Towards stochastic realization theory for Generalized Linear Switched Systems with inputs: decomposition into stochastic and deterministic components and existence and uniqueness of innovation form

Elie Rouphael, Manas Mejari, Mihaly Petreczky, Lotfi Belkoura

Abstract— **In this paper, we study a class of stochastic Generalized Linear Switched System (GLSS), which includes subclasses of jump-Markov, piecewide-linear and Linear Parameter-Varying (LPV) systems. We prove that the output of such systems can be decomposed into deterministic and stochastic components. Using this decomposition, we show existence of state-space representation in innovation form, and we provide sufficient conditions for such representations to be minimal and unique up to isomorphism.**

I. Introduction

A discrete-time stochastic *Generalized Linear Switched System (abbreviated as GLSS)* is a system of the form

$$\mathbf{S}\begin{cases}\mathbf{x}(t+1)=\sum_{i=1}^{n_\mu}(A_i\mathbf{x}(t)+B_i\mathbf{u}(t)+K_i\mathbf{v}(t))\boldsymbol{\mu}_i(t)\\ \mathbf{y}(t)=C\mathbf{x}(t)+D\mathbf{u}(t)+F\mathbf{v}(t),\quad t\in\mathbb{Z}\end{cases}\tag{1}$$

where $A_i \in \mathbb{R}^{n_x\times n_x}$, $B_i \in \mathbb{R}^{n_x\times n_u}$, $K_i \in \mathbb{R}^{n_x\times n_n}$, $i = 1,\dots,n_\mu$, $C \in \mathbb{R}^{n_y\times n_x}$ and $D \in \mathbb{R}^{n_y\times n_u}$, $F \in \mathbb{R}^{n_y\times n_n}$ are constant matrices. The stochastic process $\mathbf{x},\mathbf{u},\mathbf{y},\mathbf{v}$ and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_{n_\mu})^T$ are the state, input, output, noise and switching processes respectively, taking values respectively in $\mathbb{R}^{n_x},\mathbb{R}^{n_u},\mathbb{R}^{n_y},\mathbb{R}^{n_n},\mathbb{R}^{n_\mu}$ and defined on $\mathbb{Z}$.

Intuitively, (1) is a generalization of linear switched systems to the case of infinitely many discrete modes. If $\boldsymbol{\mu}(t)$ takes values in the set of unit vectors, i.e., $\boldsymbol{\mu}_i(t) \in \{0,1\}$, $i = 1,\dots,n_\mu$, $\sum_{i=1}^{n_\mu}\boldsymbol{\mu}_i(t) = 1$, then (1) is a *switched system [13]*. If, in addition, the process $\theta(t)$, defined by $\theta(t) = i \iff \boldsymbol{\mu}_i(t) = 1$, is a Markov chain, then (1) is a jump-Markov system [3]. If $\boldsymbol{\mu}$ takes values from a possibly infinite set and $\boldsymbol{\mu}_1 = 1$, then (1) could be viewed as a subclass of *linear parameter varying (LPV)* systems [15] with an *affine dependence* on scheduling, and $(\boldsymbol{\mu}_2,\dots,\boldsymbol{\mu}_{n_\mu})^T$ corresponds to the scheduling signal. However, in contrast to LPV systems, in general we are agnostic about the role of $\boldsymbol{\mu}$ in control, hence the use of the term GLSS.

Context and motivation: Consider the following deterministic counterpart of (1)

$$\begin{aligned}x(t+1)&=\sum_{i=1}^{n_\mu}(A_ix(t)+B_iu(t)+K_iv(t))\mu_i(t)\\ y(t)&=Cx(t)+Du(t)+Fv(t),t\in\mathbb{Z}\end{aligned}\tag{2}$$

where all the signals are deterministic and which describes the response of the true system for **any** input, switching and noise realization, and not only for the samples from $\mathbf{u},\boldsymbol{\mu},\mathbf{v}$. For the tuple of matrices $S = (\{A_i,B_i,K_i\}_{i=0}^{n_\mu},C,D,F)$ of (1) define the *deterministic behavior* $\mathcal{B}_S$ of S as the as the set of all tuples of deterministic trajectories (u,μ,y) such that there exists trajectories x and v for which (2) holds. Clearly, all samples paths of $(\mathbf{u},\boldsymbol{\mu},\mathbf{v})$ belong to $\mathcal{B}_S$.

