2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Minimal covariance realization and system identification algorithm for a class of stochastic linear switched systems with i.i.d. switching

Elie Rouphael, Manas Mejari, Mihaly Petreczky, Lotfi Belkoura

***Abstract*— In this paper, we consider stochastic realization theory of Linear Switched Systems (LSS) with i.i.d. switching. We characterize minimality of stochastic LSSs and show existence and uniqueness (up to isomorphism) of minimal LSSs in innovation form. We present a realization algorithm to compute a minimal LSS in innovation form from output and input covariances. Finally, based on this realization algorithm, by replacing true covariances with empirical ones, we propose a statistically consistent system identification algorithm.**

I. Introduction

In this paper we consider the stochastic realization problem for *stochastic linear switched systems (abbreviated as LSS)*, i.e. for state-space representations:

$$\mathcal{S}\begin{cases}\mathbf{x}(t+1)=A_{\boldsymbol{\theta}(t)}\mathbf{x}(t)+B_{\boldsymbol{\theta}(t)}\mathbf{u}(t)+K_{\boldsymbol{\theta}(t)}\mathbf{v}(t)\\ \mathbf{y}(t)=C\mathbf{x}(t)+D\mathbf{u}(t)+F\mathbf{v}(t)\end{cases}\tag{1}$$

where $A_\sigma\in\mathbb{R}^{n_x\times n_x}, B_\sigma\in\mathbb{R}^{n_x\times n_u}, K_\sigma\in\mathbb{R}^{n_x\times n_n}$, $\sigma\in\Sigma=\{1,\ldots,n_\mu\}$, $C\in\mathbb{R}^{n_y\times n_x}$, $F\in\mathbb{R}^{n_y\times n_n}$, $D\in\mathbb{R}^{n_y\times n_u}$ and $\mathbf{x}$, $\mathbf{v}$, $\mathbf{u}$, $\mathbf{y}$, $\boldsymbol{\theta}$ are the stochastic state, noise, input, output and switching processes. The process $\boldsymbol{\theta}$ takes values in the set of discrete states Σ. We call $\mathcal{S}$ from (1) a *realization* of the tuple $(\tilde{\mathbf{y}},\mathbf{u},\boldsymbol{\theta})$, if $\tilde{\mathbf{y}}=\mathbf{y}$. We call n_x the *dimension* of $\mathcal{S}$.

LSSs of the form (1) contain LTI systems as a special case, when $n_\mu=1$. Moreover, they correspond to *jump-Markov linear systems* if $\boldsymbol{\theta}$ is a Markov process [7]. Both switched and jump-Markov systems have a rich literature and a wide variety of applications, e.g., [30], [7]. We refer to (1) as switched rather than jump-Markov systems, because in contrast to the latter, in the formulation of the system identification problem, we treat the switching process as an external input, whose role is similar to that of $\mathbf{u}$. In contrast to the standard definition [7], [30] for the sake of simplicity, we assume that the matrices of the output equation do not depend on the current discrete mode.

Motivation and context: In order to motivate the contribution of the paper, we first define the system identification problem of LSSs. To this end, let us define the *deterministic behavior* $\mathcal{B}$ of a tuple $S=(\{A_i,B_i,K_i\}_{i=1}^{n_\mu},C,D,F)$ of matrices as the set of all deterministic signals (y,u,q), all defined on $\mathbb{Z}$, taking values in $\mathbb{R}^{n_y},\mathbb{R}^{n_u},\Sigma$ such that there exists a state trajectory x and noise realization v satisfying:

$$\begin{aligned}x(t+1)&=A_{q(t)}x(t)+B_{q(t)}u(t)+K_{q(t)}v(t)\\ y(t)&=Cx(t)+Du(t)+Fv(t)\end{aligned}\tag{2}$$

E. Rouphael, M. Petreczky and L. Belkoura are with Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France first.lastname@univ-lille.fr

M. Mejari is with Swiss AI lab IDSIA-SUPSI, Lugano, Switzerland. manas.mejari@supsi.ch

That is, $\mathcal{S}$ can be viewed as a stochastic version of (2), in particular, all samples paths of $(\mathbf{y},\mathbf{u},\boldsymbol{\theta})$ are elements of the deterministic behavior of S.

Assume that $\mathcal{S}$ from (1), referred to as the true system, is a realization of $(\mathbf{y},\mathbf{u},\boldsymbol{\theta})$, and we observe a finite portion $\{(y(t),u(t),q(t))\}_{t=0}^N$ of a sample path of $(\mathbf{y},\mathbf{u},\boldsymbol{\theta})$. The *identification* task is to find matrices $\hat{S}=(\{\hat{A}_i,\hat{B}_i,\hat{K}_i\}_{i=1}^{n_\mu},\hat{C},\hat{D},\hat{F})$ from $\{(y(t),u(t),q(t))\}_{t=0}^N$ such that in the limit, as $N\to\infty$, the deterministic behavior of $\hat{S}$ equals (approximates[1]) the deterministic behavior of the true systems. The motivation for this problem formulation is that in many control problems, the switching signal is viewed either as a control input or as an arbitrary external input/disturbance. That is, the output of the estimated model should be close to that of the true one for switching and inputs different from $\boldsymbol{\theta}$ and $\mathbf{u}$. For jump-Markov systems, this can be relaxed by requiring that the estimated model has the same output response as the true one for arbitrary inputs but for the fixed switching process $\boldsymbol{\theta}$. In particular, we think of the stochasticity of the observed signals as a persistency of excitation assumption, not an assumption on the desired model. Here persistence excitation is used in a general sense: an external signal is persistently exciting if the corresponding output response is sufficient to determine the input-output behavior of the underlying system.

One way to make the problem above well-posed is to ensure that both the estimated and true systems belong to a class of LSSs with the following property: if two elements of this class generate the same output for some stochastic input, switching and noise process, then their matrices differ only by a change of basis, i.e., they are isomorphic. Then if the output response of the estimated system is approximately the same as that of the true one for the designated processes $\mathbf{u}$ and $\boldsymbol{\theta}$, then the it is isomorphic to the true one, and hence the two systems have the same deterministic behavior.

For LTI systems, minimal systems in innovation form [17] represent such a class, as any two minimal LTI systems in innovation form realizing the same output are isomorphic. Moreover, under suitable conditions, any stochastic LTI system can be transformed into a minimal one in innovation form while preserving its output [17]. Hence, it is reasonable to assume that the true system is of this class. In addition, there are several algorithms which return minimal systems in innovation form, e.g., subspace identification methods [32], [15] and some parametric methods [18]. For the latter, the

[1]For the same input and switching signals, the output trajectories are close in a suitable distance ($\ell_1,\ \ell_2,\ \ell_\infty,$ etc.)

use of minimal systems in innovation form also justifies viewing LTI systems as optimal predictors of the current output based on past outputs and inputs. The uniqueness (up to isomorphism) of minimal systems in innovation form is also used for proving consistency of parametric [13], [14] and subspace methods [17], [5]. In fact, the covariance realization algorithm for LTI systems in innovation form [17] is the basis for subspace identification algorithms.

Contribution: We extend the concept of minimal (dimensional) systems in innovation form to a subclass LSSs and apply it to system identification. We assume that both $\boldsymbol{\theta}$ and $\mathbf{u}$ are i.i.d. processes. These assumptions represent the simplest case of persistently exciting data, and there is little hope to tackle the more general case without solving this one first.

We present a necessary and sufficient condition for minimality LSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$ in innovation form, and we show that all such realizations of an output are isomorphic. In addition, we present a realization algorithm for computing such LSSs from covariances of inputs and outputs.

Finally, we present a statistically consistent system identification algorithm which is based on the latter realization algorithm, and which returns, in the limit, a minimal LSS in innovation form. In particular, if the true system is a minimal LSS in innovation form, then in the limit this identification algorithm returns a LSS which is isomorphic to the true LSS, hence, it has the same deterministic behavior. We further improve this identification algorithm by combining it with Gradient-Based (GB) methods [9].

Related work: Realization theory of deterministic switched systems was studied in [23], for stochastic switched systems with no inputs in [24], [25], [28]. To the best of our knowledge, our results on realization theory of stochastic LSS with inputs are completely new.

Identification of switched and jump-Markov systems is an active research area, see the [16], [12], [3], [6], [1], [2], and the references therein. However, most of the literature assumes that the switching signal is unobserved, which makes the problem more challenging but also leads to lack of consistency results for state-space representations. A consistent subspace identification algorithms was presented in [22] for noiseless LSSs. In [29] a consistent identification algorithm for noisy LSSs was presented, but the noise gain matrix and the noise covariance were not estimated, and the identification was based on several i.i.d. time series.

LSSs can be viewed as a subclass of linear-parameter varying (LPV) systems, if the switching signal is viewed as a discrete scheduling. There is a wealth of literature on system identification for LPV systems, including subspace methods, [31], [10], [20], [34], [33] and the references therein. Stochastic realization theory of LPV systems was invetsigated in [20], [8]. In [8] the existence of LPV systems in innovation form was investigated, but the noise gain matrices of the obtained system depended on the current and past scheduling. Moreover, [8] did not address minimality and uniqueness of systems in innovation form. Unlike the identification algorithm of this paper, the existing subspace methods, see the overview [10], are not proven to be consistent or to return a minimal system in innovation form.

A statistically consistent identification algorithm was presented in [19] for noisy LPV systems with no inputs, and in [20] for noisy LPV systems with inputs. This paper is an extension of [20], the main difference w.r.t. [20] is as follows: **(1)** the conditions on the scheduling signals from [20] do not directly apply for i.i.d. switching signals (see Remark 1); **(2)** we characterize minimality and uniqueness (up to isomorphism) of systems in innovation form; **(3)** the proposed identification algorithm returns a minimal system in innovation form, while the one of [20] returns a non-minimal one; that is, the system returned by the algorithm from [20] need not have the deterministic behavior as the true system; **(4)** we present novel modifications to improve the empirical behavior of the identification algorithm.

