these two matrices; however, these decomposition techniques cannot be directly applied to parameter-varying matrices. As shown in van Wingerden and Verhaegen (2009), the main difficulty comes from the time-varying observability matrix, as the dependency structure of the reachability matrix can be absorbed in an extended input vector.

In this paper, we are interested in estimating the unknown matrices $\{A, \ldots, K_i\}_{i=0}^{n_\psi}$ corresponding to parameters $\theta_A = [\mathrm{vec}\{A_0\}^\top \ \cdots \ \mathrm{vec}\{A_{n_\psi}\}^\top]^\top$. The collection of unknown parameters is denoted by $\theta = [\theta_A^\top \ \cdots \ \theta_K^\top]^\top$ with $\theta \in \Theta = \mathbb{R}^{n_\theta}$ and $n_\theta = (1 + n_\psi)(n_x^2 + 2n_y n_x + n_u n_x + n_y n_u)$. The parameters of the data-generating system $\mathcal{S}_o$ are denoted as θ_o and we denote with $\mathcal{S}(\theta')$ the model (1) with parameters θ'. The identification problem of SS models based on a *data set* $\mathcal{D}_N = \{(y_t, p_t, u_t)\}_{t=1}^N$ has non-unique solutions up to a transformation matrix, e.g., see Cox (2018) and Verdult (2002). Hence, we aim at identifying an isomorphic, jointly state minimal $\mathcal{S}(\theta)$ w.r.t. $\mathcal{S}(\theta_0)$ defined by the following set[3]:

$$\mathcal{I}_\theta = \left\{ \theta' \;\middle|\; \exists T \in \mathbb{R}^{n_x \times n_x} \text{ s.t. } \mathrm{rank}(T) = n_x \text{ and } \theta' = S(\theta, T) \right\}, \tag{16}$$

where the indistinguishable manifold S is given in (17) inside Box II.

Given a data set $\mathcal{D}_N$ and the basis functions $\{\psi^{[i]}\}_{i=0}^{n_\psi}$, the goal of this paper is obtain a consistent estimate $\hat{\theta}$ of the data-generating system $\mathcal{S}_o$ such that $\hat{\theta} \to \theta \in \mathcal{I}_{\theta_o}$ with probability one as $N \to \infty$. For the identification setting to be well-posed, the following standard assumptions are taken

- **A.5** $\mathcal{S}(\theta_o)$ is an element of the model set, meaning that $\exists \theta \in \Theta$ such that $\theta \in \mathcal{I}_{\theta_o}$.

[3] The representation $\mathcal{S}$ is jointly state minimal if $\check{\mathcal{R}}_{n_x}$ and $\mathcal{O}_{n_x}$ have at least n_x linearly independent rows or columns, respectively, in a function sense, i.e., $\mathrm{rank}(\check{\mathcal{R}}_{n_x}) = n_x$ and $\mathrm{rank}(\mathcal{O}_{n_x}) = n_x$.

$$S(\theta,T)=\begin{bmatrix}\mathrm{vec}\{T^{-1}A_0T\}^\top & \cdots & \mathrm{vec}\{T^{-1}A_{n_\psi}T\}^\top & \mathrm{vec}\{T^{-1}B_0\}^\top & \cdots & \mathrm{vec}\{T^{-1}B_{n_\psi}\}^\top & \mathrm{vec}\{C_0T\}^\top & \cdots & \mathrm{vec}\{C_{n_\psi}T\}^\top\\ \mathrm{vec}\{D_0\}^\top & \cdots & \mathrm{vec}\{D_{n_\psi}\}^\top & \mathrm{vec}\{T^{-1}K_0\}^\top & \cdots & \mathrm{vec}\{T^{-1}K_{n_\psi}\}^\top\end{bmatrix}^\top. \tag{17}$$

Box II.

A.6 The state-minimal SS representation with static, basis affine dependency structure of the system $\mathcal{S}_o$ is structurally state-observable w.r.t. to the pair $(\mathcal{A}(p_t),\mathcal{C}(p_t))$ and structurally state-reachable w.r.t. to the pair$(\mathcal{A}(p_t),[\mathcal{B}(p_t)\ \mathcal{K}(p_t)\Xi^{-1}])$ (Cox, 2018, Lem. 2.4).

A.7 The open-loop dynamics $\mathcal{A}(p_t)$ or closed-loop dynamics $\tilde{\mathcal{A}}(p_t)$ are asymptotically stable for the open-loop or closed-loop cases, respectively.

A.8 The past window p is chosen sufficiently large, such that $\mathcal{X}_p\approx 0$ or $\tilde{\mathcal{X}}_p\approx 0$, $\forall p\in\mathbb{P}^{\mathbb{Z}}$ for the open-loop or closed-loop cases, respectively.

We can only estimate system dynamics that manifest in the data, so the system is represented with a structurally minimal IO equivalent SS representations, as formalized in Assumption A.6. With Assumption A.7, the influence of the initial state x_{t-p} can be neglected in (8), (13), or (15). This property is widely applied in subspace identification (Chiuso & Picci, 2005; Jansson, 2003; Van Overschee & De Moor, 1996; van Wingerden & Verhaegen, 2009; Verdult & Verhaegen, 2002). See Verdult and Verhaegen (2002, Lemma 5) for an upper-bound on the approximation error of this assumption. Note that we do not assume that either $(\mathcal{C}(p),\mathcal{D}(p))$ or $\mathcal{K}(p)$ are parameter independent to reduce the complexity of the IO model opposed to state-of-the-art LPV subspace schemes (van Wingerden & Verhaegen, 2009; Verdult & Verhaegen, 2005).

Next, we will develop a unified theory to extend the LTI N4SID, MOESP, CVA, SS-ARX, and PBSID principles to the LPV case. There are two significant differences with respect to the LTI case. Firstly, almost all LTI formulations apply a (partial) ARX model structure, however, in the LPV case, the LPV-ARX model comes with significantly larger parameterization compared to the MAX representation in the open-loop setting. Secondly, we apply a predictor pre-estimation step to identify the unknown quantities of the matrices $\mathcal{O}_f\check{\mathcal{R}}_p$, $\check{\mathcal{L}}_f$, $\tilde{\mathcal{O}}_f\tilde{\mathcal{R}}_p$, etc. and construct the full matrices instead of estimating the matrices $\mathcal{O}_f\check{\mathcal{R}}_p$, $\check{\mathcal{L}}_f$, $\tilde{\mathcal{O}}_f\tilde{\mathcal{R}}_p$ directly using the data-equations (8) or (13). Direct estimation of the matrices by oblique projections comes with a significant computational cost (Verdult & Verhaegen, 2002, Table 1) compared to the predictor formulation (van Wingerden & Verhaegen, 2009, Table 1), especially in the LPV case. Furthermore, direct estimation of the matrices will not take the structural restrictions of $\check{\mathcal{L}}_f$ into account, which leads to a non-causal model estimate as pointed out in Shi and Macgregor (2010). Therefore, we follow an alternative route by estimating a predictor in the pre-estimation step and construct $\mathcal{O}_f\check{\mathcal{R}}_p$, $\check{\mathcal{L}}_f$, $\tilde{\mathcal{O}}_f\tilde{\mathcal{R}}_p$ to lower the computational demand and to enforce a causal model, similar to recent literature (Chiuso, 2007; Jansson, 2005; Qin, Lin, & Ljung, 2005; Verhaegen & Verdult, 2007).

4. Subspace identification in open-loop form

In this section, we derive two methods to realize the state-sequence based on the open-loop data-equation (8). The first method is based on a maximum-likelihood argument using *canonical correlation analysis* (CCA) (Section 4.2) and the second method applies a realization based argument (Section 4.3). The latter deterministic state realization approach results in the LPV extension of various LTI schemes by using different weighting matrices in the state realization step.

4.1. Main concept

The stochastic and the deterministic approaches use the fact that the observability and reachability matrices can be decomposed into a parameter independent and a parameter dependent part. To this end, define

$$\begin{aligned}\check{P}^u_{t|p}&=\psi_t\otimes\cdots\otimes\psi_{t-p}\otimes\mathrm{I}_{n_\mathrm{u}},\\ \check{P}^\xi_{t|p}&=\psi_t\otimes\cdots\otimes\psi_{t-p}\otimes\mathrm{I}_{n_\mathrm{y}},\\ \check{M}_{t,p}&=\mathrm{diag}\big(\check{P}^u_{t-1|0},\check{P}^\xi_{t-1|0},\ldots,\check{P}^u_{t-1|p-1},\check{P}^\xi_{t-1|p-1}\big),\\ L_{t|f}&=\psi_t^\top\otimes\cdots\otimes\psi_{t+f}^\top\otimes\mathrm{I}_{n_\mathrm{y}},\\ N_{t,f}&=\mathrm{diag}\left(L_{t|0},\ldots,L_{t|f-1}\right).\end{aligned}$$

The p-step extended reachability matrix and the f-step extended observability matrix are given as

$$R_p=\begin{bmatrix}\mathtt{R}_1 & \cdots & \mathtt{R}_j\end{bmatrix},\qquad O_f=\begin{bmatrix}\mathtt{O}_1^\top & \cdots & \mathtt{O}_i^\top\end{bmatrix}^\top, \tag{18}$$

with dimensions $R_p\in\mathbb{R}^{n_\mathrm{x}\times\left(n_\mathrm{u}\sum_{l=1}^{p}(1+n_\psi)^l\right)}$ and $O_f\in\mathbb{R}^{\left(n_\mathrm{y}\sum_{l=1}^{f}(1+n_\psi)^l\right)\times n_\mathrm{x}}$ where $\mathtt{R}_k$, $\mathtt{O}_k$ are defined as

$$\mathtt{R}_1=\begin{bmatrix}B_0 & \cdots & B_{n_\psi} & K_0 & \cdots & K_{n_\psi}\end{bmatrix},$$

$$\mathtt{R}_k=\begin{bmatrix}A_0\mathtt{R}_{k-1} & \cdots & A_{n_\psi}\mathtt{R}_{k-1}\end{bmatrix}, \tag{19a}$$

$$\mathtt{O}_1=\begin{bmatrix}C_0^\top & \cdots & C_{n_\psi}^\top\end{bmatrix}^\top,$$

$$\mathtt{O}_k=\begin{bmatrix}(\mathtt{O}_{k-1}A_0)^\top & \cdots & (\mathtt{O}_{k-1}A_{n_\psi})^\top\end{bmatrix}^\top. \tag{19b}$$

Using Assumption A.8, the open-loop data-equation (8) can be decomposed as

$$\mathrm{y}_t^{t+f}=\underbrace{N_{t,f}O_f}_{(\mathcal{O}_f\diamond p)_t}\underbrace{\check{R}_p\check{M}_{t,p}}_{(\check{\mathcal{R}}_p\diamond p)_t}\check{\mathrm{z}}_t^{t-p}+(\check{\mathcal{L}}_f\diamond p)_t\check{\mathrm{z}}_t^{t+f}+\xi_t^{t+f}, \tag{20}$$

Data-equation (20) describes the IO relations of the data-generating system based on an SS form. The unknowns in this IO relation are the so-called *sub-Markov* parameters $C_iA_j\cdots A_kB_l$ and $C_iA_j\cdots A_kK_l$. Using (20), the sub-Markov parameters and the unknown noise sequence ξ_t can be estimated by LPV-MAX model estimation using convex optimization (Cox, 2018, Thm. 5.5).