In system identification of switched and LPV systems, we want to find matrices $\hat{S} = (\{\hat{A}_i,\hat{B}_i,\hat{K}_i\}_{i=0}^{n_\mu},\hat{C},\hat{D},\hat{F})$ from one single tuple of trajectories (u,μ,y) from $\mathcal{B}_S$, such that the deterministic behavior $\mathcal{B}_{\hat{S}}$ is an approximation of $\mathcal{B}_S$. For jump-Markov system, the task is similar, but in the definition of the deterministic behavior the switching and noise trajectories are sampled from processes $\boldsymbol{\mu}$ and $\mathbf{v}$.

As a tool to achieve this goal, we assume that the data (u,μ,y) used for identification, are sampled from the processes $(\mathbf{u},\boldsymbol{\mu},\mathbf{y})$, where $\mathbf{y}$ is generated by the data generator (1). Pragmatically, we cannot exclude random effects during the identification experiment (measurement error, etc.), and randomness can be used for statistical reasoning about algorithms. For the latter, we view $\hat{S}$ as a statistics for the matrices of (1), which, in turn, parameterize the joint distribution of $(\mathbf{u},\boldsymbol{\mu},\mathbf{y})$. However, the estimated models should approximate the true system for all inputs and scheduling signals, and not only those which are sampled from $\mathbf{u}$ and $\boldsymbol{\mu}$. Unfortunately, good statistical properties of $\hat{S}$, e.g., consistency, guarantee only that the output of the stochastic system determined by $\hat{S}$ is *close* to that of the stochastic data generator (1), for the specific stochastic input $\mathbf{u}$ and switching $\boldsymbol{\mu}$ and a choice of the noise process. However, this does not generally imply that the deterministic behaviors $\mathcal{B}_{\hat{S}}$ and $\mathcal{B}_S$ of $\hat{S}$ and S are close. In fact, the outputs of two GLSS may be the same for the input $\mathbf{u}$ and switching $\boldsymbol{\mu}$ used during identification, but be different for others [12].

In the LTI case, this was resolved by assuming that the data generator (1) and the stochastic system corresponding to $\hat{S}$ are minimal and in *innovation form*. Since the matrices of minimal systems in innovation form with the same outputs and inputs are related by a basis transformation [7], if the stochastic system corresponding to $\hat{S}$ is close to the data generator, then intuitively the matrices $\hat{S}$ and S are close after a suitable basis transformation. Hence, the deterministic behaviors of $\hat{S}$ and S are close. Moreover, innovation form was useful for developing system identification algorithms (e.g., subspace, prediction error minimization), and estab-

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

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

lishing a correspondence between state-space representations and optimal predictors of $\mathbf{y}$ based on its past of $\mathbf{u}$ and $\mathbf{y}$.

A key step in the proof of existence and uniqueness of minimal LTI systems in innovation form is the decomposition $\mathbf{y}'t) = \mathbf{y}^d(t) + \mathbf{y}^s(t)$ of the output into two components. The component $\mathbf{y}^d$ is determined by the input and it is the output of a noiseless LTI system. The component $\mathbf{y}^s$ is the output of an autonomous LTI system and it depends only on the noise. Then existence and uniqueness of minimal LTI systems in innovation form follow by applying realization theory [7] to the autonomous LTI system generating $\mathbf{y}^s$.

Contribution: In this paper we show an analogous result for GLSSs, i.e. we show that $\mathbf{y}(t) = \mathbf{y}^d(t) + \mathbf{y}^s(t)$, where $\mathbf{y}^d$ is the output of a GLSS with no noise $\mathbf{v}$, and $\mathbf{y}^s$ is the output of a GLSS with no input $\mathbf{u}$. Furthermore, by using results on realization theory of GLSSs with no inputs [10], we use this decomposition to show existence of an innovation form for GLSSs with inputs. Moreover, we present sufficient conditions for minimality and uniqueness (up to change of basis) of GLSSs in innovation form. This eans that minimal GLSSs in innovation form have the same useful properties for system identification as their LTI counterparts. In particular, if two minimal GLSSs in innovation form generate (approximately) the same output for the data used of identification, then they will generate (approximately) the same output for any input and switching signal, including those which do not satisfy the assumptions of the identification experiment.

Related work: System identification in general, and subspace methods in particular, of switched [6], jump-Markov [1] and LPV systems [15], [5], [17], [11], [18], [19], [14] is a well-established topic. Stochastic realization theory of GLSSs with no inputs, i.e., jump-Markov systems with no inputs, bilinear systems and autonomous stochastic LPV systems were addressed in [10]. With respect to [10] the main difference is the presence of the control input $\mathbf{u}$.

Existence of a decomposition and existence of innovation form appeared in [8], but only for the case of LPV systems with zero mean i.i.d. scheduling,. With respect to [8], the main novelty is that we allow more general switching processes, including finite Markov chains, and that we address minimality and uniqueness of innovation representations. Moreover, in [8] the proofs were not presented, they were included in the report [9], the latter can be viewed as a preliminary version of this paper.