To sum up, the identification algorithm of this paper is the first algorithm for noisy LSSs that has all of the following properties: **(a)** it estimates the noise gain and the noise covariance, **(b)** returns a minimal LSSs in innovation form, and **(c)** applies to one single long time series.

This paper uses technical results from [27] on the decomposition of outputs of LSSs into stochastic and deterministic components. The latter is used in [27] to show existence of a realization in innovation form and to provide *sufficient* conditions for minimality and uniqueness of LSSs in innovation form. In contrast, we formulate *necessary and sufficient* conditions for minimality, and we show isomorphism of a much wider class of LSSs in innovation forms. Moreover, in contrast to this paper, [27] proposes no realization or identification algorithm. A version of this paper with detailed proofs is availabe in the report [26].

II. TECHNICAL PRELIMINARIES

In this section we recall some concepts from [25] which we use to define a suitable subclass of LSSs.

Probability theory: We use the usual notation and terminology of probability theory [4]. All the random variables and stochastic processes are understood w.r.t. to a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. We use E for expected values. All the stochastic processes in this paper are discrete-time ones defined over the time-axis $\mathbb{Z}$ of the set of integers. We use bold letters for random variables and stochastic processes.

Switching process: The process $\boldsymbol{\theta}$ is independent, identically distributed taking values in $\Sigma = \{1, \ldots, n_\mu\}$, and

$$p_\sigma = \mathcal{P}(\boldsymbol{\theta}(t) = \sigma), \quad \sigma \in \Sigma. \tag{3}$$

Sequences of discrete modes: A *non empty word* over Σ is a finite sequence of letters, i.e., $w = \sigma_1\sigma_2 \cdots \sigma_k$, where $0 < k \in \mathbb{Z}$, $\sigma_1, \sigma_2, \ldots, \sigma_k \in \Sigma$. The set of *all* nonempty words is denoted by Σ^+. We denote an *empty word* by ϵ. Let $\Sigma^* = \epsilon \cup \Sigma^+$. We define the concatenation of words in the standard manner. The length of the word $w \in \Sigma^*$ is denoted by $|w|$, and $|\epsilon| = 0$.

Matrices and matrix products: Denote by I_n $n \times n$ identity matrix. Consider $n \times n$ square matrices $\{A_\sigma\}_{\sigma \in \Sigma}$. For the empty word ϵ, let $A_\epsilon = I_n$. for any word $w = \sigma_1\sigma_2 \cdots \sigma_k \in$

Σ^+, $k>0$ and $\sigma_1,\ldots,\sigma_k \in \Sigma$, we define

$$A_w = A_{\sigma_k} A_{\sigma_{k-1}} \cdots A_{\sigma_1}. \quad (4)$$

ZMWSII processes: Next, we recall the concept of zero-mean wide-sense stationary (ZMWSII) process w.r.t. $\boldsymbol{\theta}$ from [25]. To this end, we introduce the following indicator process $\boldsymbol{\mu}(t) = \begin{bmatrix}\boldsymbol{\mu}_1(t), & \ldots, & \boldsymbol{\mu}_{n_\mu}(t)\end{bmatrix}^T$ where

$$\boldsymbol{\mu}_\sigma(t) = \chi(\boldsymbol{\theta}(t) = \sigma) = \begin{cases} 1 & \text{if } \boldsymbol{\theta}(t) = \sigma \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where χ is the indicator function. The process $\boldsymbol{\mu}$ is admissible in the terminology of [25].

Remark 1 (Relationship with [20]): If we view $\boldsymbol{\mu}$ as a scheduling signal, then it does not satisfy the assumptions of [20]. By applying a suitable normalizing affine transformation to $\boldsymbol{\mu}$ we can bring it to a form which satisfies [20]. The system matrices of the thus arising LPV system are affine combinations of the system matrices of original LSS. However, this transformation does not appear to simplify the derivation of the results.

Next, we need the following products of components of $\boldsymbol{\mu}$. For every word $w \in \Sigma^+$ where $w = \sigma_1\sigma_2\cdots\sigma_k$, $k \geq 1$, $\sigma_1,\ldots,\sigma_k \in \Sigma$, we define the process $\boldsymbol{\mu}_w$ as follows:

$$\boldsymbol{\mu}_w(t) = \boldsymbol{\mu}_{\sigma_1}(t-k+1)\boldsymbol{\mu}_{\sigma_2}(t-k+2)\cdots\boldsymbol{\mu}_{\sigma_k}(t) \quad (6)$$

For an empty word $w = \epsilon$, we set $\boldsymbol{\mu}_\epsilon(t) = 1$. For the constants $\{p_\sigma\}_{\sigma\in\Sigma}$ from (3) define the products

$$p_w = p_{\sigma_1} p_{\sigma_2} \cdots p_{\sigma_k} \quad (7)$$

for any $w \in \Sigma^+$, $\sigma_i \in \Sigma$. For an empty word $w = \epsilon$, we set $p_\epsilon = 1$. For a stochastic process $\mathbf{r} \in \mathbb{R}^r$ and for each $w \in \Sigma^*$ we define the stochastic process $\mathbf{z}^{\mathbf{r}}_w$ as

$$\mathbf{z}^{\mathbf{r}}_w(t) = \mathbf{r}(t-|w|)\boldsymbol{\mu}_w(t-1)\frac{1}{\sqrt{p_w}}, \quad (8)$$

where $\boldsymbol{\mu}_w$ and p_w are as in (6) and (7). For $w = \epsilon$, $\mathbf{z}^{\mathbf{r}}_w(t) = \mathbf{r}(t)$. The process $\mathbf{z}^{\mathbf{r}}_w$ in (8) is interpreted as the product of the *past* of $\mathbf{r}$ and $\boldsymbol{\mu}$. The process $\mathbf{z}^{\mathbf{r}}_w$ will be used as predictors for future values of $\mathbf{r}$ for various choices of $\mathbf{r}$. We are now ready to recall from [25] the definition of ZMWSII process w.r.t. $\boldsymbol{\theta}$ (w.r.t; $\boldsymbol{\mu}$ in the terminology of [25]).

Definition 1 ([25]): A stochastic process $\mathbf{r}$ is *Zero Mean Wide Sense Stationary* (ZMWSSI) if

1. For $t \in \mathbb{Z}$, the σ-algebras generated by $\{\mathbf{r}(k)\}_{k\leq t}$, and $\{\boldsymbol{\mu}_\sigma(k)\}_{k\geq t,\sigma\in\Sigma}$ respectively are conditionally independent w.r.t σ-algebra generated by $\{\boldsymbol{\mu}_\sigma(k)\}_{k<t,\sigma\in\Sigma}$,
2. The processes $\{\mathbf{z}^{\mathbf{r}}_w\}_{w\in\Sigma^*}$ are zero mean, square integrable and jointly wide sense stationary, i.e. the covariances $\{E\left[\mathbf{z}^{\mathbf{r}}_w(t)(\mathbf{z}^{\mathbf{r}}_v(t))^T\right]\}_{w,v\in\Sigma^*}$ are independent of t.

The concept of ZMWSII is an extension of wide-sense stationarity, if Σ^+ is viewed as a time axis. More precisely, for any two ZMWSII processes $\mathbf{r}$ and $\mathbf{b}$ and for sequences $w, v \in \Sigma^*$, define the covariances:

$$\Lambda^{\mathbf{r},\mathbf{b}}_w = E[\mathbf{r}(t)(\mathbf{z}^{\mathbf{b}}_w(t))^T], \quad T^{\mathbf{r},\mathbf{b}}_{w,v} = E[\mathbf{z}^{\mathbf{r}}_w(t)(\mathbf{z}^{\mathbf{b}}_v(t))^T] \quad (9)$$

Then, $T^{\mathbf{r},\mathbf{r}}_{w,v} = \Lambda^{\mathbf{r},\mathbf{r}}_s$ if $w = sv$, and $T^{\mathbf{r},\mathbf{r}}_{w,v} = (\Lambda^{\mathbf{r},\mathbf{r}}_s)^T$ if $v = sw$, for some $s \in \Sigma^*$, and $T^{\mathbf{r},\mathbf{r}}_{w,v}=0$ otherwise, i.e., the covariance $T^{\mathbf{r},\mathbf{r}}_{w,v}$ depends on the difference between w and v.

III. Stationary LSSs

Below we present the definition of the class of LSSs which is studied in this paper. To this end, we recall from [27] the notion of a white noise process w.r.t. $\boldsymbol{\theta}$.

Definition 2 (White noise w.r.t. $\boldsymbol{\theta}$, [27]): A ZMWSII process $\mathbf{r}$ is a white noise w.r.t. $\boldsymbol{\theta}$, if for all $w \in \Sigma^+$, $v \in \Sigma^+$, $\Lambda^{\mathbf{r},\mathbf{r}}_w = 0$, and $T^{\mathbf{r},\mathbf{r}}_{w,v} = 0$ if $w \neq v$, and $T^{\mathbf{r},\mathbf{r}}_{w,w}$ is non-singular.

The notion of a white noise process w.r.t. $\boldsymbol{\theta}$ (w.r.t. $\boldsymbol{\mu}$ in the terminology of [27]) is an extension of the usual concept of the white noise process if Σ^+ is viewed as an additional time axis. In particular, if $\mathbf{r}$ is a white noise w.r.t. $\boldsymbol{\theta}$, the collection $\{\mathbf{z}^{\mathbf{r}}_w(t)\}_{w\in\Sigma^+}$ is a sequence of uncorrelated random variables and the covariance of $\mathbf{z}^{\mathbf{r}}_w(t)$ does not depend on t and it depends only on the first letter of w.