In this section, the state realization is accomplished by assuming that a sub-part of the structural observability matrix O_f associated with the parameter independent part of the SS representation, i.e., C_0 and A_0, is full column rank (common assumption applied in practice (Luspay et al., 2009; Schulz et al., 2016; van Wingerden et al., 2009; Verdult et al., 2002)).[4] To this end, define

[4] Any $C_iA_iK_i$ combination could be taken instead of $C_0A_0K_0$. In such case, additional assumptions should be taken on the associated scheduling variable to fulfill the observability criterion, which is not treated in this paper to simplify the discussion.

the scheduling independent observability matrices

$$O_f^0 = \begin{bmatrix} C_0 \\ C_0A_0 \\ \vdots \\ C_0A_0^{f-1} \end{bmatrix}, \quad \tilde{O}_f^0 = \begin{bmatrix} C_0 \\ C_0(A_0 - K_0C_0) \\ \vdots \\ C_0(A_0 - K_0C_0)^{f-1} \end{bmatrix},$$

for the open-loop and closed-loop setting, respectively.[5]

A.9 The scheduling independent part of the observability matrix is of rank n_x, i.e., $\operatorname{rank}(O_f^0) = n_x$ or $\operatorname{rank}(\tilde{O}_f^0) = n_x$ for the open-loop or closed-loop case, respectively.

To make use of this assumption, the observability matrix O_f in (20) is split into a part that depends on O_f^0 and a part that does not:

$$y_t^{t+f} - (\check{\mathcal{L}}_f \diamond p)_t \check{z}_t^{t+f} - N_{t,f}^* O_f^0 \check{R}_p \check{M}_{t,p} \check{z}_t^{t-p} O_f^0 \check{R}_p \check{M}_{t,p} \check{z}_t^{t-p} + \xi_t^{t+f}, \quad (21)$$

where

$$\begin{aligned} L_{t|i}^* &= \psi_t^\top \otimes \cdots \otimes \psi_{t+i-1}^\top \otimes \begin{bmatrix} \psi_{t+i}^{[1]} & \cdots & \psi_{t+i}^{[n_\psi]} \end{bmatrix} \otimes \mathrm{I}_{n_y}, \\ N_{t,i}^* &= \operatorname{diag}\left(L_{t|0}^*, \ldots, L_{t|i-1}^*\right), \\ \mathcal{O}_1^* &= \begin{bmatrix} C_1^\top & \cdots & C_{n_\psi}^\top \end{bmatrix}^\top, \\ \mathcal{O}_i^* &= \begin{bmatrix} (\mathcal{O}_{i-1}^* A_0)^\top & \cdots & (\mathcal{O}_{i-1}^* A_{n_\psi})^\top \end{bmatrix}^\top, \\ O_f^* &= \begin{bmatrix} (\mathcal{O}_1^*)^\top & \cdots & (\mathcal{O}_f^*)^\top \end{bmatrix}^\top. \end{aligned}$$

Based on (21), introduce the data-equation describing the so-called *open-loop corrected future*:

$$\check{y}_t^{t+f,(c)} = y_t^{t+f} - (\check{\mathcal{L}}_f \diamond p)_t \check{z}_t^{t+f} - N_{t,f}^* O_f^* \check{R}_p \check{M}_{t,p} \check{z}_t^{t-p}. \quad (22)$$

Using an LPV-MAX estimate of (20), the open-loop corrected future $\check{y}_t^{t+f,(c)}$ (22) can be efficiently computed from data. Then, using this surrogate variable, (21) can be reduced to a data-equation excluding the time-variation in the observability matrix:

$$\check{y}_t^{t+f,(c)} = O_f^0 \check{R}_p \check{M}_{t,p} \check{z}_t^{t-p} + \xi_t^{t+f}. \quad (23)$$

Representation (23) forms the starting point for finding an estimate of the state-bases. In the sequel, we formulate projection SID methods based on a maximum-likelihood argument (Section 4.2) and realization based argument (Section 4.3) on the corrected open-loop data-equation (23) to obtain an estimate of the state-sequence.

4.2. Maximum-likelihood estimation

The corrected formulation (23) is the fundamental data-equation to obtain an estimate of the state-sequence. In this section, the state-sequence is estimated using the canonical correlation analysis. The CCA is a well-known method in statistics that finds a (lower dimensional) space that maximizes the correlation between two random variables (Rao, 1979). In our case, this translates to the objective of finding the unknowns O_f^0, $\check{R}_p$, and the subspace of x_t by maximizing the correlation between $\check{y}_t^{t+f,(c)}$ and $\check{M}_{t,p}\check{z}_t^{t-p}$, e.g., see Chiuso (2007), Gičans (2009), Katayama (2005), Larimore (1983) and Larimore (2005) to mention a few. In Chiuso (2007) and Katayama (2005) statistical optimality of CCA in the LTI setting has been shown by formulating the optimal one-step-ahead predictor of the state based on either the past or future data. We will take an alternative viewpoint by formulating an estimate of the state-sequence by maximizing the log-likelihood function associated with the *least-squares* (LS) estimation problem of the unknowns $O_f^0\check{R}_p$ based on the signals $\check{y}_t^{t+f,(c)}$ and $\check{M}_{t,p}\check{z}_t^{t-p}$ of the model (23). Larimore (1983, 2005) claim maximum log-likelihood of the state estimation using CCA, however, the mathematical derivations are scattered within the literature and appear to be incomplete, as pointed out in Gičans (2009). In Theorem 1, we prove the maximum log-likelihood property for the LPV case. For notational simplicity, let us define the following data-matrices

$$\begin{aligned} \check{Z}_{p,N} &= \begin{bmatrix} \check{M}_{1,p}\check{z}_1^{1-p} & \cdots & \check{M}_{N,p}\check{z}_N^{N-p} \end{bmatrix}, \\ \check{Y}_{f,N}^{(c)} &= \begin{bmatrix} \check{y}_1^{1+f,(c)} & \cdots & \check{y}_N^{N+f,(c)} \end{bmatrix}. \end{aligned}$$

Theorem 1 (*CCA Based State Estimation: Open-Loop Case*)**.** *Given an LPV data-generating system (1) and an associated data set $\mathcal{D}_N$ with $\check{Z}_{p,N}\check{Z}_{p,N}^\top \succ 0$. Compute the following* singular value decomposition *(SVD)*

$$\left(\frac{1}{N}\check{Y}_{f,N}^{(c)}\left(\check{Y}_{f,N}^{(c)}\right)^\top\right)^{-\frac{1}{2}} \check{Y}_{f,N}^{(c)}\check{Z}_{p,N}^\top \left(\frac{1}{N}\check{Z}_{p,N}\check{Z}_{p,N}^\top\right)^{-\frac{1}{2}} = U\tilde{S}V^\top, \quad (24)$$

with the matrices $\tilde{U}$ and $\tilde{V}$, given by

$$\tilde{U} = \left(\frac{1}{N}\check{Y}_{f,N}^{(c)}\left(\check{Y}_{f,N}^{(c)}\right)^\top\right)^{-\frac{1}{2}} U, \qquad \tilde{V} = \left(\frac{1}{N}\check{Z}_{p,N}\check{Z}_{p,N}^\top\right)^{-\frac{1}{2}} V.$$

Under Assumptions A.1–A.2 and A.5–A.9,

$$\hat{X}_N = \tilde{V}_{n_x}^\top \check{Z}_{p,N}, \qquad \text{with} \qquad \frac{1}{N}\hat{X}_N\hat{X}_N^\top = I_{n_x}, \quad (25)$$

where $\tilde{V}_{n_x}$ defines the first n_x columns of $\tilde{V}$, is a maximum-likelihood estimate of the state-sequence. The associated log-likelihood function minimized by this estimate is

$$\begin{aligned} -\log L = \frac{f\, n_y N}{2}(\log(2\pi) + 1) - \frac{N}{2}\log\left(\det\left(\tilde{U}\right)^2\right) \\ + \frac{N}{2}\sum_{i=1}^{n_x}\log(1 - \tilde{s}_i^2), \end{aligned} \quad (26)$$

where $\tilde{S} = \operatorname{diag}(\tilde{s}_1, \ldots, \tilde{s}_{n_x})$. □

Proof. See the Appendix. ■

In case of infinite data, i.e., $N \to \infty$, $\tilde{S}$ will contain exactly n_x nonzero singular values which are equal to one (see Cox, 2018, Remark 9.1). In case of finite data, the state order n_x can be selected by a gap in magnitude between the singular values (Van Overschee & De Moor, 1996; Verhaegen & Verdult, 2007). Alternatively, the stochastic interpretation of the CCA in Theorem 1 allows for a data-driven selection of the model order n_x based on the log-likelihood function L (26) using an information criterion such as Akaike's or the Bayesian information criterion (Ljung, 1999). In depth investigation of order selection is beyond the scope of this paper.

Using the estimate of the state-sequence $\hat{x}$, the state–space matrices $\{A_i, B_i, C_i, D_i, K_i\}_{i=0}^{n_\psi}$ are estimated using two linear regression steps, e.g., see Verdult and Verhaegen (2002, Sec. 2.5). The first step is a standard ℓ_2-loss minimization problem based on the output-equation (1b) with the solution:

$$\begin{bmatrix} \hat{C}_0 & \cdots & \hat{C}_{n_\psi} & \hat{D}_0 & \cdots & \hat{D}_{n_\psi} \end{bmatrix} = Y_N \Phi_{N,\mathrm{o}}^\dagger, \quad (27)$$

where $\Phi^\dagger$ defines the right-pseudo inverse of Φ and the regression matrices are

$$\Phi_{N,\mathrm{o}} = \begin{bmatrix} \psi_1 \otimes \hat{x}_1 & \cdots & \psi_N \otimes \hat{x}_N \\ \psi_1 \otimes u_1 & \cdots & \psi_N \otimes u_N \end{bmatrix}, \quad Y_N = [y_1 \quad \cdots \quad y_N].$$

[5] The closed-loop scheduling independent observability matrix is presented here for compactness of the paper.

Using the output-equation (1b), an estimate of the innovation noise is found in the form of the residual error of (27):

$$[\hat{\xi}_1 \quad \cdots \quad \hat{\xi}_N] = Y_N - \begin{bmatrix} \hat{C}_0 & \cdots & \hat{C}_{n_\psi} & \hat{D}_0 & \cdots & \hat{D}_{n_\psi} \end{bmatrix} \Phi_{N,\mathrm{o}}. \tag{28}$$

The remaining state–space matrices are estimated by a second linear-regression step based on the state equation (1a):

$$\begin{bmatrix} \hat{A}_0 & \cdots & \hat{A}_{n_\psi} & \hat{B}_0 & \cdots & \hat{B}_{n_\psi} & \hat{K}_0 & \cdots & \hat{K}_{n_\psi} \end{bmatrix} = \hat{X}_N' \Phi_{N,s}^\dagger, \tag{29}$$

with

$$\Phi_{N,s} = \begin{bmatrix} \psi_1 \otimes \hat{x}_1 & \cdots & \psi_{N-1} \otimes \hat{x}_{N-1} \\ \psi_1 \otimes u_1 & \cdots & \psi_{N-1} \otimes u_{N-1} \\ \psi_1 \otimes \hat{\xi}_1 & \cdots & \psi_{N-1} \otimes \hat{\xi}_{N-1} \end{bmatrix}, \quad \hat{X}_N' = \begin{bmatrix} \hat{x}_2 & \cdots & \hat{x}_N \end{bmatrix}.$$

4.3. Realization based estimation

In Section 4.2, a statistical viewpoint has been taken based on only the input-scheduling-corrected output signals. Alternatively, inspired by the Ho–Kalman scheme in Tóth, Abbas et al. (2012) or the PBSID scheme in van Wingerden et al. (2009), the problem can be tackled from the realization point of view, i.e., the state is realized by decomposing the sub-Markov coefficients in $O_f^0 \check{R}_p$. Opposed to the Ho–Kalman scheme, we are interested in the state-sequence of the innovation form (1) including the noise dynamics. In the LTI case, these ideas have been extensively exploited resulting in many variants of subspace identification schemes, e.g., see Qin (2006), Van Overschee and De Moor (1996) and Verhaegen and Verdult (2007). As recognized in Van Overschee and De Moor (1996, Thm. 12), most of the LTI subspace schemes only differ by left- and right-multiplication of the Hankel matrix with different weightings. Following this concept, a unified LPV formulation of subspace schemes can be introduced:

Theorem 2 (*Unified State Realization: Open-Loop Case*)**.** *Given an LPV data-generating system (1) and an associated data set $\mathcal{D}_N$ with $\check{Z}_{p,N}\check{Z}_{p,N}^\top \succ 0$. Under Assumptions A.1–A.2 and A.5–A.9, let $O_f^0 \check{R}_p$ be the consistent estimate of the sub-Hankel matrix in (23). Compute the following SVD*

$$W_1 O_f^0 \check{R}_p W_2 = USV^\top, \tag{30}$$

where the full rank weightings can be taken as

$$HK \begin{cases} W_1 = I, \\ W_2 = I, \end{cases} \qquad N4SID \begin{cases} W_1 = I, \\ W_2 = \left(\check{Z}_{p,N}\check{Z}_{p,N}^\top\right)^{\frac{1}{2}}, \end{cases}$$

$$p\text{-}CCA \begin{cases} W_1 = \left(\check{Y}_{f,N}^{(c)}\left(\check{Y}_{f,N}^{(c)}\right)^\top\right)^{-\frac{1}{2}}, \\ W_2 = \left(\check{Z}_{p,N}\check{Z}_{p,N}^\top\right)^{\frac{1}{2}}, \end{cases}$$

where the HK stands for Ho–Kalman*, N4SID is* numerical subspace state–space system identification*, and p-CCA is* projected canonical correlation analysis *following the default naming in the subspace literature, see Van Overschee and De Moor (1996). A realization of the state-sequence is given by*

$$\hat{X}_N = S^{\frac{1}{2}} V_{n_x}^\top W_2^{-1} \check{Z}_{p,N}. \quad \square \tag{31}$$

Proof. The LPV estimation problem can be rewritten in an LTI formulation because the IO model (23) representing the SS form of the data-generating system is linear time-invariant w.r.t. the signals $\check{M}_{t,p}\check{z}_t^{t-p}$ and $\check{y}_t^{t+f,(c)}$. Hence, the derivation for LTI subspace schemes can be directly applied. For a rigorous proof, see Van Overschee and De Moor (1996, Chpt. 4.3). To illustrate, it is not difficult to show that

$$O_f^0 = W_1^{-1} U_{n_x} S_{n_x}^{\frac{1}{2}}, \qquad \check{R}_p = S_{n_x}^{\frac{1}{2}} V_{n_x}^\top W_2^{-1}. \tag{32}$$

Hence, taking (6) and Assumption A.8 into account, right-multiplying the reachability matrix $\check{R}_p$ in (32) with the data matrix $\check{Z}_{p,N}$ leads to a realization of the state as in (31). ■

An important fact in Theorem 2 is the absence of the closed-loop dynamics, contrary to the LTI case (Qin, 2006; Van Overschee & De Moor, 1996), and the required pre-estimation step to obtain $O_f^0 \check{R}_p$. Opposed to the LTI case, we do not apply oblique projections to remove the effect of future inputs, e.g., see Van Overschee and De Moor (1996, Sec. 4.2) (the oblique projection is an indirect LS estimation and prediction step). In the LPV case, we apply the pre-estimation step making the oblique projections of the future input superfluous and, therefore, a MOESP like weighting is not present. The unified formulation in Theorem 2 applies N4SID and CCA like weightings to the estimated Hankel matrix, but it is not an LPV extension of these methods, due to the missing oblique projections. In addition, it is important to note that the CCA in Theorem 1 and the p-CCA in Theorem 2 are different as CCA is based on stochastic realization theory and signal relations while p-CCA is based on pre-estimated sub-Markov coefficients. This theoretical split can also be found in the LTI literature, e.g., between Larimore (1990) and Van Overschee and De Moor (1996), respectively. Both principles are equivalent for $N \to \infty$, as the oblique projections and least-squares estimates are consistent and unbiased (Bauer & Ljung, 2002). The choice of the weightings W_1 and W_2 has been discussed by many authors. In the LTI case, it has been proven that W_1 has no influence on the asymptotic accuracy of the estimates, see Bauer and Ljung (2002), Chiuso and Picci (2004) and Gustafsson and Rao (2002). On the other hand, on finite data, the optimal choice is still an open question.

For any applied weighting, the estimated state-sequence in the unified formulation (31) is not guaranteed to have unit variance compared to the estimate by the CCA method in (25). In the LTI case, it can be shown that the resulting model estimate is stochastically balanced for any choice of the weighting (Arun & Kung, 1990), similar to deterministic Ho–Kalman realization. In the LPV case, the observability and reachability Gramians are scheduling-dependent and the authors believe that the state-sequence (31) is structurally balanced, but formally proving this property is a subject of future research.

5. Subspace identification in closed-loop form

The concepts of the presented state estimation and realization schemes for the open-loop identification setting in Theorems 1 and 2 will be extended to the closed-loop case in this section. Similar to the open-loop case, the realization problem is first tackled from an ML point of view (Section 5.2) and then from a deterministic realization viewpoint (Section 5.3). We would like to emphasize that the scheme presented in van Wingerden and Verhaegen (2009) simplifies the realization problem by considering the matrix functions C, D, and K to be constant. No such assumption will be taken next.

5.1. Main concept

Construction of the matrices $\tilde{N}_{t,f}$, $\tilde{O}_f$, $\tilde{R}_p$, and $\tilde{M}_{t,p}$ in the closed-loop case are more involved due to the multi-quadratic parameterization of $\tilde{\mathcal{A}}(p_t)$ and $\tilde{\mathcal{B}}(p_t)$ in (10). First, define all unique combinations of the scheduling induced variation $\psi_t \otimes \psi_t$ as

$$\mu_t = \begin{bmatrix} 1 & \psi_t^\top & \psi_t^{[1]}\psi_t^{[1]} & \cdots & \psi_t^{[1]}\psi_t^{[n_\psi]} & \psi_t^{[2]}\psi_t^{[2]} \\ \psi_t^{[2]}\psi_t^{[3]} & \cdots & \psi_t^{[n_\psi]}\psi_t^{[n_\psi]} \end{bmatrix}^\top, \tag{33}$$

where $\mu : \mathbb{Z} \to \mathbb{M} \subset \mathbb{R}^{n_\mu+1}$ with dimension $n_\mu = \frac{1}{2}n_\psi(n_\psi+3)$ is called the *extended scheduling variable*[6] in the sequel. In addition, define

$$\begin{bmatrix}\mathcal{A}_0 & \cdots & \mathcal{A}_{n_\mu}\end{bmatrix} = \begin{bmatrix} A_0 - K_0C_0 & A_1 - K_1C_0 - K_0C_1 & \cdots \\ A_{n_\psi} - K_{n_\psi}C_0 - K_0C_{n_\psi} & - K_1C_1 & - K_1C_2 - K_2C_1 & \cdots \\ -K_1C_{n_\psi} - K_{n_\psi}C_1 & - K_2C_2 & - K_2C_3 - K_3C_2 & \cdots\end{bmatrix},$$

and

$$\begin{bmatrix}\mathcal{B}_0 & \cdots & \mathcal{B}_{n_\mu}\end{bmatrix} = \begin{bmatrix} B_0 - K_0D_0 & B_1 - K_1D_0 - K_0D_1 & \cdots \\ B_{n_\psi} - K_{n_\psi}D_0 - K_0D_{n_\psi} & - K_1D_1 & - K_1D_2 - K_2D_1 & \cdots \\ -K_1D_{n_\psi} - K_{n_\psi}D_1 & - K_2D_2 & - K_2D_3 - K_3D_2 & \cdots\end{bmatrix},$$

Next, let us define the closed-loop p-step extended reachability matrix:

$$\begin{aligned}\tilde{\mathrm{R}}^u_1 &= \begin{bmatrix}\mathcal{B}_0 & \cdots & \mathcal{B}_{n_\mu}\end{bmatrix}, \qquad \tilde{\mathrm{R}}^y_1 = \begin{bmatrix}K_0 & \cdots & K_{n_\psi}\end{bmatrix}, \\ \tilde{\mathrm{R}}^i_j &= \begin{bmatrix}\mathcal{A}_0 & \cdots & \mathcal{A}_{n_\mu}\end{bmatrix}(\mathrm{I}_{n_\mu} \otimes \tilde{\mathrm{R}}^i_{j-1}), \\ \tilde{R}_p &= \begin{bmatrix}\tilde{\mathrm{R}}^u_1 & \tilde{\mathrm{R}}^y_1 & \cdots & \tilde{\mathrm{R}}^u_p & \tilde{\mathrm{R}}^y_p\end{bmatrix}, \end{aligned} \tag{34a}$$

with $i \in \{u, y\}$ and define the closed-loop f-step extended observability matrix as

$$\begin{aligned}\tilde{\mathrm{O}}_1 &= \begin{bmatrix}C_0^\top & \cdots & C_{n_\psi}^\top\end{bmatrix}^\top, \\ \tilde{\mathrm{O}}_j &= (\mathrm{I}_{n_\mu} \otimes \tilde{\mathrm{O}}_{j-1})\begin{bmatrix}\mathcal{A}_0^\top & \cdots & \mathcal{A}_{n_\mu}^\top\end{bmatrix}^\top, \\ \tilde{O}_f &= \begin{bmatrix}\tilde{\mathrm{O}}_1^\top & \cdots & \tilde{\mathrm{O}}_f^\top\end{bmatrix}^\top. \end{aligned} \tag{34b}$$

Finally, the scheduling dependent data-matrices are given as

$$\begin{aligned}\tilde{P}^u_{t|j} &= \mu_t \otimes \cdots \otimes \mu_{t-j} \otimes \mathrm{I}_{n_\mathrm{u}}, \\ \tilde{P}^y_{t|j} &= \mu_t \otimes \cdots \otimes \mu_{t-j-1} \otimes \psi_{t-j} \otimes \mathrm{I}_{n_\mathrm{y}}, \\ \tilde{M}_{t,j} &= \operatorname{diag}\big(\tilde{P}^u_{t-1|0}, \tilde{P}^y_{t-1|0}, \ldots, \tilde{P}^u_{t-1|j-1}, \tilde{P}^y_{t-1|j-1}\big), \\ \tilde{L}_{t|i} &= \mu_t^\top \otimes \cdots \otimes \mu_{t+i-1}^\top \otimes \psi_{t+i}^\top \otimes \mathrm{I}_{n_\mathrm{y}}, \\ \tilde{N}_{t,i} &= \operatorname{diag}\big(\tilde{L}_{t|0}, \ldots, \tilde{L}_{t|i-1}\big). \end{aligned}$$

By applying the aforementioned matrices and Assumption A.8, Eq. (13) reads as

$$\mathrm{y}_t^{t+f} = \underbrace{\tilde{N}_{t,f}\tilde{O}_f}_{(\tilde{\mathcal{O}}_f \diamond p)_t}\underbrace{\tilde{R}_p\tilde{M}_{t,p}}_{(\tilde{\mathcal{R}}_p \diamond p)_t}\tilde{z}_t^{t-p} + (\tilde{\mathcal{L}}_f \diamond p)_t\tilde{z}_t^{t+f} + \xi_t^{t+f}. \tag{35}$$

It can be proven that the LPV-SS representation (1) with state dimension n_x is *stochastically structurally state-observable* if and only if $\operatorname{rank}(\tilde{\mathrm{O}}_{n_\mathrm{x}}) = n_\mathrm{x}$ (Cox, 2018, Lem. 9.4). Similarly, the SS representation (1) is *stochastically structurally state-reachable* if and only if $\operatorname{rank}(\tilde{\mathrm{R}}_{n_\mathrm{x}}) = n_\mathrm{x}$ (Cox, 2018, Lem. 9.4). In addition, based on the extended Hankel matrix $\tilde{O}_f\tilde{R}_p$, the existence of a stochastic realization of $\mathcal{S}_\mathrm{o}$ with finite model order n_x can be proven (Cox, 2018, Lem. 9.5). Hence, the data-equation (35) allows us to obtain a state–space realization of the data-generating system $\mathcal{S}_\mathrm{o}$ by representation (1) . The unknown coefficients $C_i\mathcal{A}_0\cdots\mathcal{A}_0\mathcal{B}_0$, $C_i\mathcal{A}_{ij}\cdots\mathcal{A}_{ij}\mathcal{B}_{ij}$, $C_i\mathcal{A}_0\cdots\mathcal{A}_0K_i$, $C_i\mathcal{A}_{ij}\cdots\mathcal{A}_{ij}K_i$ for $i,j = 0,\ldots,n_\psi$ found in $\tilde{O}_f\tilde{R}_p$ and $\tilde{\mathcal{L}}_f$ are the sub-Markov coefficients of the multi-quadratic parameterization of the closed-loop formulation (35). These unknown quantities can be estimated by a linear regression of an LPV-ARX model.