The existence of innovation representation was studied for LPV systems in [4], [5]. In contrast to this paper, in [4], [5] the noise gain of the innovation representation has a dynamical dependence on the scheduling, and there is no claim on minimality and uniqueness of such representations. In particular, it is unclear when the innovation representation from [4], [5] generates the same output as the original model for all scheduling signals. However, [4], [5] has the advantage that it is valid for any scheduling signal.

II. PRELIMINARIES

Below we present the notation used in the paper. In addition, we recall a number of concepts from [10], and then we use them to define the subclass of stationary GLSS for which our main results hold.

Probability theory: We use the standard terminology of probability theory [2]. All the random variables and stochastic processes are understood w.r.t. to a fixed probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where $\mathcal{F}$ is a σ-algebra over the sample space Ω. The expected value of a random variable $\mathbf{r}$ is denoted by $E[\mathbf{x}]$ and conditional expectation w.r.t. σ-algebra $\mathcal{F}$ is denoted by $E[\mathbf{r} \mid \mathcal{F}]$. The stochastic processes in this paper are discrete-time ones defined over the time-axis $\mathbb{Z}$.

Finite sequences: In what follows $\Sigma = \{1, \ldots, n_\mu\}$. A *non empty word* over Σ is a finite sequence of letters (elements) of Σ, i.e., $w = \sigma_1\sigma_2\cdots\sigma_k$, for some $k \in \mathbb{N}$, $k > 0$, $\sigma_1, \sigma_2, \ldots, \sigma_k \in \Sigma$; $|w| := k$ is the length of w. The set of *all* nonempty words is denoted by Σ^+. We denote the *empty word* by ϵ and by convention $|\epsilon| = 0$. Let $\Sigma^* = \{\epsilon\} \cup \Sigma^+$. The concatenation of two nonempty words $v = a_1a_2\cdots a_m$ and $w = b_1b_2\cdots b_n$ is defined as $vw = a_1\cdots a_m b_1\cdots b_n$ for some $m, n > 0$. By convention $v\epsilon = \epsilon v = v$ for all $v \in \Sigma^*$.

Notation for matrices We denote by I_n the $n \times n$ identity matrix. Consider $n \times n$ square matrices $\{A_\sigma\}_{\sigma\in\Sigma}$. For any $w = \sigma_1\sigma_2\cdots\sigma_k \in \Sigma^+$, $k > 0$ and $\sigma_1, \ldots, \sigma_k \in \Sigma$, we define $A_w = A_{\sigma_k}A_{\sigma_{k-1}}\cdots A_{\sigma_1}$. For an empty word ϵ, set $A_\epsilon = I_n$.

Notions from [10]: admissible switching, ZMWSII, SII processes: We first formulate our assumptions for the switching process. 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:

$$\boldsymbol{\mu}_w(t) = \boldsymbol{\mu}_{\sigma_1}(t-k+1)\boldsymbol{\mu}_{\sigma_2}(t-k+2)\cdots\boldsymbol{\mu}_{\sigma_k}(t) \tag{3}$$

For an empty word $w = \epsilon$, we set $\boldsymbol{\mu}_\epsilon(t) = 1$.

Definition 1 (Admissible process, [10]): The switching process $\boldsymbol{\mu}$ is called *admissible*, if the following holds:

1. There exists a set $\mathcal{E} \subseteq \Sigma \times \Sigma$ such that:

- $\forall \sigma \in \Sigma, \exists \sigma' \in \Sigma : (\sigma, \sigma') \in \mathcal{E}$.
- Define the set of admissible words L as the set of all words $w \in \Sigma^+$ such that $w = \sigma_1\cdots\sigma_k \in \Sigma^+, \sigma_1, \ldots, \sigma_k \in \Sigma, k > 0$ and for all $i = 1, \ldots, k-1$, $(\sigma_i, \sigma_{i+1}) \in \mathcal{E}$. Then for all $w \in \Sigma^+, w \notin L, \boldsymbol{\mu}_w = 0$.