Next, we state the assumptions on the input and output.

Assumption 1 (Inputs and outputs): **(1)** $\mathbf{u}$ is a white noise w.r.t. $\boldsymbol{\theta}$ and the covariance $T^{\mathbf{u},\mathbf{u}}_{\sigma,\sigma} = Q_{\mathbf{u}}$ does not depend on $\sigma \in \Sigma$. **(2)** The process $\begin{bmatrix}\mathbf{y}^T, & \mathbf{u}^T\end{bmatrix}^T$ is a ZMWSSI.

Now, we are ready to define the class of stationary LSS, which is a special case of stationary generalized switched linear systems defined in [27].

Definition 3 (Stationary LSS): A *stationary LSS* (abbreviated sLSS) is a system (1), such that

1. $\mathbf{w} = \begin{bmatrix}\mathbf{v}^T, & \mathbf{u}^T\end{bmatrix}^T$ is a white noise process w.r.t. $\boldsymbol{\theta}$.
2. The process $\begin{bmatrix}\mathbf{x}^T, & \mathbf{w}^T\end{bmatrix}^T$ is a ZMWSSI, and $T^{\mathbf{x},\mathbf{w}}_{\sigma,\sigma} = 0$, $\Lambda^{\mathbf{x},\mathbf{w}}_w = 0$ for all $\sigma \in \Sigma$, $w \in \Sigma^+$.
3. The eigenvalues of the matrix $\sum_{\sigma\in\Sigma} p_\sigma A_\sigma \otimes A_\sigma$ are inside the open unit circle.

If $B_\sigma = 0$, $\sigma \in \Sigma$, and $D = 0$ the we call (1) an *autonomous stationary* LSS (asLSS), and in this case we say it is an asLSS realization of $(\mathbf{y}, \boldsymbol{\theta})$.

In the terminology of [25], an sLSS corresponds to a stationary **GBS** (Generalized Bilinear System) w.r.t. inputs $\{\boldsymbol{\mu}_\sigma\}_{\sigma\in\Sigma}$ and with noise $w = [\mathbf{v}^T, \mathbf{u}^T]^T$. In the terminology of [27], a sLSS is a stationary generalized switched system for which the switching process is i.i.d. Note that the processes $\mathbf{x}$ and $\mathbf{y}$ are ZMWSII, in particular, they are wide-sense stationary, and that $\mathbf{x}$ is orthogonal to the future values of the noise process $\mathbf{v}$. We need stationarity, as even for LTI case stochastic realization theory exists only for the stationary case [17].

As it was noted in [27], the state of an sLSS is uniquely determined by its matrices and noise and input process, i.e.,

$$\mathbf{x}(t) = \sum_{\sigma\in\Sigma, w\in\Sigma^*} \sqrt{p_{\sigma w}} A_w \left(K_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t) + B_\sigma \mathbf{z}^{\mathbf{u}}_{\sigma w}(t)\right) \quad (10)$$

where the right-hand side is absolutely convergent in the mean-square sense. This motivates the following notation.

Notation 1: We identify a sLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$ of the form (1) with the tuple $\mathcal{S} = (\{A_\sigma, B_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, D, F, \mathbf{v})$.

IV. Minimality of sLSS in innovation form

In this section, we present the main results on existence and uniqueness of minimal sLSSs in innovation form. To this end, first, we recall realization theory of deterministic linear switched systems, and then relate it with realization theory of sLSSs.

A. Review of deterministic realization theory of LSSs

A *deterministic linear switched system (abbreviated as dLSS)*, is a system of the form

$$\mathscr{S}\begin{cases} \mathrm{x}(t+1)=\mathcal{A}_{q(t)}\mathrm{x}(t)+\mathcal{B}_{q(t)}\mathrm{u}(t) \\ \mathrm{y}(t)=\mathcal{C}\mathrm{x}(t)+\mathcal{D}\mathrm{u}(t) \end{cases} \tag{11}$$

where $\{\mathcal{A}_\sigma,\mathcal{B}_\sigma\}_{\sigma\in\Sigma},\mathcal{C},\mathcal{D}$ are matrices of suitable dimensions, $q:\mathbb{Z}\to\Sigma$ is the switching signal, $\mathrm{x}:\mathbb{Z}\to\mathbb{R}^{n_\mathrm{x}}$ is the state trajectory $\mathrm{u}:\mathbb{Z}\to\mathbb{R}^{n_\mathrm{u}}$ is the input trajectory $\mathrm{y}:\mathbb{Z}\to\mathbb{R}^{n_\mathrm{y}}$ is the output trajectory with finite support [2]. We identify a dLSS of the form (11) with the tuple

$$\mathscr{S}=(\{\mathcal{A}_i,\mathcal{B}_i\}_{i=1}^{n_\mu},\mathcal{C},\mathcal{D}) \tag{12}$$

We refer to [23], [21] for a detailed definition input-output behavior of dLSSs, their realization theory, minimality, etc. In particular, let us call any function $M:\Sigma^*\to\mathbb{R}^{n_\mathrm{y}\times n_\mathrm{u}}$ a *Markov function*. The dLSS $\mathscr{S}$ *realizes* M, if for all $w\in\Sigma^*$,

$$M(w)=\begin{cases} \mathcal{C}\mathcal{A}_s\mathcal{B}_\sigma, & w=\sigma s,\ \sigma\in\Sigma,\ s\in\Sigma^* \\ \mathcal{D}. & w=\epsilon \end{cases} \tag{13}$$

If $\mathscr{S}$ is a realization of M, then M is referred to as the *Markov function of* $\mathscr{S}$ and it is denoted by $M_{\mathscr{S}}$. The values $\{M(w)\}_{w\in\Sigma^*}$ of $M_{\mathscr{S}}$ are the *Markov parameters* of $\mathscr{S}$. From [23], [21] it then follows that two dLSSs have the same input-output behavior, if and only if their Markov functions are equal. For dLSS (11) the integer n_x is called the *dimension* of $\mathscr{S}$, and we say that a dLSS is *minimal*, if there exists no other dLSS of smaller dimension which realizes the same Markov-function. From [23], [21], it follows that a dLSS is minimal if and only if it is span-reachable and observable, and the latter properties are equivalent to rank conditions of the extended reachability and observability matrices [21, Definition 23, Theorem 1, Theorem 2]. Furthermore, any dLSS can be transformed to a minimal one with the same Markov function, using a minimization algorithm [21, Procedure 3 1][3]. For a more detailed discussion see [21]. Moreover, any two minimal dLSS which are input-output equivalent are isomorphic, i.e. they are related by a linear change of coordinates [21, Theorem 1].

B. Relationship between sLSSs and dLSSs

The idea behind relating sLSSs and dLSSs is to express the covariances of outputs of sLSSs as Markov parameters of dLSSs. To this end, we need to recall [27] the decomposition of the output $\mathbf{y}$ into deterministic and stochastic parts, and the notion of orthogonal projection from [27].

Notation 2 (E_l): Recall from [4] that the set $\mathcal{H}_1$ of real valued square integrable random variables is a Hilbert-space with the scalar product $\langle \mathbf{z}_1,\mathbf{z}_2\rangle=E[\mathbf{z}_1\mathbf{z}_2]$. If M is a closed subspace of $\mathcal{H}_1$, then denote by $E_l[\mathbf{h}\,|\,M]$ the orthogonal projection, in the usual sense for Hilbert-spaces, of $\mathbf{h}\in\mathcal{H}_1$ onto M. Let $\mathbf{z}=(\mathbf{z}_1,\dots,\mathbf{z}_k)^T$ be a k-dimensional random variable such that $\mathbf{z}_i$ belongs to $\mathcal{H}_1$ for all $i=1,\dots,k$. The orthogonal projection of $\mathbf{z}$ onto M, *denoted by* $E_l[\mathbf{z}|M]$, is defined as k-dimensional random variable $(E_l[\mathbf{z}_1\mid M],E_l[\mathbf{z}_2|M],\dots,E_l[\mathbf{z}_k|M])^T$. If $\mathfrak{S}$ is a subset of vector valued random variables, coordinates of which all belong to $\mathcal{H}_1$, and M is generated by the coordinates of the elements of $\mathfrak{S}$, then instead of $E_l[\mathbf{z}\mid M]$ we use $E_l[\mathbf{z}\mid\mathfrak{S}]$.
Intuitively, $E_l[\mathbf{z}\mid\mathfrak{S}]$ is *minimal variance linear prediction* of $\mathbf{z}$ based on the elements of $\mathfrak{S}$.