The developed concepts of the open-loop setting can be applied to obtain a realization of the model in the closed-loop setting. This concept has successfully been used in the LPV literature, e.g., in Larimore (2005), Verdult and Verhaegen (2002) and van Wingerden and Verhaegen (2009) to mention a few. To this end, the closed-loop counterpart of (21) is

$$\mathrm{y}_t^{t+f} - (\tilde{\mathcal{L}}_f \diamond p)_t\tilde{z}_t^{t+f} - \tilde{N}^*_{t,f}\tilde{O}^*_f\tilde{R}_p\tilde{M}_{t,p}\tilde{z}_t^{t-p} = \tilde{O}^0_f\tilde{R}_p\tilde{M}_{t,p}\tilde{z}_p^{p+t} + \xi_t^{t+f}, \tag{36}$$

where

$$\begin{aligned}\tilde{\mathrm{O}}^*_1 &= \begin{bmatrix}C_1^\top & \cdots & C_{n_\psi}^\top\end{bmatrix}^\top, \\ \tilde{\mathrm{O}}^*_j &= (\mathrm{I}_{n_\mu} \otimes \tilde{\mathrm{O}}^*_{j-1})\begin{bmatrix}\mathcal{A}_0^\top & \cdots & \mathcal{A}_{n_\psi}^\top\end{bmatrix}^\top, \\ \tilde{O}^*_f &= \begin{bmatrix}(\tilde{\mathrm{O}}^*_1)^\top & \cdots & (\tilde{\mathrm{O}}^*_f)^\top\end{bmatrix}^\top, \\ \tilde{L}^*_{t|i} &= \mu_t^\top \otimes \cdots \otimes \mu_{t+i-1}^\top \otimes \begin{bmatrix}\psi_{t+i}^{[1]} & \cdots & \psi_{t+i}^{[n_\psi]}\end{bmatrix} \otimes \mathrm{I}_{n_\mathrm{y}}, \\ \tilde{N}^*_{t,i} &= \operatorname{diag}\big(\tilde{L}^*_{t|0}, \ldots, \tilde{L}^*_{t|i-1}\big). \end{aligned}$$

From (36), the *closed-loop corrected future* can be introduced as

$$\tilde{\mathrm{y}}_t^{t+f,(c)} = \mathrm{y}_t^{t+f} - (\tilde{\mathcal{L}}_f \diamond p)_t\tilde{z}_t^{t+f} - \tilde{N}^*_{t,f}\tilde{O}^*_f\tilde{R}_p\tilde{M}_{t,p}\tilde{z}_t^{t-p}, \tag{37}$$

then using the same principle as in the open-loop case, (36) is reduced to a data-equation where the time-variation in the observability matrix disappears due to the use of the corrected future

$$\tilde{\mathrm{y}}_t^{t+f,(c)} = \tilde{O}^0_f\tilde{R}_p\tilde{M}_{t,p}\tilde{z}_t^{t-p} + \xi_t^{t+f}. \tag{38}$$

The closed-loop form (38) enables to treat the state realization problem equivalent to the open-loop case in Section 4.

5.2. Maximum-likelihood estimation

The corrected formulation (38) is the fundamental data-equation to obtain an estimate of the state-sequence. In this section, the state-sequence is estimated using the ML point of view introduced in Section 4.2. For notational simplicity, let us define the following data-matrices

$$\begin{aligned}\tilde{Z}_{p,N} &= \begin{bmatrix}\tilde{M}_{1,p}\tilde{z}_1^{1-p} & \cdots & \tilde{M}_{N,p}\tilde{z}_N^{N-p}\end{bmatrix}, \\ \tilde{Y}^{(c)}_{f,N} &= \begin{bmatrix}\tilde{\mathrm{y}}_1^{1+f,(c)} & \cdots & \tilde{\mathrm{y}}_N^{N+f,(c)}\end{bmatrix}. \end{aligned}$$

For the closed-loop case, we can reformulate Theorem 1 as:

Theorem 3 (*CCA Based State Estimation: Closed-Loop Case*)**.** *Given an LPV data-generating system (1) and an associated data set $\mathcal{D}_N$ with $\tilde{Z}_{p,N}\tilde{Z}_{p,N}^\top \succ 0$. Compute the SVD*

$$\left(\frac{1}{N}\tilde{Y}^{(c)}_{f,N}\left(\tilde{Y}^{(c)}_{f,N}\right)^\top\right)^{-\frac{1}{2}}\tilde{Y}^{(c)}_{f,N}\tilde{Z}_{p,N}^\top\left(\frac{1}{N}\tilde{Z}_{p,N}\tilde{Z}_{p,N}^\top\right)^{-\frac{1}{2}} = U\tilde{S}V^\top, \tag{39}$$

and the matrices $\tilde{U}$ and $\tilde{V}$ given by

$$\tilde{U} = \left(\frac{1}{N}\tilde{Y}^{(c)}_{f,N}\left(\tilde{Y}^{(c)}_{f,N}\right)^\top\right)^{-\frac{1}{2}}U, \quad \tilde{V} = \left(\frac{1}{N}\tilde{Z}_{p,N}\tilde{Z}_{p,N}^\top\right)^{-\frac{1}{2}}V.$$

Under Assumptions A.3–A.4 and A.5–A.9,

$$\hat{X}_N = \tilde{V}_{n_\mathrm{x}}^\top\tilde{Z}_{p,N}, \qquad \text{with} \qquad \frac{1}{N}\hat{X}_N\hat{X}_N^\top = I_{n_\mathrm{x}}, \tag{40}$$

provides a maximum-likelihood estimate of the state-sequence. The associated log-likelihood function minimized by this estimate is

$$\begin{aligned}-\log L = &\frac{fn_yN}{2}(\log(2\pi)+1) - \frac{N}{2}\log\left(\det(\tilde{U})^2\right) \\ &+\frac{N}{2}\sum_{i=1}^{n_x}\log(1-\tilde{s}_i^2), \end{aligned} \tag{41}$$

where $\tilde{S} = \operatorname{diag}(\tilde{s}_1, \ldots, \tilde{s}_{n_x})$. □

[6] n_μ is given by $\binom{n_\psi+1}{1} + \binom{n_\psi+1}{2} - 1$ where $\binom{n}{k}$ denotes the binomial coefficient.

Proof. Follows the same reasoning as in Theorem 1 with trivial adaptations. The complete proof is found in Cox (2018). ■

Note that, Theorem 3 is the LPV counterpart of the LTI SS-ARX scheme presented in Jansson (2003). Hence, an important contribution of our framework is the extension and the generalization of the CCA to the LPV setting making it possible to directly extend the SS-ARX scheme. In addition, as a contribution, the derived CCA setting allows to prove the maximum-likelihood property and to obtain the log-likelihood function of the estimate, which have not been formally proven in the LTI case, see Jansson (2003).

5.3. Realization based estimation

The state estimation problem has been tackled from the input-scheduling-corrected output statistics point of view in Theorem 3. Alternatively, the state-sequence realization problem can be interpreted as a weighted decomposition of the stochastic, closed-loop Hankel matrix $\tilde{O}_f^0\tilde{R}_p$. More specifically, the concepts introduced for the open-loop case in Section 4.3 can be directly extended to the closed-loop case leading to a unified theory, which immediately extends various LTI subspace methods to the LPV case:

Theorem 4 (*Unified State Realization: Closed-Loop Case*)**.** *Given an LPV data-generating system (1) and an associated data set $\mathcal{D}_N$ with $\tilde{Z}_{p,N}\tilde{Z}_{p,N}^\top \succ 0$. Under Assumptions A.3–A.4 and A.5–A.9, let $\tilde{O}_f^0\tilde{R}_p$ be the consistent estimate of the sub-Hankel matrix in (38). Compute the following SVD*

$$W_1\tilde{O}_f^0\tilde{R}_pW_2 = USV^\top, \tag{42}$$

where the full rank weightings can be taken as

$$HK \begin{cases} W_1 = I, \\ W_2 = I, \end{cases} \qquad PBSID \begin{cases} W_1 = I, \\ W_2 = \left(\tilde{Z}_{p,N}\tilde{Z}_{p,N}^\top\right)^{\frac{1}{2}}, \end{cases}$$

$$p\text{-}SS\text{-}ARX \begin{cases} W_1 = \left(\tilde{Y}_{f,N}^{(c)}\left(\tilde{Y}_{f,N}^{(c)}\right)^\top\right)^{-\frac{1}{2}}, \\ W_2 = \left(\tilde{Z}_{p,N}\tilde{Z}_{p,N}^\top\right)^{\frac{1}{2}}. \end{cases}$$

The HK, predictor based subspace identification *(PBSID) and* projected space state autoregressive exogenous method *(p-SS-ARX) follow the default naming in subspace literature, see Jansson (2003), Van Overschee and De Moor (1996) and van Wingerden and Verhaegen (2009). Then, a realization of the state-sequence is given by*

$$\hat{X}_N = S^{\frac{1}{2}}V_{n_x}^\top W_2^{-1}\tilde{Z}_{p,N}. \quad \square \tag{43}$$

Proof. Based on a similar argument as for Theorem 2. ■

In Chiuso (2007) and van Wingerden and Verhaegen (2009) the implementation and derivation of the PBSID method is accomplished differently. Without exploring the theoretical basis, similar to the above given general theory, the authors in Chiuso (2007) and van Wingerden and Verhaegen (2009) aimed at realizing a computationally efficient estimator by computing the SVD on $W_1\tilde{O}_f^0\tilde{R}_p\tilde{Z}_{p,N} = USV^\top$ and realize the state by $\hat{X}_N = S^{\frac{1}{2}}V_{n_x}^\top$. Obviously, this method is equivalent to the above defined PBSID weighting, but it is computationally more efficient as it avoids the square root operation in (42). Additionally, Chiuso (2007) proves asymptotic equivalence between LTI PBSID and LTI SS-ARX. Extension of this proof to the LPV case has not been accomplished yet, but it is likely to hold. A so-called "optimal" formulation of Theorem 4 can also be derived (Cox, 2018, Sec. 9.8) based on the LTI formulation (Chiuso, 2007). The general idea of Chiuso (2007) is to prove that the initial condition $\mathcal{X}_p$ on the data-equation falls within the variance of the estimator and it can be neglected if the past window p is chosen large enough. This concept translates to taking the assumption that $[\prod_{i=1}^{p}\mathcal{A}(p_{t-i})] \approx 0$ or $[\prod_{i=1}^{p}(\tilde{\mathcal{A}}\diamond p)_{t-i}] \approx 0$ for all $p \in \mathbb{P}^{\mathbb{Z}}$ instead of Assumption A.8.

Remark 5. Theorems 1, 2, 3, and 4 can straightforwardly be modified such that $\mathcal{A}(p_t), \ldots, \mathcal{K}(p_t)$ are affinely dependent on individual basis functions $\{(\alpha^{[i]} \diamond p)_t\}_{i=1}^{n_\alpha}, \ldots, \{(\kappa^{[j]} \diamond p)_t\}_{j=1}^{n_\kappa}$ with dynamic dependency.