2. Denote by $\mathscr{F}_t^{\boldsymbol{\mu},-}$ the σ-algebra generated by the random variables $\{\boldsymbol{\mu}(k)\}_{k<t}$. There exists positive numbers $\{p_\sigma\}_{\sigma\in\Sigma}$ such that for any $w, v \in \Sigma^+, \sigma, \sigma' \in \Sigma$, $t \in \mathbb{Z}$:

$$E[\boldsymbol{\mu}_{w\sigma}(t)\boldsymbol{\mu}_{v\sigma'}(t) \mid \mathscr{F}_t^{\boldsymbol{\mu},-}] = \begin{cases} p_\sigma\boldsymbol{\mu}_w(t-1)\boldsymbol{\mu}_v(t-1) & \sigma = \sigma' \text{and } w\sigma, v\sigma \in L \\ 0 & \text{otherwise} \end{cases}$$

$$E[\boldsymbol{\mu}_{w\sigma}(t)\boldsymbol{\mu}_{\sigma'}(t) \mid \mathscr{F}_t^{\boldsymbol{\mu},-}] = \begin{cases} p_\sigma\boldsymbol{\mu}_w(t-1) & \sigma = \sigma' \text{ and } w\sigma \in L \\ 0 & \text{otherwise} \end{cases}$$

$$E[\boldsymbol{\mu}_\sigma(t)\boldsymbol{\mu}_{\sigma'}(t) \mid \mathscr{F}_t^{\boldsymbol{\mu},-}] = 0 \quad \text{if } \sigma \neq \sigma'$$

3. There exist real numbers $\{\alpha_\sigma\}_{\sigma\in\Sigma}$ such that $\sum_{\sigma\in\Sigma}\alpha_\sigma\boldsymbol{\mu}_\sigma(t) = 1$ for all $t \in \mathbb{Z}$.

4. For each $w, v \in \Sigma^+$, the process $\left[\boldsymbol{\mu}_w, \boldsymbol{\mu}_v\right]^T$ is wide-sense stationary.

Below we recall from [10] a number of examples of admissible processes.

Example 1 (White noise): If $\boldsymbol{\mu} = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_{n_\mu}]^T$ is an i.i.d. process such that $\boldsymbol{\mu}_1 = 1$, for all $i, j = 2, \ldots, n_\mu$, $t \in \mathbb{Z}$, $\boldsymbol{\mu}_i(t), \boldsymbol{\mu}_j(t)$ are independent and $\boldsymbol{\mu}_i(t)$ is zero mean, then $\boldsymbol{\mu}$ is admissible with $\mathcal{E} = \Sigma \times \Sigma$ and $p_\sigma = E[\boldsymbol{\mu}_\sigma^2(t)]$.

Example 2 (Discrete valued i.i.d process): Let $\boldsymbol{\theta}$ be an i.i.d process with values in $\Sigma = \{1, \ldots, n_\mu\}$. Let $\boldsymbol{\mu}_\sigma(t) = \chi(\boldsymbol{\theta}(t) = \sigma)$ for all $\sigma \in \Sigma$, $t \in \mathbb{Z}$, where χ is the indicator function. Let $\mathcal{E} = \Sigma \times \Sigma$, and $p_\sigma = P(\boldsymbol{\theta}(t) = \sigma)$, $\alpha_\sigma = 1$ for all $\sigma \in \Sigma$. Then $\boldsymbol{\mu}$ is admissible.

Example 3 (Markov chain): Assume that $\boldsymbol{\theta}$ is a stationary and ergodic Markov process whose state space is the finite set Θ. Assume $P(\boldsymbol{\theta}(t) = q_2 \mid \boldsymbol{\theta}(t-1) = q_1) = p_{(q_2,q_1)} > 0$, $q_1, q_2 \in \Theta$. Let us take $\Sigma = \Theta \times \Theta$, $\boldsymbol{\mu}_{(q_2,q_1)}(t) = \chi(\boldsymbol{\theta}(t+1) = q_2, \boldsymbol{\theta}(t) = q_1)$ for all $q_1, q_2 \in \Theta, t \in \mathbb{Z}$, and let $\mathcal{E} = \{(\sigma_1, \sigma_2) \in \Sigma \times \Sigma \mid \sigma_1 = (q_2, q_1), \sigma_2 = (q_3, q_2), q_1, q_2, q_3 \in \Theta\}$. Define $\alpha_\sigma = 1$ for all $\sigma \in \Sigma$. Let us identify Σ with the set $\{1, \ldots, n_\mu\}$, where n_μ is the square of cardinality of Θ. Then $\boldsymbol{\mu}$ is admissible.

Assumption 1: The switching process $\boldsymbol{\mu}$ is admissible.

This assumption imposes restrictions on the data used for system identification, but not necessarily for the model class which will be identified. It can be thought of as a persistence of excitation condition. In particular, binary and white noises, which are the simplest persistently exciting signals, satisfy our assumption. We believe that for developing realization theory for general switching signals, these simple cases must be understood first. Moreover, admissible switching signals cover the fairly general case of Markov chains.