The *deterministic component* $\mathbf{y}^d$ and the *stochastic component* $\mathbf{y}^s$ of $\mathbf{y}$ are defined as

$$\begin{aligned}\mathbf{y}^d(t)&=E_l[\mathbf{y}(t)\mid\{\mathbf{z}^{\mathbf{u}}_w(t)\}_{w\in\Sigma^+}\cup\{\mathbf{u}(t)\}],\\ \mathbf{y}^s(t)&=\mathbf{y}(t)-\mathbf{y}^d(t).\end{aligned} \tag{14}$$

Intuitively, $\mathbf{y}^d$ is the best linear prediction of $\mathbf{y}$ based on the present and past values of $\mathbf{u}$ multiplied by indicator functions of past discrete modes. In fact, by [27, Theorem 1], $\mathbf{y}^d$ (resp. $\mathbf{y}^s$) is the output of the asLSS obtained from (1), by setting $K_\sigma=0$ (resp. $B_\sigma=0$), for all $\sigma\in\Sigma$

Define the Markov function $M_{\mathbf{y},\mathbf{u}}:\Sigma^*\to\mathbb{R}^{n_y\times(n_y+n_u)}$

$$M_{\mathbf{y},\mathbf{u}}(w)=\begin{cases}\begin{bmatrix}\Lambda^{\mathbf{y}^d,\mathbf{u}}_w Q_{\mathbf{u}}^{-1}, & \Lambda^{\mathbf{y}^s}_w\end{bmatrix} & w\neq\epsilon\\ \begin{bmatrix}\Lambda^{\mathbf{y}^d,\mathbf{u}}_\epsilon Q_{\mathbf{u}}^{-1}, & I_{n_y}\end{bmatrix} & w=\epsilon\end{cases} \tag{15}$$

It can be shown that the first components of $M_{\mathbf{y},\mathbf{u}}$ can be expressed as Markov parameters of a suitably defined dLSS. Since, by [27], $\mathbf{y}^s$ is the output of an asLSS, and hence of a GBS in terminology of [25], then by [25, Lemma 4] $\Lambda^{\mathbf{y}^s}_w$ can be shown to be the Markov function of a suitable dLSS. That is, $M_{\mathbf{y},\mathbf{u}}$ has a dLSS realization.

Formally, define the process $\mathbf{x}^s(t)=\mathbf{x}(t)-E_l[\mathbf{x}(t)\mid\{\mathbf{z}^{\mathbf{u}}_w(t)\}_{w\in\Sigma^+}\cup\{\mathbf{u}(t)\}]$ and the matrices

$$G_\sigma=\sqrt{p_\sigma}(A_\sigma T^{\mathbf{x}^s,\mathbf{x}^s}_{\sigma,\sigma}C^T+K_\sigma T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma}F^T) \tag{16}$$

Then define the *dLSS associated with sLSS* as

$$\mathscr{S}_{\mathcal{S}}=(\{\sqrt{p_\sigma}A_\sigma,\begin{bmatrix}\sqrt{p_\sigma}B_\sigma & G_\sigma\end{bmatrix}\}_{\sigma=1}^{n_\mu},C,\begin{bmatrix}D & I_{n_y}\end{bmatrix}),$$

Note that by [27, proof of Theorem 1] $\mathbf{x}^s(t)$ is the state of the asLSS obtained from $\mathcal{S}$ by considering $B_\sigma=0$, and hence $\mathbf{x}^s$ is a ZMWSII process and $T^{\mathbf{x}^s,\mathbf{x}^s}_{\sigma,\sigma}$ is well-defined.

Lemma 1: If $\mathcal{S}$ is a sLSS realization of $(\mathbf{y},\mathbf{u},\boldsymbol{\mu})$, then the associated dLSS $\mathscr{S}_{\mathcal{S}}$ is a realization of $M_{\mathbf{y},\mathbf{u}}$.
The proof, see [26], uses [25, Lemma 4], [27, Theorem 2].

Remark 2 (Computing G_σ): In order to compute G_σ, we can use that by [27, proof Theorem 1] $\mathbf{x}^s$ is the state of the asLSS obtained from (1) by setting $B_\sigma=0$, $\sigma\in\Sigma$. Hence, by [25, Lemma 5], the covariance of $\mathbf{x}^s$ satisfies $p_\sigma T^{\mathbf{x}^s,\mathbf{x}^s}_{\sigma,\sigma}=\lim_{N\to\infty}P^N_\sigma$, where $P^0_\sigma=0$ and $P^{N+1}_\sigma=p_\sigma\sum_{\sigma_1\in\Sigma}(A_{\sigma_1}P^N_{\sigma_1}A^T_{\sigma_1}+p_{\sigma_1}K_{\sigma_1}T^{\mathbf{v},\mathbf{v}}_{\sigma_1,\sigma_1}K^T_{\sigma_1})$.
Conversely, we can associate with any dLSS realization of the Markov function $M_{\mathbf{y},\mathbf{u}}$ a sLSS realization of $(\mathbf{y},\mathbf{u},\boldsymbol{\theta})$. In order to define the latter, we need to specify its noise process,

[2] A function $g:\mathbb{Z}\longrightarrow\mathbb{R}^p$ has a finite support, then it $\exists\ t_0\in\mathbb{Z}$, such that $\forall\ t<t_0$, $g(t)=0$

[3] [21, Corollary 1] should be applied with zero initial state

which happens to be the *innovation noise* $\mathbf{e}^s$ of $\mathbf{y}^s$ as defined in [25, eq. (16)]. By [27, Theorem 2],

$$\begin{aligned}\mathbf{e}^s(t) &= \mathbf{y}(t) - \hat{\mathbf{y}}(t)\\ \hat{\mathbf{y}}(t) &= E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w^{\mathbf{y}}(t), \mathbf{z}_w^{\mathbf{u}}(t)\}_{w\in\Sigma^+} \cup \{\mathbf{u}(t)\}]\end{aligned} \tag{17}$$

That is, $\mathbf{e}^s(t)$ is the prediction error of the best linear predictor $\hat{\mathbf{y}}(t)$ of $\mathbf{y}(t)$ given the products of past outputs, inputs and discrete modes, i.e., $\mathbf{e}^s$ is direct extension of the classical innovation noise. We say that $\mathbf{y}$ is *full rank*, if for all $\sigma \in \Sigma$, the covariance $T_{\sigma,\sigma}^{\mathbf{e}^s,\mathbf{e}^s}$ is invertible. This is a direct extension of the classical notion of a full rank process. If $\mathbf{y}$ is full rank, for an observable dLSS

$$\mathscr{S} = (\{\hat{A}_\sigma, \begin{bmatrix}\hat{B}_\sigma & \hat{G}_\sigma\end{bmatrix}\}_{\sigma=1}^{n_\mu}, \hat{C}, \begin{bmatrix}D & I_{n_y}\end{bmatrix})$$

for which $\sum_{\sigma=1}^{n_\mu} \hat{A}_\sigma \otimes \hat{A}_\sigma$ is a Schur matrix[4], define the *sLSS $\mathcal{S}_{\mathscr{S}}$ associated with $\mathscr{S}$ as*

$$\mathcal{S}_{\mathscr{S}} = (\{\frac{1}{\sqrt{p_i}}\hat{A}_\sigma, \frac{1}{\sqrt{p_\sigma}}\hat{B}_\sigma, \hat{K}_\sigma\}_{\sigma=1}^{n_\mu}, \hat{C}, \hat{D}, I_{n_y}, \mathbf{e}^s),$$

where $\hat{K}_\sigma = \lim_{\mathcal{I}\to\infty} \hat{K}_\sigma^{\mathcal{I}}$, and $\{\hat{K}_\sigma^{\mathcal{I}}\}_{\sigma\in\Sigma,\mathcal{I}\in\mathbb{N}}$ satisfies

$$\begin{aligned}\hat{P}_\sigma^{\mathcal{I}+1} &= p_\sigma \sum_{\sigma_1\in\Sigma} \frac{1}{p_{\sigma_1}} \left(\hat{A}_{\sigma_1}\hat{P}_{\sigma_1}^{\mathcal{I}}\hat{A}_{\sigma_1}^T + \hat{K}_{\sigma_1}^{\mathcal{I}}\hat{Q}_{\sigma_1}^{\mathcal{I}}(\hat{K}_{\sigma_1}^{\mathcal{I}})^T\right)\\ \hat{Q}_\sigma^{\mathcal{I}} &= p_\sigma T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s} - \hat{C}\hat{P}_\sigma^{\mathcal{I}}(\hat{C})^T\\ \hat{K}_\sigma^{\mathcal{I}} &= \left(\hat{G}_\sigma\sqrt{p_\sigma} - \frac{1}{\sqrt{p_\sigma}}\hat{A}_\sigma\hat{P}_\sigma^{\mathcal{I}}(\hat{C})^T\right)\left(\hat{Q}_\sigma^{\mathcal{I}}\right)^{-1}\end{aligned} \tag{18}$$

with $\hat{P}_\sigma^0 = 0$. Note that the noise process of $\mathcal{S}_{\mathscr{S}}$ is the innovation noise, and its noise and state covariances satisfy

$$p_\sigma T_{\sigma,\sigma}^{\mathbf{e}^s,\mathbf{e}^s} = \lim_{\mathcal{I}\to\infty} \hat{Q}_\sigma^{\mathcal{I}}, \quad p_\sigma T_{\sigma,\sigma}^{\hat{\mathbf{x}},\hat{\mathbf{x}}} = \lim_{\mathcal{I}\to\infty} \hat{P}_\sigma^{\mathcal{I}} \tag{19}$$

where $\hat{\mathbf{x}}$ is the unique state process of $\mathcal{S}_{\mathscr{S}}$. The convergence of (18) and (19) follow from the proof of [25, Theorem 3].

Lemma 2: If $\mathbf{y}$ is full rank and $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$ has a realization by sLSS, and $\mathscr{S}$ is a minimal dLSS realization of $M_{\mathbf{y},\mathbf{u}}$, then the associated sLSS $\mathcal{S}_{\mathscr{S}}$ is a sLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$. The proof, see [26], uses [27, Theorem 2], [25, Theorem 3].

C. Main result on minimal sLSSs in innovation form

A sLSS of the form (1) is said to be in *innovation form*, if its noise process $\mathbf{v}$ equals $\mathbf{e}^s$ from (17) and $F = I_{n_y}$. Note that the sLSS associated with a dLSS is in innovation form. The sLSS (1) in innovation form is a predictor

$$\begin{aligned}\mathbf{x}(t+1) &= (A_{\boldsymbol{\theta}(t)} - K_{\boldsymbol{\theta}(t)}C)\mathbf{x}(t) + B_{\boldsymbol{\theta}(t)}\mathbf{u}(t) + K_{\boldsymbol{\theta}(t)}\mathbf{y}_u(t)\\ \hat{\mathbf{y}}(t) &= C\mathbf{x}(t) + D\mathbf{u}(t), \ \mathbf{y}_u(t) = \mathbf{y}(t) - D\mathbf{u}(t)\end{aligned}$$

which generates the best linear prediction $\hat{\mathbf{y}}(t)$ of the output $\mathbf{y}(t)$, see (17), while being driven by past outputs and inputs.