Remark 6. To lower the computational load w.r.t. the IO estimation and realization, we can apply the kernelization based computation similar to Verdult and Verhaegen (2005) and van Wingerden and Verhaegen (2009).

6. Simulation example

In this section, we will demonstrate the performance of the discussed LPV subspace identification schemes on the benchmark example given in Verdult and Verhaegen (2005). The developed subspace schemes are compared to the $\text{PBSID}_{\text{opt}}$ method (van Wingerden & Verhaegen, 2009).

6.1. Identification setting

The benchmark is based on a MIMO LPV-SS model with input dimension $n_\text{u} = 2$, scheduling dimension $n_\text{p} = 2$, state dimension $n_\text{x} = 2$, and output dimension $n_\text{y} = 2$. To be able to compare the developed approaches to existing LPV subspace methods, we consider the simplified setting with a scheduling independent innovation noise model

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

The innovation noise model is chosen such that the open-loop and closed-loop dynamics are asymptotically input-to-state stable on the domain $p_t \in \mathbb{P} = [-1, 1]^2$ with a quadratic Lyapunov function defined by a constant symmetric matrix. The noise process $\boldsymbol{\xi}$ is taken as a white noise with distribution $\boldsymbol{\xi}_t \sim \mathcal{N}(0, \Xi)$ where Ξ is diagonal and it is chosen such that the *signal-to-noise ratio* (SNR)[7]

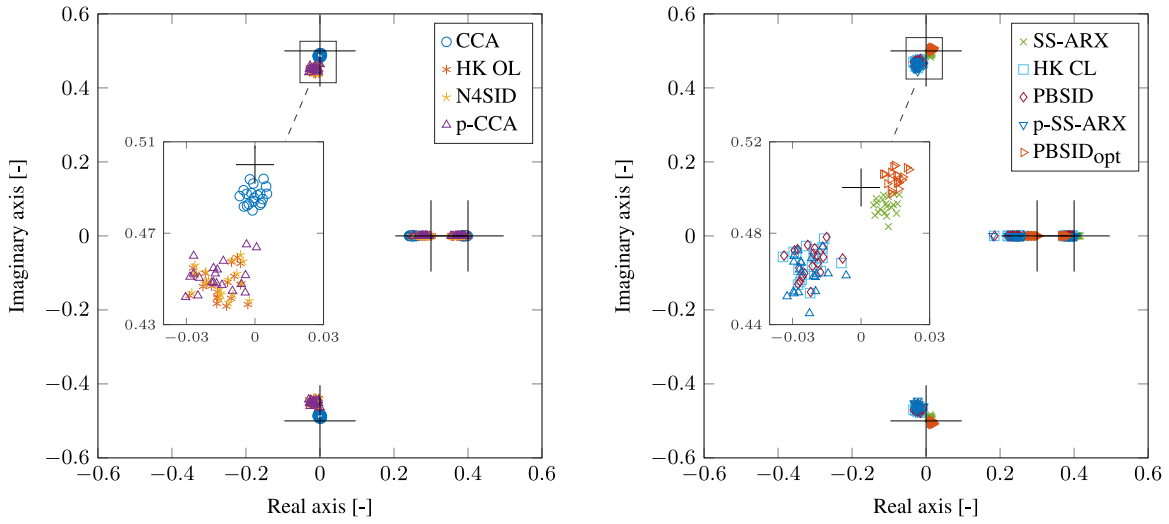
$$\text{SNR}_y^{[i]} = 10\log_{10}\frac{\sum_{t=1}^{N}(y_t^{[i]})^2}{\sum_{t=1}^{N}(w_t^{[i]})^2},$$

is set for various Monte Carlo experiments as $\text{SNR}_y^{[i]} = \{\infty, 25, 10, 0\}$ dB for all $i = 1, \ldots, n_\text{y}$. The $[i]$ denotes the ith channel, i.e., element of the vector signal, and $\text{SNR}_y^{[i]}$ is the SNR of the output $y^{[i]}$. To evaluate the statistical properties of the subspace schemes, we will carry out two simulation studies with $N = \{10^3, 10^4\}$ samples in the identification data set $\mathcal{D}_N$ and in each simulation study $N_\text{MC} = 100$ Monte Carlo runs are executed. In each run, new realizations of the input and scheduling signals are used. The simulation output or one-step-ahead predicted output $\hat{y}$ of the estimated model is compared to the measured output and the one-step-ahead predicted output y of the true system (oracle), respectively, by means of the *best fit rate* (BFR)[8]

$$\text{BFR} = \max\left\{1 - \frac{\frac{1}{N}\sum_{t=1}^{N}\|y_t - \hat{y}_t\|_2}{\frac{1}{N}\sum_{t=1}^{N}\|y_t - \bar{y}\|_2}, 0\right\}\cdot 100\%, \tag{44}$$

[7] The noise w_t is a colored noise signal with state-equation $x_{t+1}^w = \mathcal{A}(p_t)x_t^w + \mathcal{K}(p_t)\xi_t$ and output-equation $w_t = \mathcal{C}(p_t)x_t^w + \xi_t$.

[8] Usually the BFR are defined per channel. Eq. (44) is the average performance criteria over all channels.



(a) Open-loop methodologies.

(b) Closed-loop methodologies.

Fig. 1. The eigenvalues of the estimated A_0 and A_1 matrices for 20 experiments with $\text{SNR}_y^{[i]} = 10$ dB and $N = 10^4$ samples in the data set $\mathcal{D}_N$. The figure displays the subspace methodologies using canonical correlation analysis (CCA), Ho–Kalman like projection in open-loop formulation (HK OL), Ho–Kalman like projection in closed-loop formulation (HK CL), N4SID projection, CCA like projection (p-CCA), canonical correlation analysis in closed-loop (SS-ARX) predictor based subspace (PBSID), SS-ARX like projection (p-SS-ARX) and $\text{PBSID}_{\text{opt}}$ of van Wingerden and Verhaegen (2009).

using a validation data set $\mathcal{D}_{\text{val}}$ as in Verdult and Verhaegen (2005). In (44), $\bar{y}$ defines the mean of the simulation output or the one-step-ahead predicted output y of the oracle.

For the open-loop setting, the MAX model is estimated using *pseudo linear regression* (PLR) where the update is determined by the enhanced Gauss–Newton method (Cox, 2018, Appx. B.3). Note that this optimization problem is convex. The PLR is initialized with a FIR model estimate using ℓ_2-regularized least squares with *generalized cross validation* (GCV) to estimate the optimal regularization parameter (Golub & Van Loan, 2013, Sec. 6.1.4). For the closed-loop setting, an ARX model is estimated using ℓ_2-regularized least squares with GCV. The $\text{PBSID}_{\text{opt}}$ (van Wingerden & Verhaegen, 2009) uses Tikhonov regularization with GCV.

Next, we will provide a summary of the used design parameters, which are optimized to provide the highest BFR on $\mathcal{D}_{\text{val}}$. The MAX model orders are $n_{\text{b}} = 4$ and $n_{\text{c}} = 4$ for $N = 10^3$ and $n_{\text{b}} = 6$ and $n_{\text{c}} = 6$ for $N = 10^4$. The ARX model orders are $n_{\text{a}} = 6$ and $n_{\text{b}} = 6$ for both $N = \{10^3, 10^4\}$. The future and past window for the open-loop CCA (Theorem 1) are $f = 3$ and $p = 4$ for $N = 10^3$ and $f = 3$ and $p = 5$ for $N = 10^4$. For open-loop unified method (Theorem 2), the future and past windows are $f = 1$ and $p = 4$ for $N = 10^3$ and $f = 3$ and $p = 4$ for $N = 10^4$. The future and past windows for the closed-loop methods (Theorems 3 and 4 and $\text{PBSID}_{\text{opt}}$) are chosen as $f = 3$ and $p = 4$ for both $N = \{10^3, 10^4\}$. For the enhanced Gauss–Newton method of the MAX estimation, we enforce a minimal ascension step size of $\beta_1 = 10^{-4}$ per iteration (Armijo–Goldstein condition), the initial value for singular value truncation for the search direction is $\gamma = 10^{-8}$, the minimum of the regularization parameter is $\lambda_{\min} = 10^{-5}$ (Levenberg–Marquardt regularization), the minimum orthogonality requirement of the search direction is $\nu = 0.01$, the minimum step length for backtracking is $\alpha_{\min} = 0.001$, and the first-order termination condition is $\epsilon = 10^{-6}$ with a maximum of 20 iterations. See Cox (2018, Algo. 7.1 and Appx. B.3) for details.

6.2. Analysis of the results

Table 1 shows the mean and the standard deviation (between parentheses) of the BFR on $\mathcal{D}_{\text{val}}$ of the estimation algorithms per Monte Carlo run for various $\text{SNR}_y^{[i]} = \{\infty, 25, 10, 0\}$ dB. The SS-ARX like weighting of Theorem 4 (p-SS-ARX) experiences numerical problems for the data set with $N = 10^3$ samples and, therefore, the BFR is substantially lower. The table shows that the state realization methods based on the maximum-likelihood argument (CCA and SS-ARX) outperform the state realization schemes. Most likely, this difference comes from the fact that the CCA argument obtains a minimum variance estimate of the state given the hypothesized noise. In addition, Fig. 1 also shows that the structural estimation bias of the realization based schemes is bigger than the structural bias of the maximum-likelihood schemes. The structural bias is caused by the fact that the initial condition in Assumption A.8 is not yet small enough. The bias can be further reduced by increasing the past window p; however, this will increase the parameter variance and, therefore, decrease the overall BFR on the estimate.

Compared with $\text{PBSID}_{\text{opt}}$ proposed in van Wingerden and Verhaegen (2009), we can see that direct implementation of the CCA and SS-ARX schemes have comparable performance. However theoretically, $\text{PBSID}_{\text{opt}}$ should be close to the BFR performance of the standard implementation of PBSID, but Table 1 and Fig. 1 highlight a clear difference. This is caused by the kernel trick of $\text{PBSID}_{\text{opt}}$ that significantly improves numerical accuracy. In addition, the difference in BFR between some realization techniques is in the order of 10^{-9}. For example, the case of HK OL, N4SID, and p-CCA for a data set with sample size $N = 10^3$. This indicates that the IO estimation step is dominant over the realization step in terms of the BFR for these particular cases. These observations indicate how important it is to develop a numerically efficient implementation of the developed subspace identification schemes to enhance their performance beyond the theoretical developments of this paper. Therefore, extending the kernel implementation to Theorems 1, 2, 3 and 4 is an important objective for future research. Furthermore, while the comparison is provided here with p-independent innovation noise models, the developed subspace schemes in this paper are capable to accomplish state estimation with p-dependent noise scenarios that are beyond the capabilities of the current state-of-the-art.

7. Conclusion

In this paper, we have presented a unified framework to formulate extensions of subspace identification methods for LPV

Table 1
Mean and standard deviation (between parentheses) of the BFR of the estimation algorithms per Monte Carlo run for different $\text{SNR}_y^{[i]} = \{\infty, 25, 10, 0\}$ dB. The table displays the subspace methodologies using canonical correlation analysis (CCA) in open-loop, the Ho–Kalman like projection (HK) in open-loop (OL) or closed-loop (CL), N4SID projection, CCA like projection (p-CCA), canonical correlation analysis in closed-loop (SS-ARX) predictor based subspace (PBSID), SS-ARX like projection (p-SS-ARX) and $\text{PBSID}_{\text{opt}}$ of van Wingerden and Verhaegen (2009).

(a) BFR using the simulated output.