Remark 1 (LPV systems): For LPV systems, our assumption implies that the scheduling signal used for identification is sampled from a stochastic process. In addition to this being a persistence of excitation condition, we argue that in certain cases this assumption is justified [11]: in the presence of measurement errors, or when the scheduling is externally generated, or it is a function of the stochastic states/inputs.

Next, we recall the concept of *ZMWSII process w.r.t* $\boldsymbol{\mu}$ from [10]. To this end, let $\{p_\sigma\}_{\sigma\in\Sigma}$ be the constants from Definition 1. For any $w = \sigma_1 \cdots \sigma_k \in \Sigma^+$, $\sigma_1, \ldots, \sigma_k \in \Sigma$, for a process $\mathbf{r}$, define the product p_w and the process $\mathbf{z}_w^{\mathbf{r}}$

$$\begin{aligned} p_w &= p_{\sigma_1} p_{\sigma_2} \cdots p_{\sigma_k}, \\ \mathbf{z}_w^{\mathbf{r}}(t) &= \mathbf{r}(t - |w|)\boldsymbol{\mu}_w(t-1)\frac{1}{\sqrt{p_w}}, \end{aligned} \tag{4}$$

where $\boldsymbol{\mu}_w$ is as in (3).The process $\mathbf{z}_w^{\mathbf{r}}$ in (4) is interpreted as the product of the *past* of $\mathbf{r}$ and $\boldsymbol{\mu}$.

Definition 2 (ZMWSSI, [10]): A process $\mathbf{r}$ is *Zero Mean Wide Sense Stationary w.r.t.* $\boldsymbol{\mu}$ (ZMWSSI) if

(1) For $t \in \mathbb{Z}$, the σ-algebras generated by the variables $\{\mathbf{r}(k)\}_{k\le t}$, $\{\boldsymbol{\mu}_\sigma(k)\}_{k<t,\sigma\in\Sigma}$ and $\{\boldsymbol{\mu}_\sigma(k)\}_{k\ge t,\sigma\in\Sigma}$, denoted by $\mathcal{F}_t^{\mathbf{r}}$, $\mathcal{F}_t^{\boldsymbol{\mu},-}$ and $\mathcal{F}_t^{\boldsymbol{\mu},+}$ respectively, are such that $\mathcal{F}_t^{\mathbf{r}}$ and $\mathcal{F}_t^{\boldsymbol{\mu},+}$ are conditionally independent w.r.t. $\mathcal{F}_t^{\boldsymbol{\mu},-}$.

(2) The processes $\{\mathbf{r}, \{\mathbf{z}_w^{\mathbf{r}}\}_{w\in\Sigma^+}\}$ are zero mean, square integrable and are jointly wide sense stationary.

Intuitively, ZMWSII is an extension of wide-sense stationarity, where Σ^+ is viewed as time axis: ZMWSII implies the covariances $E\left[\mathbf{z}_w^{\mathbf{r}}(t)(\mathbf{z}_v^{\mathbf{r}}(t))^T\right]$ do not depend on t, and they depend on the difference between v and w. Next, we recall the definition of a square integrable process w.r.t. $\boldsymbol{\mu}$.

Definition 3 (SII process, [10]): A process $\mathbf{r}$ is *Square Integrable w.r.t.* $\boldsymbol{\mu}$ *(SII)*, if for all $w \in \Sigma^*$, $t \in \mathbb{Z}$, the random variable $\mathbf{r}(t + |w|)\boldsymbol{\mu}_w(t + |w| - 1)$ is square integrable.

As it was mentioned in [10, Section III, Remark 2], if $\boldsymbol{\mu}$ is essentially bounded, then any ZMWSII process is SII.

Assumptions, inputs and outputs and on GLSSs: First, we define the notion of white noise processes w.r.t. $\boldsymbol{\mu}$, which will be used for stating our assumptions on GLSSs.

Definition 4 (White noise w.r.t. $\boldsymbol{\mu}$*):* A ZMWSII process $\mathbf{r}$ is a white noise w.r.t. $\boldsymbol{\mu}$, if for all $w \in \Sigma^+$, $v \in \Sigma^*$, $\sigma \in \Sigma$,

$$E[\mathbf{z}_w^{\mathbf{r}}(t)(\mathbf{z}_{\sigma v}^{\mathbf{r}}(t))^T] = \begin{cases} 0 & \text{if } w \neq \sigma v \\ E[\mathbf{z}_\sigma^{\mathbf{r}}(t)(\mathbf{z}_\sigma^{\mathbf{r}}(t))^T] & \text{if } w = \sigma v \end{cases},$$

and $E[\mathbf{z}_\sigma^{\mathbf{r}}(t)(\mathbf{z}_\sigma^{\mathbf{r}}(t))^T]$ is nonsingular for all $\sigma \in \Sigma$.