We say that a sLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$ is *minimal*, if it has the smallest dimension among all sLSS realizations of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$. In order to study minimality, we need the concepts of reachability and observability. We call a sLSS $\mathcal{S}$ *reachable* (resp. *observable*), if the associated dLSS $\mathscr{S}_{\mathcal{S}}$ is span-reachable, (resp. observable) according to the definition of [21, Definition 19], applied to zero initial state.

[4]All its eigenvalues are inside the unit disk

In order to investigate uniqueness of minmal sLSS in innovation form, we need the concept of isomorphism Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be sLSSs realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form such that $\mathcal{S}$ is of the form (1) and $\widetilde{\mathcal{S}} = (\{\tilde{A}_\sigma, \tilde{B}_\sigma, \tilde{K}_\sigma\}_{\sigma=1}^{n_\mu}, \tilde{C}, \tilde{D}, I_{n_y}, \mathbf{e}^s)$. We say that $\mathcal{S}$ and $\widetilde{S}$ are *isomorphic*, if there exists a nonsingular matrix T such that

$$\begin{aligned}&\tilde{A}_\sigma = TA_\sigma T^{-1}, \tilde{K}_\sigma = TK_\sigma, \ \tilde{B}_\sigma = TB_\sigma, \quad \sigma \in \Sigma\\ &\tilde{C} = CT^{-1}, D = \tilde{D}\end{aligned} \tag{20}$$

Theorem 1 (Main result: minimal innovation form): Assume that $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$ has a sLSS realization and that $\mathbf{y}$ is full rank w.r.t. $\mathbf{u}$ and $\boldsymbol{\theta}$. Then the following holds.

1. A sLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$ is minimal if and only if it is reachable and observable.
2. If a dLSS $\mathscr{S}_m$ is a minimal realization of $M_{\mathbf{y},\mathbf{u}}$, then the associated sLSS $\mathcal{S}_{\mathscr{S}_m}$ is a minimal sLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$ in innovation form.
3. There exists a minimal sLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$ in innovation form.
4. Any two minimal sLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$ in innovation form are isomorphic.

The proof is presented in [26], it relies on the fact that any dLSS can be transformed to a minimal one, that minimal input-output dLSS are isomorphic and on Lemma 1-2.

Moreover, an sLSS can be transformed to a minimal one in innovation form as follows: first we compute the associated dLSS, we minimize it using [21, Procedure 3], and then we compute the sLSS associated with the latter minimal dLSS.

Remark 3 (Relationship with [27]): For a sLSS (1), the minimality condition of [27] is a sufficient condition for reachability and observability of the associated dLSS. Likewise, [27] shows isomorphism between sLSSs in innovation form which satisfy the above sufficient condition for minimality, and for which the image of B_σ belongs to the image of K_σ. In contrast, Theorem 1 establishes isomorphism for a much wider class of sLSSs in innovation form.

V. Minimial covariance realization algorithm

In this section we present a Ho-Kalman-like algorithm for computing a minimal sLSS realization innovation form from $M_{\mathbf{y},\mathbf{u}}$. To this end, we first recall the Ho-Kalman realization algorithm for dLSS.

A. Reduced basis Ho-Kalman algorithm for dLSS

Below we recall from [8] an adaptation of the Ho-Kalman-like algorithm. Define a (n, n_y, n_u)-*selection (selection for short)* as a pair of word-index sets (α, β) such that,

$$\alpha = \{(v_i, k_i)\}_{i=1}^n, \ \beta = \{(\sigma_j, \nu_j, l_j)\}_{j=1}^n, \tag{21}$$

for some $v_i \in \Sigma^+$, $|v_i| \le n$, $k_i \in \{1, 2, \cdots, n_y\}$, $\sigma_j \in \Sigma$, $\nu_j \in \Sigma^+$, $|\nu_j| \le n$, $l_j \in \{1, 2, \cdots, n_u\}$, $i, j \in \{1, \ldots, n\}$ Intuitively, the components of α, β determine the choice of Markov parameters, while the other components determine a choice of row and column indices of the chosen Markov parameters, see [20] for an example. Formally, let $M : \Sigma^* \to \mathbb{R}^{n_y \times n_u}$ be a Markov function. Define the Hankel matrix

$\mathcal{H}^M_{\alpha,\beta} \in \mathbb{R}^{n\times n}$ and the matrices $\mathcal{H}^M_{\sigma,\alpha,\beta} \in \mathbb{R}^{n\times n}$, $\mathcal{H}^M_{\alpha,\sigma} \in \mathbb{R}^{n\times n_\mathrm{u}}$ and $\mathcal{H}^M_{\beta} \in \mathbb{R}^{n_\mathrm{y}\times n}$ as follows:

$$\left[\mathcal{H}^M_{\alpha,\beta}\right]_{i,j} = [M(\sigma_j\nu_j\upsilon_i)]_{k_i,l_j}, \quad i,j=1,\ldots,n \tag{22}$$

$$\left[\mathcal{H}^M_{\sigma,\alpha,\beta}\right]_{i,j} = [M(\sigma_j\nu_j\sigma\upsilon_i)]_{k_i,l_j}, \ i,j=1,\ldots,n \tag{23}$$

$$\left[\mathcal{H}^M_{\alpha,\sigma}\right]_{i,j} = [M(\sigma\upsilon_i)]_{k_i,j}, \ j=1,\ldots,n_\mathrm{u}, i=1,\ldots,n \tag{24}$$

$$\left[\mathcal{H}^M_{\beta}\right]_{i,j} = [M(\sigma_j\nu_j)]_{i,l_j}, \ i=1,\ldots,n_\mathrm{y}, j=1,\ldots n \tag{25}$$

where $(\upsilon_i,k_i)\in\alpha, (\sigma_j,\nu_j,l_j)\in\beta$ are as in the ordering in (21). Finally, let $\mathcal{L}_{\alpha,\beta}$ be the set of all words $w\in\Sigma^*$ such that $M(w)$ occurs in one of the matrices, (22) – (23), i.e.,

$$\mathcal{L}_{\alpha,\beta} = \{\upsilon_i, \sigma_j\nu_j, \sigma_j\nu_j\upsilon_i, \sigma_j\nu_j\sigma\upsilon_i, \sigma\upsilon_i \mid \sigma\in\Sigma\}_{i,j=1}^n$$

The promised algorithm is presented in Algorithm 1.

Algorithm 1 Ho-Kalman for dLSSs

Input: $(n,n_\mathrm{y},n_\mathrm{u})$-selection (α,β), $\{M(w)\}_{w\in\mathcal{L}_{\alpha,\beta}}$, $M(\epsilon)$

Construct the matrices $\mathcal{H}^M_{\alpha,\beta}$, $\mathcal{H}^M_{\sigma,\alpha,\beta}$ $\mathcal{H}^M_{\alpha,\sigma}$ and $\mathcal{H}^M_{\beta}$, and set

$$\hat{A}_\sigma = (\mathcal{H}^M_{\alpha,\beta})^{-1}\mathcal{H}^M_{\sigma,\alpha,\beta}, \ \hat{B}_\sigma = (\mathcal{H}^M_{\alpha,\beta})^{-1}\mathcal{H}^M_{\alpha,\sigma}, \ \hat{C} = \mathcal{H}^M_{\beta}.$$

Output: dLSS $(\{\hat{A}_\sigma,\hat{B}_\sigma\}_{\sigma=1}^{n_\mu}, \hat{C}, M(\epsilon))$

Lemma 3 ([8]): Assume that there exists a minimal dLSS realization of M of dimension n. Then there exists an $(n,n_\mathrm{y},n_\mathrm{u})$-selection (α,β) such that rank$(\mathcal{H}^M_{\alpha,\beta})=n$. For any $(n,n_\mathrm{y},n_\mathrm{u})$-selection (α,β) for which rank$(\mathcal{H}^M_{\alpha,\beta})=n$, Algorithm 1 returns a minimal dLSS realization of M.

B. Covariance realization algorithm

Theorem 1, together with Algorithm 1 and Lemma 3 suggests that in order to compute a sLSS realization of $(\mathbf{y},\mathbf{u},\boldsymbol{\theta})$, one could apply Algorithm 1 to Hankel-matrices corresponding to $M_{\mathbf{y},\mathbf{u}}$. However, the definition of $M_{\mathbf{y},\mathbf{u}}$ uses the covariances of the processes $\mathbf{y}^d$ and $\mathbf{y}^s$, which are not directly available. Below we show how to compute $M_{\mathbf{y},\mathbf{u}}$ using only the covariances of $\mathbf{y}$ and $\mathbf{u}$.

To this end, define the Markov function $\Psi_{\mathbf{u},\mathbf{y}}$ such that $\Psi_{\mathbf{u},\mathbf{y}}(w)$ is formed by the first n_u columns of $M_{\mathbf{y},\mathbf{u}}(w)$, i.e. $\Psi_{\mathbf{u},\mathbf{y}}(w)=\Lambda^{\mathbf{y}^d,\mathbf{u}}_w Q_u^{-1}$. Then $\Psi_{\mathbf{u},\mathbf{y}}$ has a realization by a dLSS: by Lemma 1 $(\{\sqrt{p_\sigma}A_\sigma,\sqrt{p_\sigma}B_\sigma\}_{\sigma=1}^{n_\mu},C,D)$ is a dLSS realization of $\Psi_{\mathbf{u},\mathbf{y}}$.