BFR [%]	$N = 10^3$								$N = 10^4$							
	∞ dB		25 dB		10 dB		0 dB		∞ dB		25 dB		10 dB		0 dB	
CCA	98.45	(0.2202)	97.85	(0.2568)	91.10	(0.8816)	78.79	(2.544)	99.43	$(2.749 \cdot 10^{-2})$	99.05	(0.0654)	94.95	(0.3470)	87.36	(1.017)
HK OL	98.33	(0.1940)	97.90	(0.2674)	90.12	(1.163)	72.57	(2.318)	97.50	$(9.764 \cdot 10^{-2})$	97.40	(0.1238)	93.78	(0.5231)	78.32	(1.592)
N4SID	98.33	(0.1940)	97.90	(0.2674)	90.12	(1.163)	72.57	(2.318)	97.51	$(9.694 \cdot 10^{-2})$	97.41	(0.1235)	93.91	(0.5031)	77.32	(1.716)
p-CCA	98.33	(0.1940)	97.90	(0.2674)	90.12	(1.163)	72.57	(2.318)	97.70	$(9.652 \cdot 10^{-2})$	97.50	(0.1324)	93.77	(0.6117)	74.37	(2.298)
SS-ARX	95.88	(0.8259)	96.02	(0.7916)	90.17	(0.957)	82.59	(2.243)	99.93	$(1.636 \cdot 10^{-2})$	99.36	$(6.611 \cdot 10^{-2})$	95.78	(0.3136)	92.71	(0.5857)
HK CL	94.53	(1.0340)	92.39	(1.397)	71.75	(4.340)	60.72	(2.645)	99.85	$(4.567 \cdot 10^{-2})$	96.97	(0.3260)	86.15	(0.9586)	78.74	(1.203)
PBSID	94.57	(1.0205)	92.45	(1.382)	74.55	(3.162)	62.16	(3.255)	99.85	$(4.567 \cdot 10^{-2})$	97.02	(0.3223)	86.27	(0.9683)	78.70	(1.212)
p-SS-ARX	63.02	(11.79)	51.18	(6.672)	46.00	(2.728)	45.38	(2.664)	99.86	$(4.510 \cdot 10^{-2})$	96.58	(0.4401)	85.64	(1.005)	78.50	(1.228)
$\text{PBSID}_{\text{opt}}$	99.93	$(2.533 \cdot 10^{-2})$	98.98	(0.1961)	94.82	(0.8025)	86.47	(1.915)	99.92	$(1.744 \cdot 10^{-2})$	99.58	$(5.473 \cdot 10^{-2})$	97.53	(0.2391)	92.94	(0.5976)

(b) BFR using the predicted output.

BFR [%]	$N = 10^3$								$N = 10^4$							
	∞ dB		25 dB		10 dB		0 dB		∞ dB		25 dB		10 dB		0 dB	
CCA	98.92	(0.1708)	96.21	(0.1470)	85.20	(0.8021)	65.47	(1.843)	99.50	$(2.504 \cdot 10^{-2})$	97.00	$(3.581 \cdot 10^{-2})$	87.00	(0.2044)	67.88	(0.4496)
HK OL	98.90	(0.1557)	96.27	(0.1596)	84.17	(0.8873)	58.84	(1.592)	97.63	(0.1072)	95.65	$(8.794 \cdot 10^{-2})$	84.15	(0.6078)	57.24	(0.8006)
N4SID	98.90	(0.1557)	96.27	(0.1596)	84.17	(0.8873)	58.84	(1.592)	97.63	(0.1068)	95.65	$(8.748 \cdot 10^{-2})$	84.13	(0.6025)	57.37	(0.9458)
p-CCA	98.90	(0.1557)	96.27	(0.1596)	84.17	(0.8873)	58.84	(1.592)	97.72	(0.1085)	95.68	$(9.926 \cdot 10^{-2})$	84.79	(0.6815)	57.60	(1.215)
SS-ARX	97.24	(0.5553)	95.31	(0.5791)	84.30	(1.064)	72.19	(1.657)	99.94	$(1.323 \cdot 10^{-2})$	97.02	(0.0760)	88.63	(0.2212)	77.42	(0.5231)
HK CL	95.86	(0.7027)	93.17	(0.8058)	72.53	(2.883)	51.40	(2.905)	99.88	$(3.682 \cdot 10^{-2})$	96.75	(0.2393)	85.93	(0.6049)	72.27	(0.6701)
PBSID	95.88	(0.7007)	93.20	(0.798)	74.50	(2.093)	53.48	(3.061)	99.88	$(3.682 \cdot 10^{-2})$	96.77	(0.2351)	85.98	(0.6128)	72.25	(0.6748)
p-SS-ARX	68.09	(12.87)	54.42	(8.925)	44.78	(3.105)	29.11	(1.757)	99.88	$(3.627 \cdot 10^{-2})$	95.58	(0.3735)	84.79	(0.7385)	71.56	(0.7321)
$\text{PBSID}_{\text{opt}}$	99.95	$(1.710 \cdot 10^{-2})$	97.23	(0.1720)	87.95	(0.5796)	74.09	(1.624)	99.94	$(1.397 \cdot 10^{-2})$	97.12	(0.0689)	89.13	(0.1646)	76.34	(0.3061)

identification by systematically developing LPV subspace identification theory. Based on the derived open-loop, closed-loop, and predictor-based data-equations, several methods have been proposed to estimate LPV-SS models in one unified framework based on a maximum-likelihood or realization argument. Hence, we have shown how to extend LTI CVA, SS-ARX, PBSID, and N4SID to the LPV setting. The effectiveness of the presented subspace identification methods is demonstrated in a Monte Carlo study by identifying a MIMO LPV benchmark system. An important future direction of research is to improve numerical efficiency and reduce computational load of the developed methods.

Acknowledgments

We would like to thank the authors of van Wingerden and Verhaegen (2009) for providing their code to make the simulation study possible.

Appendix. Proof of Theorem 1

The idea of the proof is based on the reasoning used in Gičans (2009) for the LTI case, however, we will also fix some of the inconsistencies found in Gičans (2009). Representation (23) reads as:

$$\check{y}_t^{t+f,(c)} = \underbrace{O_f^0 \check{R}_p}_{\mathcal{H}_{f,p}^0} \check{M}_{t,p} \check{z}_t^{t-p} + \xi_t^{t+f} = O_f^0 x_t + \xi_t^{t+f}. \tag{A.1}$$

Note that ξ_t^{t+f} is a sample path realization of a Gaussian white noise with variance $\Sigma_\xi^2 = \mathrm{I}_f \otimes \Xi^2$. The model (A.1) is dependent on the unknown sub-Markov parameters in $\mathcal{H}_{f,p}^0$, which are parameterized as $\mathcal{H}$ and the unknown noise variance Σ_ξ^2 with symmetric parameterization Σ^2. Due to the Markov property of representation (A.1) and by employing Bayes' rule, the maximum likelihood estimate can be obtained by maximizing the likelihood or, equivalently, by minimizing the log-likelihood based on the data set $\mathcal{D}_N$

$$\min_{\substack{\mathcal{H}\in\Theta \\ \Sigma^2\in\mathbb{S}}} -\log L(\mathcal{H}, \Sigma^2) = \min_{\substack{\mathcal{H}\in\Theta \\ \Sigma^2\in\mathbb{S}}} \frac{1}{2} f n_y N \log(2\pi) + \frac{1}{2} N \log\left(\det\left(\Sigma^2\right)\right) + \frac{1}{2} \sum_{t=1}^{N} \varepsilon_{t|\mathcal{H}}^\top \Sigma^{-2} \varepsilon_{t|\mathcal{H}}, \tag{A.2}$$

where the one-step-ahead prediction-error is $\varepsilon_{t|\mathcal{H}} = \check{y}_t^{t+f,(c)} - \mathcal{H}\check{M}_{t,p}\check{z}_t^{t-p}$ and the future window is f. As the signals are assumed to be persistently exciting, i.e., $\check{Z}_{p,N}(\check{Z}_{p,N})^\top \succ 0$, the well-known unique stationary point of (A.2) is obtained at (Gibson, Wills, & Ninness, 2005, Lem. 3.3)[9]

$$\hat{\mathcal{H}}_{f,p}^0 = \frac{1}{N} \check{Y}_{f,N}^{(c)} \check{Z}_{p,N}^\top (\check{Z}_{p,N} \check{Z}_{p,N}^\top)^{-1}, \tag{A.3a}$$

$$\hat{\Sigma}_\xi^2 = \frac{1}{N} \check{Y}_{f,N}^{(c)} (\check{Y}_{f,N}^{(c)})^\top - \frac{1}{N} \hat{\mathcal{H}}_{f,p}^0 \check{Z}_{p,N} \check{Z}_{p,N}^\top (\hat{\mathcal{H}}_{f,p}^0)^\top. \tag{A.3b}$$

The solution (A.3) is a consistent estimate which is efficient in terms of the parameter variance. The interest is not in an estimate of $\mathcal{H}_{f,p}^0$, but to attain a realization of the state, O_f^0, and $\check{R}_p$ separately that together will maximize the likelihood (A.2). To start, assume that $\check{R}_p \check{Z}_{p,N}$ is known a-priori in (A.1). Then, similar to (A.2), the solution to the linear least-squares problem is

$$\hat{O}_f^0 = \frac{1}{N} \check{Y}_{f,N}^{(c)} \check{Z}_{p,N}^\top \check{R}_p^\top \left(\frac{1}{N} \check{R}_p \check{Z}_{p,N} \check{Z}_{p,N}^\top \check{R}_p^\top \right)^{-1}, \tag{A.4a}$$

[9] The estimate $\hat{\Sigma}_\xi^2 = \frac{1}{N} \varepsilon_{t|\hat{\mathcal{H}}} \varepsilon_{t|\hat{\mathcal{H}}}^\top$ is simplified as

$$\frac{1}{N} \check{Z}_{p,N} (\check{Y}_{f,N}^{(c)} - \hat{\mathcal{H}}_{f,p} \check{Z}_{p,N})^\top = \frac{1}{N} \check{Z}_{p,N} \left(\check{Y}_{f,N}^{(c)} - \frac{1}{N} \check{Y}_{f,N}^{(c)} \check{Z}_{p,N}^\top \left(\frac{1}{N} \check{Z}_{p,N} \check{Z}_{p,N}^\top \right)^{-1} \check{Z}_{p,N} \right)^\top = 0.$$

$$\hat{\Sigma}_\xi^2 = \frac{1}{N}\check{Y}_{f,N}^{(c)}(\check{Y}_{f,N}^{(c)})^\top - \frac{1}{N^2}\check{Y}_{f,N}^{(c)}\check{Z}_{p,N}^\top \check{R}_p^\top \times \left(\frac{1}{N}\check{R}_p\check{Z}_{p,N}\check{Z}_{p,N}^\top\check{R}_p^\top\right)^{-1}\check{R}_p\check{Z}_{p,N}(\check{Y}_{f,N}^{(c)})^\top. \tag{A.4b}$$

The log-likelihood function associated with (A.4), given $\check{R}_p$, $\check{Z}_{p,N}$ and $\check{Y}_{f,N}^{(c)}$, is

$$-\log L(\hat{O}_f^0\check{R}_p, \hat{\Sigma}_\xi^2) = \frac{1}{2} f\, n_y N\,(\log(2\pi)+1) + \frac{N}{2}\log\left(\det\left(\frac{1}{N}\check{Y}_{f,N}^{(c)}(\check{Y}_{f,N}^{(c)})^\top - \frac{1}{N^2}\check{Y}_{f,N}^{(c)}\check{Z}_{p,N}^\top\check{R}_p^\top \times\left(\frac{1}{N}\check{R}_p\check{Z}_{p,N}\check{Z}_{p,N}^\top R_p^\top\right)^{-1}\check{R}_p\check{Z}_{p,N}(\check{Y}_{f,N}^{(c)})^\top\right)\right). \tag{A.5}$$