Intuitively, if $\mathbf{r}$ is a white noise process w.r.t. $\boldsymbol{\mu}$, then $\{\mathbf{z}_w^{\mathbf{r}}(t)\}_{w\in\Sigma^+}$ is a sequence of uncorrelated random variables. Due to **3.** of Definition 1, the product $\mathbf{r}(t-k)\mathbf{r}(t)$ is a linear combination of $\{\mathbf{z}_w^{\mathbf{r}}(t)\}_{w\in\Sigma^+}$, hence a white noise process w.r.t. $\boldsymbol{\mu}$ is also a white noise process in the classical sense. Conversely, if $\mathbf{r}$ is a white noise and it is independent of $\{\boldsymbol{\mu}(s)\}_{s\in\mathbb{Z}}$, then it is a white noise process w.r.t. $\boldsymbol{\mu}$.

Assumption 2 (Inputs and outputs): **(1)** $\mathbf{u}$ is a white noise w.r.t. $\boldsymbol{\mu}$, **(2)** the process $\left[\mathbf{y}^T, \mathbf{u}^T\right]^T$ is a ZMWSSI.

The assumption that $\mathbf{u}$ is white noise was made for the sake of simplicity, we conjecture that the results of the paper can be extended to more general inputs, e.g., inputs generated by autoregressive models driven by white noise. Next, we define the class of systems considered in this paper.

Definition 5 (Stationary GLSS): A *stationary GLSS* (abbreviated sGLSS) of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ is a system (1), such that

1. $\mathbf{w} = \left[\mathbf{v}^T, \mathbf{u}^T\right]^T$ is a white noise process w.r.t. $\boldsymbol{\mu}$.
2. The process $\left[\mathbf{x}^T, \mathbf{w}^T\right]^T$ is a ZMWSSI, and $E[\mathbf{z}_\sigma^{\mathbf{x}}(t)(\mathbf{z}_\sigma^{\mathbf{w}}(t))^T] = 0$, $E[\mathbf{x}(t)(\mathbf{z}_w^{\mathbf{w}}(t))^T] = 0$, 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.
4. For all $\sigma_1, \sigma_2 \in \Sigma$, if $(\sigma_1, \sigma_2) \notin \mathcal{E}$, then $A_{\sigma_2} A_{\sigma_1} = 0$ and $A_{\sigma_2}\begin{bmatrix} B_{\sigma_1} & K_{\sigma_1}\end{bmatrix} E[\mathbf{z}_{\sigma_1}^{\mathbf{w}}(t)(\mathbf{z}_{\sigma_1}^{\mathbf{w}}(t))^T] = 0$.

If $B_i = 0$, $i \in \Sigma$, and $D = 0$ the we call (1) an *autonomous stationary* GLSS (asGLSS) of $(\mathbf{y}, \boldsymbol{\mu})$.

Intuitively, sGLSSs are introduced in order to ensure that all the relevant stochastic processes are stationary in an suitable sense. The latter assumption is widespread in stochastic realization theory and system identification.

In the terminology of [10], a sGLSS (resp. asGLSS) corresponds to a stationary *Generalized Bilinear System* (**GBS**) with noise $\begin{bmatrix}\mathbf{v}^T & \mathbf{u}^T\end{bmatrix}^T$ (resp. $\mathbf{v}$). From [10] it follows that the state and output process $\mathbf{x}$ and $\mathbf{y}$ are ZMWSII, and hence Assumption 2 is satisfied for all $(\mathbf{u}, \boldsymbol{\mu}, \mathbf{y})$ generated by sGLSS. Moreover, from [10, Lemma 2] it follows that

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

where the infinite sum converges in the mean square sense. Hence, the state $\mathbf{x}$ is uniquely determined by the system matrices and the input $\mathbf{u}$ and noise $\mathbf{v}$, and it is the limit of any state trajectory started from some initial state.

Notation 1: We identify the sGLSS $\mathbf{S}$ of the form (1) with the tuple $\mathbf{S} = (\{A_\sigma, K_\sigma, B_\sigma\}_{\sigma=1}^{n_\mu}, C, D, F, \mathbf{v})$, and if $\mathbf{S}$ is a asGLSS, i.e. $B_\sigma = 0$, $\sigma \in \Sigma$, $D = 0$, then we will identify it with the tuple $\mathbf{S} = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, F, \mathbf{v})$.

III. Main result

We start by recalling from [10] the following terminology.