Lemma 4: If there exists a sLSS of $(\mathbf{y},\mathbf{u},\boldsymbol{\theta})$. Then

$$\begin{aligned}&\Lambda^{\mathbf{y}^d,\mathbf{u}}_w=\Lambda^{\mathbf{y},\mathbf{u}}_w, \ \Lambda^{\mathbf{y}^s,\mathbf{y}^s}_{\sigma w}=\Lambda^{\mathbf{y},\mathbf{y}}_{\sigma w}-\Lambda^{\mathbf{y}^d,\mathbf{y}^d}_{\sigma w},\\ &T^{\mathbf{y}^s,\mathbf{y}^s}_{\sigma,\sigma}=T^{\mathbf{y},\mathbf{y}}_{\sigma,\sigma}-T^{\mathbf{y}^d,\mathbf{y}^d}_{\sigma,\sigma}\end{aligned} \tag{26}$$

for all $w\in\Sigma^*$, $\sigma\in\Sigma$. Moreover, if $(\{\tilde{A}^d_\sigma,\tilde{B}^d_\sigma\}_{\sigma=1}^{n_\mu},\tilde{C}^d,\tilde{D}^d)$ is a minimal dLSS realization of $\Psi_{\mathbf{u},\mathbf{y}}$, then

$$\begin{aligned}\Lambda^{\mathbf{y}^d,\mathbf{y}^d}_{\sigma w} &= \frac{1}{p_{\sigma w}}\tilde{C}^d\tilde{A}^d_w(\tilde{A}^d_\sigma\tilde{P}_\sigma(\tilde{C}^d)^T+\tilde{B}^d_\sigma Q_u(\tilde{D}^d)^T)\\ T^{\mathbf{y}^d,\mathbf{y}^d}_{\sigma,\sigma} &= \frac{1}{p_\sigma}(\tilde{C}^d\tilde{P}_\sigma(\tilde{C}^d)^T+\tilde{D}^dQ_u(\tilde{D}^d)^T)\end{aligned} \tag{27}$$

where $\tilde{P}_\sigma = \lim_{\mathcal{I}\to\infty}\tilde{P}^{\mathcal{I}}_\sigma$, and $\tilde{P}^0_\sigma=0$ and $\tilde{P}^{\mathcal{I}+1}_\sigma = p_\sigma\sum_{\sigma_1\in\Sigma}\left(\frac{1}{p_{\sigma_1}}\tilde{A}^d_{\sigma_1}\tilde{P}^{\mathcal{I}}_{\sigma_1}(\tilde{A}^d_{\sigma_1})^T+\tilde{B}^d_{\sigma_1}Q_u(\tilde{B}^d_{\sigma_1})^T\right)$.

The proof of Lemma is presented in [26]. The first part of the lemma relies on orthogonality of $\mathbf{y}^s(t)$ and $\{\mathbf{z}^{\mathbf{u}}_w(t)\}_{w\in\Sigma^*}$ established in [27]. The second part follows by identifying dLSS realizations of $\Psi_{\mathbf{u},\mathbf{y}}$ with asLSS realizations of $\mathbf{y}^d$ whose noise is $\mathbf{u}$, and using the equations in [25, Lemma 4] for the covariances of $\mathbf{y}^d$.

We can then use Lemma 4 to compute the necessary values of $M_{\mathbf{y},\mathbf{u}}$ as follows. First, we compute a minimal dLSS realization of $\Psi_{\mathbf{u},\mathbf{y}}$ using Algorithm 1, and then use (26) – (27) to compute those values of $\Lambda^{\mathbf{y}^s,\mathbf{y}^s}_w$, and hence of $M_{\mathbf{y},\mathbf{u}}$, which are necessary for applying Algorithm 1 to $M_{\mathbf{y},\mathbf{u}}$. This idea is formalized in Algorithm 2.

Algorithm 2 Minimal covariance realization algorithm

Input: (n_x,n_y,n_u+n_y)-selection (α,β); $(\bar{n},n_y,n_u)$-selection $(\bar{\alpha},\bar{\beta})$; covariances $\{\Lambda^{\mathbf{y},\mathbf{u}}_w\}_{w\in\mathcal{L}_{\alpha,\beta}\cup\mathcal{L}_{\bar{\alpha},\bar{\beta}}}$, $\{\Lambda^{\mathbf{y},\mathbf{y}}_w\}_{w\in\mathcal{L}_{\alpha,\beta}}$, $\Lambda^{\mathbf{y},\mathbf{u}}_\epsilon$, $\{T^{\mathbf{y},\mathbf{y}}_{\sigma,\sigma}\}_{\sigma\in\Sigma}$; integer $\mathcal{I}>0$.

1. Use (26) to compute $\{\Psi_{\mathbf{u},\mathbf{y}}(w)\}_{w\in\mathcal{L}_{\bar{\alpha},\bar{\beta}}}$ and $\Psi_{\mathbf{u},\mathbf{y}}(\epsilon)$.
2. Run Algorithm 1 with $M=\Psi_{\mathbf{u},\mathbf{y}}$ and selection $(\bar{\alpha},\bar{\beta})$ and denote the result by $\mathscr{S}_d=(\{\tilde{A}_i,\tilde{B}_i\}_{i=1}^{n_\mu},\tilde{C},\tilde{D})$.
3. Compute $\{\Lambda^{\mathbf{y}^s,\mathbf{y}^s}_w,\Lambda^{\mathbf{y}^d,\mathbf{u}}_w\}_{w\in\mathcal{L}_{\alpha,\beta}}$, using (26), (27) and $\{\Lambda^{\mathbf{y},\mathbf{y}^s}_w\}_{w\in\mathcal{L}_{\alpha,\beta}}$.
4. Compute $M_{\mathbf{y},\mathbf{u}}(w)$ from $\{\Lambda^{\mathbf{y}^s,\mathbf{y}^s}_w,\Lambda^{\mathbf{y}^d,\mathbf{u}}_w\}$ for $w\in\mathcal{L}_{\alpha,\beta}$.
5. Run Algorithm 1 with $M=M_{\mathbf{y},\mathbf{u}}$ and selection (α,β), and denote by $\mathscr{S}=(\{\hat{A}_i,\hat{B}_i,\hat{G}_i\}_{i=1}^{n_\mu},\hat{C},\hat{D})$ the result.
6. From $\{T^{\mathbf{y},\mathbf{y}}_{\sigma,\sigma}\}_{\sigma\in\Sigma}$ compute $\{T^{\mathbf{y}^s,\mathbf{y}^s}_{\sigma,\sigma}\}_{\sigma\in\Sigma}$ using (26) and (27). Let $\hat{K}^{\mathcal{I}}_\sigma$, $\hat{Q}^{\mathcal{I}}_\sigma$ as in (18).

Output: Matrices $\{\frac{\hat{A}_\sigma}{\sqrt{p_\sigma}},\frac{\hat{B}_\sigma}{\sqrt{p_\sigma}},\hat{K}^{\mathcal{I}}_\sigma,\hat{Q}^{\mathcal{I}}_\sigma\}_{\sigma=1}^{n_\mu},\hat{C},\hat{D}$.

Corollary 1: Assume that $\mathbf{y}$ is full rank and there exists a minimal sLSS realization of $(\mathbf{y},\mathbf{u},\boldsymbol{\theta})$ of dimension n_x, and a minimal dLSS realization of $\Psi_{\mathbf{y},\mathbf{u}}$ of dimension $\bar{n}$. If the selections (α,β) and $(\bar{\alpha},\bar{\beta})$ satisfy

$$\mathrm{rank}H^{M_{\mathbf{y},\mathbf{u}}}_{\alpha,\beta}=n_x, \quad \mathrm{rank}H^{\Psi_{\mathbf{u},\mathbf{y}}}_{\bar{\alpha},\bar{\beta}}=\bar{n}. \tag{28}$$

then Algorithm 2 returns matrices such that the sLSS $\tilde{S}=(\{\hat{A}_\sigma/\sqrt{p_\sigma},\hat{B}_\sigma/\sqrt{p_\sigma},\hat{K}_\sigma,\}_{\sigma=1}^{n_\mu},\hat{C},\hat{D},\mathbf{e}^s)$ is a minimal realization of $(\mathbf{y},\mathbf{u},\boldsymbol{\theta})$ in innovation form, where $\hat{K}_\sigma = \lim_{\mathcal{I}\to\infty}\hat{K}^{\mathcal{I}}_\sigma$ and $T^{\mathbf{e}^s,\mathbf{e}^s}_{\sigma,\sigma}p_\sigma=\lim_{\mathcal{I}\to+\infty}\hat{Q}^{\mathcal{I}}_\sigma$, $\sigma\in\Sigma$. Moreover, there exist an (n_x,n_y,n_u+n_y)-selection (α,β) and $(\bar{n},n_y,n_u)$-selection $(\bar{\alpha},\bar{\beta})$ which satisfy (28).
For the proof of Corollary 1 see [26].

That is, Algorithm 2 returns a minimal sLSS in innovation form based on the input and output covariances, if the selections (α,β), $(\bar{\alpha},\bar{\beta})$ give Hankel-matrices of a correct rank, and there always exists such selections.

That is, by finding a minimal sLSS of $(\mathbf{y},\mathbf{u},\boldsymbol{\theta})$ in innovation form, Algorithm 2 finds a sLSS which has the same deterministic behavior as the true system, i.e. the same output response as the true system for any inputs, noise and switching signals, if the true system is assumed to be minimal in innovation form.