Note that the last product in (A.2) can be simplified to

$$\frac{1}{2}\sum_{t=1}^{N}\varepsilon_{t|\hat{O}_f^0\check{R}_p}^\top\hat{\Sigma}_\xi^{-2}\varepsilon_{t|\hat{O}_f^0\check{R}_p} = \frac{1}{2}\mathrm{Tr}\left\{N\hat{\Sigma}_\xi^2\hat{\Sigma}_\xi^{-2}\right\} = \frac{1}{2} f\, n_y N.$$

Next, the focus will be on obtaining the p-step extended reachability matrix $\check{R}_p$ that minimizes the log-likelihood (A.5). Similar to Larimore (2005), take the following constrained SVD

$$\begin{aligned}&\tilde{S} = \tilde{U}^\top\frac{1}{N}\check{Y}_{f,N}^{(c)}\check{Z}_{p,N}^\top\tilde{V},\\ \text{s.t.}\quad &\mathrm{I}_{n_y f} = \tilde{U}^\top\frac{1}{N}\check{Y}_{f,N}^{(c)}(\check{Y}_{f,N}^{(c)})^\top\tilde{U},\quad \mathrm{I}_{n_Z} = \tilde{V}^\top\frac{1}{N}\check{Z}_{p,N}\check{Z}_{p,N}^\top\tilde{V},\end{aligned} \tag{A.6}$$

where $\tilde{U}\in\mathbb{R}^{n_y f\times n_y f}$, $\tilde{S}\in\mathbb{R}^{n_y f\times n_Z}$, and $\tilde{V}\in\mathbb{R}^{n_Z\times n_Z}$, and $n_Z = (n_u+n_y)\sum_{i=1}^{p}(1+n_\psi)^i$. This constrained SVD is unique in case $\check{Y}_{f,N}^{(c)}(\check{Y}_{f,N}^{(c)})^\top \succ 0$ and $\check{Z}_{p,N}\check{Z}_{p,N}^\top\succ 0$, which comes naturally under the persistency of excitation conditions. The matrix $\tilde{S}$ is a diagonal matrix with ordered singular values $\tilde{s}_1^2\geq\cdots\geq\tilde{s}_n^2\geq 0$ and $n=\min(n_y f, n_Z)$. In the following discussion, we assume that $n_y f\leq n_Z$ to simplify notation. Note that (A.6) is equivalent to

$$\begin{aligned}&\frac{1}{N}\check{Y}_{f,N}^{(c)}(\check{Y}_{f,N}^{(c)})^\top = (\tilde{U}\tilde{U}^\top)^{-1},\quad \frac{1}{N}\check{Z}_{p,N}\check{Z}_{p,N}^\top = (\tilde{V}\tilde{V}^\top)^{-1},\\ &\frac{1}{N}\check{Y}_{f,N}^{(c)}\check{Z}_{p,N}^\top = (\tilde{U}\tilde{U}^\top)^{-1}\tilde{U}\tilde{S}\tilde{V}^\top(\tilde{V}\tilde{V}^\top)^{-1} = \tilde{U}^\dagger\tilde{S}(\tilde{V}^\dagger)^\top,\end{aligned} \tag{A.7}$$

by applying the left pseudo-inverses. Substituting the relations (A.7) into the second line of (A.5) leads to

$$\frac{N}{2}\log\left(\det(\tilde{U}^\dagger)\det\left(\mathrm{I}_{n_y f} - \tilde{S}(\tilde{V}^\dagger)^\top\check{R}_p^\top \times\left(\check{R}_p(\tilde{V}\tilde{V}^\top)^{-1}\check{R}_p^\top\right)^{-1}\check{R}_p\tilde{V}^\dagger\tilde{S}^\top\right)\det\left((\tilde{U}^\dagger)^\top\right)\right). \tag{A.8}$$

In (A.8), the product property of the determinant is applied: $\det(AB) = \det(A)\cdot\det(B)$. It is important to see that the constrained SVD (A.6) only decomposes the signal relations based on the (co)variances and the decomposition does not change the minimization of (A.5). To simplify the notation, define

$$Q \triangleq (\tilde{V}^\dagger)^\top\check{R}_p^\top\in\mathbb{R}^{n_Z\times n_x}, \tag{A.9}$$

which represents an injective mapping of $\check{R}_p$ to Q as $\tilde{V}$ is full rank. Applying this transformation, (A.8) read as

$$-\frac{1}{2}N\log\left(\det\left((\tilde{U}\tilde{U}^\top)^{-1}\right)\times\det\left(\mathrm{I}_{n_y f} - \tilde{S}Q(Q^\top Q)^{-1}Q^\top\tilde{S}^\top\right)\right). \tag{A.10}$$

Note that minimization of (A.5) is equivalent to minimizing (A.10) with a change of variables. Furthermore, the inverted expression $Q^\top Q$ in (A.10) can be written as

$$Q^\top Q = \check{R}_p(\tilde{V}\tilde{V}^\top)^{-1}\check{R}_p^\top = \frac{1}{N}\check{R}_p\check{Z}_{p,N}\check{Z}_{p,N}^\top\check{R}_p^\top = \frac{1}{N}X_N X_N^\top, \tag{A.11}$$

which is the sample variance of the to-be-chosen state variable. The realized state is an isomorphic representation with respect to the original SS form of the underlying data-generating system, see Section 3. Hence, its sample variance $\frac{1}{N}X_N X_N^\top$ can be chosen to be any arbitrary positive definite matrix. In CCA, the variance is chosen to be identity, hence, all states have equal magnitude and are maximally uncorrelated. As such, minimizing the log-likelihood (A.10) renders

$$\begin{aligned}&\min_{Q^\top Q=\mathrm{I}_{n_x}}\log\left(\det\left(\mathrm{I}_{n_y f} - \tilde{S}QQ^\top\tilde{S}^\top\right)\right)\\ &= \min_{Q^\top Q=\mathrm{I}_{n_x}}\log\left(\det\left(\mathrm{I}_{n_y f} - Q^\top\tilde{S}^\top\tilde{S}Q\right)\right),\end{aligned} \tag{A.12}$$

by applying Sylvester's determinant identity. To find an expression for minimizing the log-likelihood, note that the determinant of a matrix is the product of its eigenvalues. From Poincaré separation theorem (Lemma 7) it follows that the choice $Q^\top Q = \mathrm{I}_{n_x}$ will not lead to a single solution of (A.12). Using Lemma 7, let us investigate the individual descending sorted eigenvalues

$$\begin{aligned}\lambda_{n_y f-i+1}\left\{\mathrm{I}_{n_x} - Q^\top\tilde{S}^\top\tilde{S}Q\right\} &= 1-\lambda_i\left\{Q^\top\tilde{S}^\top\tilde{S}Q\right\}\\ &\geq 1-\lambda_i\left\{\tilde{S}^\top\tilde{S}\right\} = 1-\tilde{s}_i^2,\end{aligned} \tag{A.13}$$

for $i = 1,\ldots,n_x$. Hence, minimization of the marginal likelihood (A.12) has a lower-bound based on the product of the singular values in $\tilde{S}$. The lower-bound is clearly obtained if $Q = [\mathrm{I}_{n_x}\quad 0]^\top$, which also satisfies $Q^\top Q = \mathrm{I}_{n_x}$. There might be other solutions to the minimization problem of (A.12), however, we take the solution equal to the CVA solution of Larimore (1983). Hence, the latter choice of Q maximizes the marginal likelihood function (A.5). An estimate of the reachability matrix $\hat{R}_p$ is obtained by reformulating (A.9) as

$$\hat{R}_p = Q^\top\tilde{V}^\top = [\mathrm{I}_{n_x}\quad 0]\tilde{V}^\top, \tag{A.14}$$

which results in selecting the first n_x columns of $\tilde{V}$. Then, the estimates of the observability (A.4a) and noise covariance (A.4b) are equivalent to

$$\hat{O}_f = \tilde{U}^\dagger\tilde{S}(\tilde{V}^\dagger)^\top\hat{R}_p^\top = \tilde{U}^\dagger\tilde{S}Q, \tag{A.15a}$$

$$\begin{aligned}\hat{\Sigma}_\xi^2 &= (\tilde{U}\tilde{U}^\top)^{-1} - \tilde{U}^\dagger\tilde{S}(\tilde{V}^\dagger)^\top\hat{R}_p^\top\hat{R}_p\tilde{V}^\dagger\tilde{S}^\top(\tilde{U}^\dagger)^\top\\ &= \tilde{U}^\dagger(\mathrm{I}_{n_y f} - \tilde{S}QQ^\top\tilde{S}^\top)(\tilde{U}^\dagger)^\top,\end{aligned} \tag{A.15b}$$

where $(\tilde{V}^\dagger)^\top\tilde{V} = \frac{1}{N}\check{R}_p\check{Z}_{p,N}\check{Z}_{p,N}^\top\check{R}_p^\top = \mathrm{I}_{n_x}$ due to (A.6) and (A.11). Note that the multiplication $\tilde{S}Q$ selects only the first n_x singular values of $\tilde{S}$. Combining the estimates (A.14) and (A.15a) results in an estimate of $\mathcal{H}_{f,p}^0$:

$$\hat{\mathcal{H}}_{f,p}^0 = \tilde{U}^\dagger\tilde{S}QQ^\top\tilde{V}^\top. \tag{A.16}$$

The log-likelihood function corresponding to the estimates (A.14)–(A.16) is

$$\begin{aligned}-\log L(\hat{\mathcal{H}}_{f,p}^0, \hat{\Sigma}_\xi^2) &= \frac{f\, n_y N}{2}(\log(2\pi)+1)\\ &+\frac{N}{2}\log\left(\det\left(\frac{1}{N}\check{Y}_{f,N}^{(c)}(\check{Y}_{f,N}^{(c)})^\top\right)\prod_{i=1}^{n_x}(1-\tilde{s}_i^2)\right) =\\ &-\frac{N}{2}\log\left(\det(\tilde{U})^2\right) + \frac{f\, n_y N}{2}(\log(2\pi)+1) + \frac{N}{2}\sum_{i=1}^{n_x}\log(1-\tilde{s}_i^2).\end{aligned}$$

The last remaining step is to show that the estimates of the Hankel matrix $\hat{\mathcal{H}}_{f,p}^0$ (A.16) and the noise $\hat{\Sigma}_\xi^2$ (A.15b) are equivalent to the estimates in (A.3) that characterize the minimum of (A.2). Substitute (A.7) in (A.3)

$$\hat{\mathcal{H}}_{f,p}^0 = \frac{1}{N}\check{Y}_{f,N}^{(c)}\check{Z}_{p,N}^\top\left(\frac{1}{N}\check{Z}_{p,N}\check{Z}_{p,N}^\top\right)^{-1} = \tilde{U}^\dagger\tilde{S}\tilde{V}^\top, \tag{A.17a}$$

$$\hat{\Sigma}_{\xi}^{2} = \frac{1}{N}\check{Y}_{f,N}^{(c)}(\check{Y}_{f,N}^{(c)})^{\top} - \frac{1}{N}\hat{\mathcal{H}}_{f,p}^{0}\check{Z}_{p,N}\check{Z}_{p,N}^{\top}\hat{\mathcal{H}}_{f,p}^{\top}$$
$$= \tilde{U}^{\dagger}\left(\mathrm{I}_{n_y f} - \tilde{S}\tilde{S}^{\top}\right)(\tilde{U}^{\dagger})^{\top}. \tag{A.17b}$$

The estimates (A.15b) and (A.16) are identical to (A.17) when the all singular values are selected, due to $\tilde{S}Q$. In case the number of data points goes to infinity, i.e., $N \to \infty$, then $\tilde{S}$ will contain exactly n_x nonzero singular values. In the finite data case, the number of states in the realization is selected based on Q. In conclusion, the SVD (26) maximizes the marginal-likelihood function of the linear estimation problem (23) w.r.t. the unknowns O_f^0, $\check{K}_p$ and the covariance Ξ^2 with state-sequence (25) and log-likelihood function (26).