Notation 2 (Orthogonal projection E_l): Recall from [2] that the set of real valued square integrable random variables, denoted by $\mathcal{H}_1$, forms a Hilbert-space with the scalar product defined as $\langle \mathbf{z}_1, \mathbf{z}_2 \rangle = E[\mathbf{z}_1 \mathbf{z}_2]$. Let $\mathbf{z}$ be a square integrable random variable taking its values in $\mathbb{R}^k$. Let M be a closed subspace of $\mathcal{H}_1$. The orthogonal projection of $\mathbf{z}$ onto M, *denoted by* $E_l[\mathbf{z} \mid M]$, is defined as the random variable $\mathbf{z}^* = \begin{bmatrix}\mathbf{z}_1^*, \ldots, \mathbf{z}_k^*\end{bmatrix}^T$ such that $\mathbf{z}_i^* \in M$ is the orthogonal projection of the ith coordinate $\mathbf{z}_i$ of $\mathbf{z}$, viewed as an element of $\mathcal{H}_1$ onto M. If $\mathfrak{S}$ is a subset of square integrable random variables in $\mathbb{R}^p$, and M is generated by the coordinates of the elements of $\mathfrak{S}$, then instead of $E_l[z \mid M]$ we use $E_l[\mathbf{z} \mid \mathfrak{S}]$. Intuitively, $E_l[\mathbf{z} \mid \mathfrak{S}]$ is *the best (minimal variance) linear prediction* of $\mathbf{z}$ based on the elements of $\mathfrak{S}$.

Using the notation above, let us define the *deterministic component* $\mathbf{y}^d$ of $\mathbf{y}$ as follows

$$\mathbf{y}^d(t) = E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w^{\mathbf{u}}(t)\}_{w \in \Sigma^+} \cup \{\mathbf{u}(t)\}]. \tag{5}$$

Also, define the *stochastic component* of $\mathbf{y}$ as

$$\mathbf{y}^s(t) = \mathbf{y}(t) - \mathbf{y}^d(t). \tag{6}$$

Intuitively, $\mathbf{y}^d$ represent the best prediction of $\mathbf{y}$ which is linear in the present and past values of $\mathbf{u}$ and non-linear in the past values of $\boldsymbol{\mu}$. In fact, $\mathbf{y}^d$ is the output of the asGLSS obtained from (1) by considering $\mathbf{v} = 0$ and viewing $\mathbf{u}$ as noise, and $\mathbf{y}^s$ is the output of the asGLSS obtained from (1) by taking $\mathbf{u} = 0$ and viewing $\mathbf{v}$ as noise.

Theorem 1: For a sGLSS of the form (1), $\mathbf{S}_d = (\{A_\sigma, B_\sigma\}_{\sigma=1}^{n_\mu}, C, D, \mathbf{u})$ is an asGLSS of $(\mathbf{y}^d, \boldsymbol{\mu})$ and $\mathbf{S}_s = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, F, \mathbf{v})$ is an asGLSS of $(\mathbf{y}^s, \boldsymbol{\mu})$.

The proof of Theorem 1 is presented in the Appendix V-A.

In fact, the converse of Theorem 1 also holds. To this end, recall from [10] the definition of *innovation process* of $\mathbf{y}^s$:

$$\mathbf{e}^s(t) = \mathbf{y}^s(t) - E_l[\mathbf{y}^s(t) \mid \{\mathbf{z}_w^{\mathbf{y}^s}(t)\}_{w \in \Sigma^+}] \tag{7}$$

Intuitively, $\mathbf{e}^s(t)$ is the difference between $\mathbf{y}(t)$ and the best linear prediction of $\mathbf{y}^s(t)$ based on its own past multiplied with past values of the switching process.

Theorem 2: Assume that there exists a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ and that the following holds:

1. $\hat{\mathbf{S}}_d = (\{\hat{A}_i, \hat{B}_i\}_{i=1}^{n_\mu}, \hat{C}, \hat{D}, \mathbf{u})$ is an asGLSS of $(\mathbf{y}^d, \boldsymbol{\mu})$.
2. $\hat{\mathbf{S}}_s = (\{\hat{A}_i, \hat{K}_i\}_{i=1}^{n_\mu}, \hat{C}, I_{n_y}, \mathbf{v})$ is an asGLSS of $(\mathbf{y}^s, \boldsymbol{\mu})$ in *innovation form*, i.e. $\mathbf{v} = \mathbf{e}^s$.