VI. Identification algorithm

We formulate an identification algorithm based on Algorithm 2, by using empirical covariances instead of the true ones $\Lambda_w^{\mathbf{y},\mathbf{u}}$, $\Lambda_w^{\mathbf{y},\mathbf{y}}$, $T_{\sigma,\sigma}^{\mathbf{y},\mathbf{y}}$ when applying Algorithm 2. To this end, we make the following assumptions.

Assumption 2: **(1)** The $(n_x, n_y, n_u + n_y)$-selection (α, β) and the $(\bar{n}, n_y, n_u)$-selection $(\bar{\alpha}, \bar{\beta})$ satisfy (28), $\mathbf{y}$ is full rank, and n_x and $\bar{n}$ satisfy the hypothesis of Corollary 1.

(2) The process $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$ is ergodic, and the observed sample paths $y : \mathbb{Z} \to \mathbb{R}^{n_y}$, $u : \mathbb{Z} \to \mathbb{R}^{n_u}$ and $q : \mathbb{Z} \to \Sigma$ of $\mathbf{y}$, $\mathbf{u}$ and $\boldsymbol{\theta}$ respectively satisfy the following. For all $w, v \in \mathcal{L}_{\alpha,\beta} \cup \mathcal{L}_{\bar{\alpha},\bar{\beta}} \cup \Sigma \cup \{\epsilon\}$ define the *empirical covariances*

$$T_{v,w,N}^{\mathbf{y},\mathbf{b}} = \frac{\sum_{t=N_0}^{N} z_v^y(t)(z_w^b(t))^T}{N - N_0}, \tag{29}$$

where $b = \begin{cases} y & \mathbf{b} = \mathbf{y} \\ u & \mathbf{b} = \mathbf{u} \end{cases}$, and for all $\mu_\sigma(t) = \chi(q(t) = \sigma)$, $\sigma \in \Sigma$, and for all $w = \sigma_1\sigma_2 \ldots \sigma_r \in \Sigma^+$, $r > 0$, $\sigma_1, \ldots, \sigma_r \in \Sigma$, $b \in \{y, u\}$

$$\mu_w(t) = \mu_{\sigma_1}(t - k + 1)\mu_{\sigma_2}(t - k + 2) \cdots \mu_{\sigma_r}(t)$$
$$z_w^b(t) = b(t - r)\mu_w(t - 1)\frac{1}{\sqrt{p_w}}, \quad z_\epsilon^b(t) = b(t),$$

and N_0 is an upper bound on the length of words in $\mathcal{L}_{\alpha,\beta} \cup \mathcal{L}_{\bar{\alpha},\bar{\beta}} \cup \Sigma$. Then we assume that for all $\mathbf{b} \in \{\mathbf{y}, \mathbf{u}\}$,

$$\Lambda_w^{\mathbf{y},\mathbf{u}} = \lim_{N\to\infty} T_{\epsilon,w,N}^{\mathbf{y},\mathbf{b}}, \quad T_{\sigma,\sigma}^{\mathbf{y},\mathbf{y}} = \lim_{N\to\infty} T_{\sigma,\sigma,N}^{\mathbf{y},\mathbf{y}} \tag{30}$$

The first assumption is not restrictive, such selections always exist, and they can be found by an exhaustive search through all the possible selections. The assumption (30) says that the observed sample (y, u, q) satisfies the law of large numbers; ergodicity, which is a standard assumption in identification, implies that almost any sample satisfies (30).

Algorithm 3 Identification sLSS

Input: Data $\{y(t), u(t), q(t)\}_{t=1}^{N}$, (n_x, n_y, n_u+n_y)-selection (α, β) and $(\bar{n}, n_y, n_u)$-selection $(\bar{\alpha}, \bar{\beta})$; integer $\mathcal{I} > 0$.

1. Run Algorithm 2 with the covariances $\Lambda_w^{\mathbf{y},\mathbf{b}}$, $T_{\sigma,\sigma}^{\mathbf{y},\mathbf{y}}$ being replaced by the empirical covariances $T_{\epsilon,w,N}^{\mathbf{y},\mathbf{b}}$, $T_{\sigma,\sigma,N}^{\mathbf{y},\mathbf{y}}$ respectively for $\mathbf{b} \in \{\mathbf{y}, \mathbf{u}\}$ and $w \in \mathcal{L}_{\alpha,\beta} \cup \mathcal{L}_{\bar{\alpha},\bar{\beta}}$, $\sigma \in \Sigma$.

Output: The matrices $\{\tilde{A}_\sigma^N, \tilde{B}_\sigma^N, \tilde{K}_\sigma^{N,\mathcal{I}}, \tilde{Q}_\sigma^{N,\mathcal{I}}\}_{\sigma=1}^{n_\mu}, \tilde{C}^N, \tilde{D}^N$ returned by Algorithm 2.

Lemma 5 (Consistency): Under Assumption 2 the matrices returned by Algorithm 3 satisfy the following:

$$\tilde{K}_\sigma = \lim_{\mathcal{I}\to\infty} \lim_{N\to\infty} \tilde{K}_\sigma^{N,\mathcal{I}}, \quad p_\sigma T_{\sigma,\sigma}^{\mathbf{e}^s,\mathbf{e}^s} = \lim_{\mathcal{I}\to\infty} \lim_{N\to\infty} \tilde{Q}_\sigma^{N,\mathcal{I}}$$
$$[\tilde{A}_\sigma, \tilde{B}_\sigma] = \lim_{N\to\infty} [\tilde{A}_\sigma^N, \tilde{B}_\sigma^N], \ [\tilde{C}, \tilde{D}] = \lim_{N\to\infty} [\tilde{C}^N, \tilde{D}^N]$$

and $\mathcal{S} = (\{\tilde{A}_\sigma, \tilde{B}_\sigma, \tilde{K}_\sigma,\}_{\sigma=1}^{n_\mu}, \tilde{C}, \tilde{D}, \hat{\mathbf{x}}, \mathbf{e}^s)$ is a minimal sLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$ in innovation form.

That is, Algorithm 3 is a statistically consistent. Moreover, if the true system is a minimal sLSS in innovation form, which, by Theorem 1 we can assume w.l.g, then the sLSS returned by Algorithm 3 will be isomorphic to the true system, as $N \to \infty$. In particular, in the limit, the identified system has *the same deterministic behavior* as the true system. That is, while the data used for identification had to be sampled from certain distributions, the identified model recreates the behavior of the true one for any switching signal and input.

Algorithm 3 can be improved as follows.

Remark 4 (Computing empirical covariances): Using (29) directly for computing empirical covariances $T_{\epsilon,w,N}^{\mathbf{y},\mathbf{b}}$, $\mathbf{b} \in \{\mathbf{y}, \mathbf{u}\}$ results in slow convergence. Instead, we propose to compute the empirical covariances by solving a linear regression problem. More precisely, assume that $\mathcal{L}_{\alpha,\beta} \cup \mathcal{L}_{\bar{\alpha},\bar{\beta}} = \{w_1, \ldots, w_k\}$ and $\mathcal{L}_{\alpha,\beta} = \{w_1, \ldots, w_r\}$ for some $k, r > 0$. Let us then define the following matrices

$$R = \begin{bmatrix} y(N_0) \\ \vdots \\ y(N) \end{bmatrix}, \quad \Phi_{\mathbf{b}} = \begin{bmatrix} \mathbf{z}_{w_1}^{y}(N_0) & \cdots & \mathbf{z}_{w_{l(\mathbf{b})}}^{\mathbf{b}}(N_0) \\ \vdots & & \vdots \\ \mathbf{z}_{w_1}^{\mathbf{b}}(N) & \cdots & \mathbf{z}_{w_{l(\mathbf{b})}}^{\mathbf{b}}(N) \end{bmatrix}$$

where for $\mathbf{b} = \mathbf{y}$, $l(\mathbf{b}) = r$ and $b = y$, and for $\mathbf{b} = \mathbf{u}$, $l(\mathbf{b}) = k$ and $b = u$. By the well-known formula, $\hat{\Theta}^{\mathbf{b}} = (\Phi_{\mathbf{b}}^T \Phi_{\mathbf{b}})^{-1} \Phi_{\mathbf{b}}^T R$ is the least squares solution of the equation $R = \Phi_{\mathbf{b}} \hat{\Theta}^{\mathbf{b}}$. Hence, $\frac{1}{N-N_0}(\Phi_{\mathbf{b}}^T \Phi_{\mathbf{b}})\hat{\Theta}^{\mathbf{b}} = \frac{1}{N-N_0}\Phi_{\mathbf{b}}^T R$, and the latter contains all covariances $T_{\epsilon,w,N}^{\mathbf{y},\mathbf{b}}$, $\mathbf{b} \in \{\mathbf{u}, \mathbf{y}\}$ required for Algorithm 3. In turn, $\hat{\Theta}^{\mathbf{b}}$ can be computed using standard linear regression tools.

Remark 5 (Refinement using gradient descend): In order to enhance the performance of Algorithm 3, inspired by [9], we propose to use the sLSS returned by Algorithm 3 as initial value for the Gradient-Based (GB) algorithm of [9, Algorithm 7.1]. Note that [9] uses zero matrices as initial values for the noise gain matrices, whereas Algorithm 3 returns a non-zero estimate of those matrices.