Note that in early literature on CVA SID (Larimore, 1983, 1990), the constrained SVD (A.6) was performed with arbitrary positive-definite weight $\Lambda \in \mathbb{S}$ such that $\mathrm{I} = \tilde{V}^{\top}\Lambda\tilde{V}$, which is called the CVA method. The CCA and CVA methods coincide with the weighting choice in (A.6). The CVA method with $\Lambda \neq \frac{1}{N}\check{Z}_{p,N}(\check{Z}_{p,N})^{\top}$ and $\Lambda \succ 0$ leads to a minimal prediction-error solution of $\varepsilon_{t|\hat{\mathcal{H}}}$ (Larimore, 1990, Eq. (10)), but will not lead to a maximum-likelihood estimate of $\mathcal{H}$. ■

Lemma 7 (*Poincaré Separation Theorem*). *Let $A \in \mathbb{S}^n$ and $B \in \mathbb{R}^{n\times r}$ be matrices such that $B^{\top}B = I_r$. Let $\lambda_i\{.\}$ represent the eigenvalues of a matrix sorted in descending order. Then,*

$$\lambda_i\left\{B^{\top}AB\right\} \leq \lambda_i\{A\}, \qquad i = 1, \ldots, r. \quad \square$$

Proof. See Rao (1973, p. 64). ■

References

Arun, K. S., & Kung, S. Y. (1990). Balanced approximation of stochastic systems. *SIAM Journal on Matrix Analysis and Applications*, *11*, 42–68.

Bauer, D., & Ljung, L. (2002). Some facts about the choice of the weighting matrices in Larimore type of subspace algorithms. *Automatica*, *38*, 763–773.

Chiuso, A. (2007). On the relation between CCA and predictor-based subspace identification. *IEEE Transactions on Automatic Control*, *52*(10), 1795–1812.

Chiuso, A., & Picci, G. (2004). The asymptotic variance of subspace estimates. *Journal of Econometrics*, *118*(1–2), 257–291.

Chiuso, A., & Picci, G. (2005). Consistency analysis of some closed-loop subspace identification methods. *Automatica*, *41*, 377–391.

Cox, P. B. (2018). *Towards efficient identification of linear parameter-varying state-space models* (Phd thesis), Eindhoven University of Technology.

Cox, P. B., & Tóth, R. (2016). Alternative form of predictor based identification of LPV-SS models with innovation noise. In *Proc. of the 55th IEEE Conf. on decision and control* (pp. 1223–1228). Las Vegas, NV, USA.

Eykhoff, P. (1974). *System identification parameter and state estimation*. John Wiley & Sons Ltd..

Felici, F., van Wingerden, J.-W., & Verhaegen, M. (2007). Subspace identification of MIMO LPV systems using a periodic scheduling sequence. *Automatica*, *43*(10), 1684–1697.

Gibson, S., Wills, A., & Ninness, B. (2005). Maximum-likelihood parameter estimation of bilinear systems. *IEEE Transactions on Automatic Control*, *50*(10), 1581–1596.

Gičans, O. (2009). Herleitung der linearen canonical variate analysis (CVA) auf basis der multivariaten statistischen analyse. Studienarbeit 219, Institut für Regelungs- und Steuerungssysteme, Universität Karlsruhe.

Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations* (4th ed.). The Johns Hopkins University Press.

Gustafsson, T., & Rao, B. D. (2002). Statistical analysis of subspace-based estimation of reduced-rank linear regression. *IEEE Transactions on Signal Processing*, *50*, 151–159.

Jansson, M. (2003). Subspace identification and ARX modeling. In *Proc. of the 13th IFAC symposium on system identification* (pp. 1585–1590). Rotterdam, The Netherlands.

Jansson, M. (2005). A new subspace identification method for open and closed loop data. In *Proc. of the 16th IFAC world congress* (pp. 500–505). Prague, Czech Republic.

Katayama, T. (2005). *Subspace methods for system identification*. Springer-Verlag London.

Larimore, W. E. (1983). System identification, reduced-order filtering and modeling via canonical variate analysis. In *Proc. of the American control conf.* (pp. 445–451). San Francisco, CA, USA.

Larimore, W. E. (1990). Canonical variate analysis in identification, filtering, and adaptive control. In *Proc. of the 29th Conf. on decision and control* (pp. 596–604). Honolulu, HI, USA.

Larimore, W. E. (2005). Maximum likelihood subspace identification for linear, nonlinear, and closed-loop systems. In *Proc. of the American control conference* (pp. 2305–2319). Portland, OR, USA.

Larimore, W. E., Cox, P. B., & Tóth, R. (2015). CVA identification of nonlinear systems with LPV state-space models of affine dependence. In *Proc. of the American control conf.* (pp. 831–837). Chicago, IL, USA.

Lee, L. H., & Poolla, K. P. (1997). Identifiability issues for parameter-varying and multidimensional linear systems. In *Proc. of the ASME design, engineering technical conf.*, Sacramento, CA, USA.

Lindquist, A., & Picci, G. (2015). *Linear stochastic systems: A geometric approach to modeling, estimation and identification*. Springer.

Ljung, L. (1999). *System identification: Theory for the user* (2nd ed.). Springer.

Lopes Dos Santos, P., Azevedo-Perdicoulis, T.-P., & Ramos, J. A. (2010). Gas pipelines LPV modelling and identification for leakage detection. In *Proc. of the American control conf.* (pp. 1211–1216). Baltimore, Maryland, USA.

Lopes dos Santos, P., Ramos, J. A., & Martins de Carvalho, J. L. (2008). Subspace identification of linear parameter-varying systems with innovation-type noise models driven by general inputs and a measurable white noise time-varying parameter vector. *International Journal of Systems Science*, *39*(9), 897–911.

Luspay, T., Kulcsár, B., Van Wingerden, J.-W., & Verhaegen, M. (2009). On the identification of LPV traffic flow model. In *Proc. of the European control conf.* (pp. 1752–1757). Budapest, Hungary.

Mercère, G., Markovsky, I., & Ramos, J. A. (2016). Innovation-based subspace identification in open- and closed-loop. In *Proc. of the 55th IEEE Conf. on decision and control* (pp. 2951–2956). Las Vegas, NV, USA.

Qin, S. J. (2006). An overview of subspace identification. *Computers and Chemical Engineering*, *30*, 1502–1513.

Qin, S. J., Lin, W., & Ljung, L. (2005). A novel subspace identification approach with enforced causal models. *Automatica*, *41*, 2043–2053.

Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). Wiley, New York.

Rao, C. R. (1979). Separation theorems for singular values of matrices and their applications in multivariate analysis. *Journal of Multivariate Analysis*, *9*(3), 362–377.

Schulz, E., Bussa, A., & Werner, H. (2016). Identification of linear parameterivarying systems via IO and subspace identification - A comparison. In *Proc. of the 55th Conf. on decision and control* (pp. 7147–7152). Las Vegas, NV, USA.

Shi, R., & Macgregor, J. F. (2010). A framework for subspace identification Methods. In *Proc. of the American control conf.* (pp. 3678–3683). Arlington, VA, USA.

Tóth, R., Abbas, H. S., & Werner, H. (2012). On the state-space realization of LPV input-output models: practical approaches. *IEEE Transactions on Control Systems Technology*, *20*(1), 139–153.

Tóth, R., Laurain, V., Gilson, M., & Garnier, H. (2012). Instrumental variable scheme for closed-loop LPV model identification. *Automatica*, *48*(9), 2314–2320.

Van Overschee, P., & De Moor, B. (1996). *Subspace identification for linear systems: Theory - implementation - applications* (p. 272). Springer US.

van Wingerden, J. W., Felici, F., & Verhaegen, M. (2007). Subspace identification of MIMO LPV systems using a piecewise constant scheduling sequence with hard/soft switching. In *Proc. of the European control conf.* (pp. 927–934). Kos, Greece.

van Wingerden, J. W., Houtzager, I., Felici, F., & Verhaegen, M. (2009). Closed-loop identification of the time-varying dynamics of variable-speed wind turbines. *International Journal of Robust and Nonlinear Control*, *19*(1), 4–21.

van Wingerden, J. W., & Verhaegen, M. (2009). Subspace identification of bilinear and LPV systems for open- and closed-loop data. *Automatica*, *45*(2), 372–381.

Verdult, V. (2002). Nonlinear system identification: a state-space approach. (Phd thesis), (p. 279). University of Twente.

Verdult, V., Bergboer, N., & Verhaegen, M. (2003). Identification of fully parameterized linear and nonlinear state-space systems by projected gradient search. In *Proc. of the 13th IFAC symposium on system identification* (pp. 737–742). Rotterdam, The Netherlands.

Verdult, V., Ljung, L., & Verhaegen, M. (2002). Identification of composite local linear state-space models using a projected gradient search. *International Journal of Control*, *75*(16/17), 1125–1153.

Verdult, V., & Verhaegen, M. (2002). Subspace identification of multivariable linear parameter-varying systems. *Automatica*, *38*(5), 805–814.
Verdult, V., & Verhaegen, M. (2005). Kernel methods for subspace identification of multivariable LPV and bilinear systems. *Automatica*, *41*, 1557–1565.
Verhaegen, M., & Verdult, V. (2007). *Filtering and system identification: a least squares approach*. Cambridge University Press.
Wills, A., & Ninness, B. (2008). On gradient-based search for multivariable system estimates. *IEEE Transactions on Automatic Control*, *53*(1), 298–306.
Wills, A., & Ninness, B. (2011). System identification of linear parameter varying state-space models. In P. Lopes dos Santos, T. P. Azevedo Perdicoúlis, C. Novara, J. A. Ramos, & D. E. Rivera (Eds.), *Linear parameter-varying system identification: New developments and trends* (pp. 295–316). World Scientific, ch. 11.

Pepijn Bastiaan Cox was born in Arnhem, The Netherlands, in 1989. He has received his B.Sc. degree in Mechanical Engineering (cum laude) in 2010 and his M.Sc. degree in Systems and Control Engineering (cum laude) in 2013, both at the Delft University of Technology (TUDelft), The Netherlands. In 2018, he obtained his Ph.D. degree in the Control Systems group at the Eindhoven University of Technology (TUe), The Netherlands. His Ph.D. topic was "Towards Efficient Identification of Linear Parameter-Varying State–Space Models" under the supervision of prof. dr. P.M.J. Van den Hof and dr. R. Tóth. In 2018, he was a postdoctoral researcher in the Control Systems group at the TUe. Currently, he works as a signal processing expert at the Radar Technology group at TNO the Netherlands. Pepijn Cox's main research interests are in linear parameter-varying (LPV) and nonlinear system modeling and identification as well as compressive sensing and numerical optimization.

Roland Tóth was born in 1979 in Miskolc, Hungary. In 2004, he received the B.Sc. degree in Electrical Engineering and the M.Sc. degree in Information Technology in parallel with cum laude distinction at the University of Pannonia, Veszprém, Hungary and, in 2008, the Ph.D. degree with cum laude distinction at the Delft Center for Systems and Control (DCSC), Delft University of Technology (TUDelft), Delft, The Netherlands, He was a Post-Doctoral Research Fellow at DCSC, TUDelft in 2009 and at the Berkeley Center for Control and Identification, University of California, Berkeley in 2010. He held a position at DCSC, TUDelft in 2011–12. Currently, he is an Associate Professor at the Control Systems Group, Eindhoven University of Technology (TU/e) and he is a senior researcher at the Systems and Control Laboratory, Institute for Computer Science and Control in Budapest, Hungary. He is an Associate Editor of the IEEE Transactions on Control Systems Technology. His research interests are in linear parameter-varying (LPV) and nonlinear system identification and control, machine learning for modeling and control, model predictive control and behavioral system theory. Dr. Tóth received the TUDelft Young Researcher Fellowship Award in 2010, the VENI award of The Netherlands Organisation for Scientific Research in 2011 and the Starting Grant of the European Research Council in 2016.