VII. Numerical Example

In this section, we illustrate Algorithm 3 on a numerical example. All computations are carried out on an i7 2.11-GHz Intel core processor with 32GB of RAM running MATLAB R2023b. For data generation we use the following randomly generatated sLSS (1) with $\Sigma = \{1, 2\}$

$$A_1 = \begin{bmatrix} 0.1039 & 0.0255 & 0.5598 \\ 0.4338 & 0.0067 & 0.0078 \\ 0.3435 & 0.0412 & 0.0776 \end{bmatrix}, B_1 = \begin{bmatrix} 1.6143 \\ 5.9383 \\ 7.3671 \end{bmatrix},$$
$$A_2 = \begin{bmatrix} 0.1834 & 0.2456 & 0.0511 \\ 0.0572 & 0.2445 & 0.0642 \\ 0.1395 & 0.6413 & 0.5598 \end{bmatrix}, B_2 = \begin{bmatrix} 6.0624 \\ 4.9800 \\ 3.1372 \end{bmatrix},$$
$$K_1 = [0.4942\ 0.2827\ 0.8098]^T,$$
$$K_2 = [0.6215\ 0.1561\ 0.7780]^T$$
$$C = \begin{bmatrix} 0.1144 & 0.7623 & 0.0020 \end{bmatrix}, D = 1, F = 1.$$

Training and validation data of length N and N_{val}, respectively, are generated using the data generator sLSS with white-noise input process $\mathbf{u}$ with uniform distribution $\mathcal{U}(-1, 1)$, an i.i.d. process $\boldsymbol{\theta}$ sucht that $\mathcal{P}(\boldsymbol{\theta}(t) = q) = 0.5$, $q = 1, 2$, and a Gaussian white noise $\mathbf{v}$ with variance $\sigma_{\mathbf{v}}^2$. The corresponding *Signal-to-Noise Ratio* (SNR) are shown in Table I and calculated as follows: $\text{SNR} = 20\log_{10}(\frac{||\mathbf{y}||_2}{||\mathbf{v}||_2})$.

We run the version of Algorithm 3 using the Least-Square method from Remark 4, with $(\alpha, \beta) = (\bar{\alpha}, \bar{\beta})$,

$$\alpha=\{(11,1),(1,1),(\epsilon,1)\}, \quad \beta=\{(2,\epsilon,1),(1,2,1),(1,1,1)\},$$

and the GB refinement step from Remark 5. For the latter we used the LPVcore toolbox [11]. The basis are chosen so that it satisfied **(1)** of Assumption 2.

For validation, the true output minus the measurement noise $Fv(t)$, denoted by $y(t)$, was compared with the output $\hat{y}(t)$ predicted by the estimated model in innovation form using past inputs and outputs, as explained in Section IV-C. The quality of the match is evaluated via *Best Fit Ratio* criterion $\mathrm{BFR} = \max\left\{1 - \sqrt{\frac{\sum_{t=1}^{N_{\mathrm{val}}}(y(t)-\hat{y}(t))^2}{\sum_{t=1}^{N_{\mathrm{val}}}(y(t)-\bar{y})^2}}, 0\right\} \times 100\%$, where $\bar{y}$ denotes the sample mean of the output in the validation set.

TABLE I

BFR ON A NOISE-FREE VALIDATION DATA ($N_{\mathrm{val}} = 500$)

Method	$N = 5000$		$N = 10000$	
	SNR $= 6.1$ dB, $\sigma_{\mathbf{v}} = 1.5$			
	BFR[%]	time[s]	BFR[%]	time[s]
GB + zero noise gain	83.13	197	82.42	361
Algorithm 3	85.67	2.60	90.86	2.03
Algorithm 3 + GB	90.5	256	91.71	323
	SNR $= 0.5$ dB, $\sigma_{\mathbf{v}} = 3$			
GB + zero noise gain	63.25	188	66.11	316
Algorithm 3	76.97	1.94	87.29	1.90
Algorithm 3 + GB	84.17	236	88.03	376

The results are reported in Table I. We compare Algorithm 3 with using only the GB search algorithm from [9] where the initial value of the noise gain is zero, while the other matrices are initialized by the outcome of Algorithm 3. This is then equivalent with combining the corelation analysis (CRA) algorithm of [8] with GB. Algorithm 3 appears to perform better than using only GB search, especially for noisy data and a larger number of data points. Moreover, Algorithm 3 is faster than GB search. If combined with GB search, the performance of Algorithm 3 improves further, but at the expense of computational time.

VIII. Conclusion

In this paper, we presented a characterization of minimality and uniqueness of LSSs in innovation form, and we proposed a realization algorithm for computing minimal LSS in innovation form. Using this realization algorithm, we formulated a system identification algorithm proven to be statistically consistent. The latter algorithm was evaluated on a numerical example and demonstrated promising performance, both in terms of runtime complexity and estimation accuracy.

References

[1] A. R. P. Andriën and D. J. Antunes. Near-optimal map estimation for markov jump linear systems using relaxed dynamic programming. *IEEE L-CSS*, 4(4):815–820, 2020.

[2] L. Bako, G. Mercère, and S. Lecoeuche. On-line structured subspace identification with application to switched linear systems. *International Journal of Control*, 82(8):1496–1515, 2009.

[3] M. P. Balenzuela, A. G. Wills, C. Renton, and B. Ninness. Parameter estimation for jump markov linear systems. *Automatica*, 135:109949, 2022.

[4] P. Bilingsley. *Probability and measure*. Wiley, 1986.

[5] A. Chiuso and G. Picci. Consistency analysis of some closed-loop subspace identification methods. *Automatica*, 41(3):377–391, 2005.

[6] E. Cinquemani, R. Porreca, G. Ferrari-Trecate, and J. Lygeros. A general framework for the identification of jump markov linear systems. In *IEEE CDC*, 2007.

[7] O.L.V. Costa, M.D. Fragoso, and R.P. Marques. *Discrete-Time Markov Jump Linear Systems*. Springer Verlag, 2005.

[8] P. Cox, M. Petreczky, and R. Tóth. Towards efficient maximum likelihood estimation of LPV-SS models. *Automatica*, 97(9):392–403, 2018.

[9] P. B. Cox. *Towards efficient identification of linear parameter-varying state-space models*. PhD thesis, Eindhoven Univ. of Technology, 2018.

[10] P. B. Cox and R. Tóth. Linear parameter-varying subspace identification: A unified framework. *Automatica*, 123:109296, 2021.

[11] P. den Boef, P. B. Cox, and R. Tóth. LPVcore: MATLAB toolbox for LPV modelling, identification and control. 2021. IFAC SYSID.

[12] A. Garulli, S. Paoletti, and A. Vicino. A survey on switched and piecewise affine system identification. In *IFAC SYSID*, 2012.

[13] E. J. Hannan and M. Deistler. *The Statistical Theory of Linear Systems*. Classics in Applied Mathematics. SIAM, 1988.

[14] B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems*, volume 63,64 of *CWI Tract*. 1989.

[15] T. Katayama. *Subspace Methods for System Identification*. Springer-Verlag, 2005.

[16] F. Lauer and G. Bloch. *Hybrid System Identification: Theory and Algorithms for Learning Switching Models*, volume 478 of *LNCIS*. Springer, 2019.

[17] A. Lindquist and G. Picci. *Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification*. Springer, 2015.

[18] L. Ljung. *System Identification: Theory for the user (2nd Ed.)*. PTR Prentice Hall., Upper Saddle River, USA, 1999.

[19] M. Mejari and M. Petreczky. Consistent and computationally efficient estimation for stochastic lpv state-space models: realization based approach. In *IEEE CDC*, 2019.

[20] M. Mejari and M. Petreczky. Realization and identification algorithm for stochastic LPV state-space models with exogenous inputs. In *IFAC LPVS*, Eindhoven, Netherlands, 2019.

[21] M. Petreczky. Realization theory of linear hybrid systems. In *Hybrid Dynamical Systems: Observation and control*, Lecture Notes in Control and Information Sciences. Springer-Verlag, 2015.

[22] M. Petreczky and L. Bako. On the notion of persistence of excitation for linear switched systems. *Nonlinear Analysis: Hybrid Systems*, 48:101308, 2023.

[23] M. Petreczky, L. Bako, and J.H. van Schuppen. Realization theory for discrete-time linear switched systems. *Automatica*, 49(11):3337–3344, 2013.

[24] M. Petreczky and R. Vidal. Realization theory of stochastic jump-markov linear systems. In *IEEE CDC*, 2007.

[25] M. Petreczky and R. Vidal. Realization theory for a class of stochastic bilinear systems. *IEEE TAC*, 63(1):69–84, 2018.

[26] E. Rouphael, M. Mejari, M. Petreczky, and L. Belkoura. Minimal covariance realization and system identification algorithm for a class of stochastic linear switched systems with i.i.d. switching, 2024. ArXiv 2403.14259.

[27] E. Rouphael, M. Mejari, M. Petreczky, and L. Belkoura. Towards stochastic realization theory for generalized linear switched systems with inputs: decomposition into stochastic and deterministic components and existence and uniqueness of innovation form. *Submitted to CSS Control Systems Letters*, 2024. 2403.11012.

[28] E. Rouphael, M. Petreczky, and L.Belkoura. On minimal LPV state-space representations in innovation form: an algebraic characterization. In *IEEE CDC*, 2022.

[29] T. Sarkar, A. Rakhlin, and M. Dahleh. Nonparametric system identification of stochastic switched linear systems. In *IEEE CDC*, 2019.

[30] Zh. Sun and Sh. S. Ge. *Switched linear systems : control and design*. Springer, London, 2005.

[31] R. Tóth. Modeling and identification of linear parameter-varying systems. In *LNCIS, Vol. 403*, Heidelberg, 2010. Springer.

[32] P. van Overschee and B. De Moor. *Subspace Identification for Linear Systems*. Kluwer Academic Publishers, 1996.

[33] V. Verdult and M. Verhaegen. Subspace identification of multivariable linear parameter-varying systems. *Automatica*, 38(5):805–814, 2002.

[34] V. Verdult and M. Verhaegen. Subspace identification of piecewise linear systems. In *43rd IEEE CDC*, 2004.