Then $\hat{\mathbf{S}} = (\{\hat{A}_i, \hat{K}_i, \hat{B}_i\}_{i=1}^{n_\mu}, \hat{C}, \hat{D}, I, \mathbf{e}^s)$ is a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$, and the innovation process $\mathbf{e}^s$ satisfies $\mathbf{e}^s(t) = \mathbf{y}(t) - \hat{\mathbf{y}}(t)$, where

$$\hat{\mathbf{y}}(t) = E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w^{\mathbf{y}}(t), \mathbf{z}_w^{\mathbf{u}}(t)\}_{w \in \Sigma^+} \cup \{\mathbf{u}(t)\}]. \tag{8}$$

The proof of Theorem 2 is presented in Section V-B.

Remark 2: The condition that the $\{A_i\}_{i=1}^{n_\mu}$ and C matrices of $\hat{\mathbf{S}}_d$ of $\hat{\mathbf{S}}_d$ can be relaxed: if $\bar{\mathbf{S}}_d = (\{\hat{A}_i^d, \hat{B}_i^d\}_{i=1}^{n_\mu}, \hat{C}^d, \hat{D}, \mathbf{u})$ and $\bar{\mathbf{S}}_s = (\{\hat{A}_i^s, \hat{B}_i^s\}_{i=1}^{n_\mu}, \hat{C}^s, I, \mathbf{e}^s)$ are asGLSS of $(\mathbf{y}^s, \boldsymbol{\mu})$ and $(\mathbf{y}^s, \boldsymbol{\mu})$ respectively, then with

$$\hat{A}_i = \begin{bmatrix} \hat{A}_i^d & 0 \\ 0 & \hat{A}_i^s \end{bmatrix}, \ \hat{B}_i = \begin{bmatrix} \hat{B}_i^d \\ 0 \end{bmatrix}, \ \hat{K}_i = \begin{bmatrix} 0 \\ \hat{K}_i^s \end{bmatrix}, \ \hat{C} = \begin{bmatrix} (C^d)^T \\ (C^s)^T \end{bmatrix}^T,$$

$\hat{\mathbf{S}}_d$ and $\hat{\mathbf{S}}_s$ from Theorem 2 are asGLSSs of $(\mathbf{y}^d, \boldsymbol{\mu})$ and $(\mathbf{y}^s, \boldsymbol{\mu})$ respectively and Theorem 2 applies.

Thus, Theorem 1 – 2 means that finding sGLSSs of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ is equivalent to finding an asGLSS representations of the deterministic and stochastic components respectively.

Theorem 2 suggests that $\mathbf{e}^s(t)$ can be viewed as the innovation process of $\mathbf{y}$. Indeed, $\hat{\mathbf{y}}(t)$ from (8) is the best linear prediction of $\mathbf{y}(t)$ based on past values of $\mathbf{y}$ and past and current values of $\mathbf{u}$ multiplied by past values of the switching process. Then $\mathbf{e}^s(t)$ is the prediction error $\mathbf{y}(t) - \hat{\mathbf{y}}(t)$. This motivates the following definition.

Definition 6 (Innovation form): The sGLSS (1) is in *innovation form*, if F is the identity matrix and $\mathbf{v} = \mathbf{e}^s$.

Similarly to the LTI case [7], an sGLSS in innovation form can be viewed as a recursive filter driven by $\mathbf{y}$, $\mathbf{u}$, $\boldsymbol{\mu}$, whose output is the optimal prediction $\hat{\mathbf{y}}(t)$ from (8). Indeed, from $\mathbf{e}^s(t) = \mathbf{y}(t) - C\mathbf{x}(t) - D\mathbf{u}(t)$ it follows that $\mathbf{x}(t+1)$ is a function of $\mathbf{x}(t), \mathbf{u}(t), \mathbf{y}(t)$ and $\boldsymbol{\mu}(t)$, and $\hat{\mathbf{y}}(t) = C\mathbf{x}(t) + D\mathbf{u}(t)$

Corollary 1 (Existence): Any sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ can be transformed to a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form.

Proof: From Theorem 1 it follows that $\mathbf{S}_s$ is an asGLSS of $(\mathbf{y}^s, \boldsymbol{\mu})$ and $\mathbf{S}_d$ is an asGLSS of $(\mathbf{y}^d, \boldsymbol{\mu})$. From [10, Theorem 2] it follows that there exists a (minimal) asGLSS $\bar{\mathbf{S}}_s$ of $(\mathbf{y}^s, \boldsymbol{\mu})$ in innovation form and by [10, Theorem 3] it can be computed from $\mathbf{S}_s$ using [10, Algorithm 1]. Then using Remark 2 and Theorem 2, it follows that $\hat{\mathbf{S}}$ defined in Theorem 2 is a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form. ■

We can also provide conditions for minimality of sGLSSs. To this end, for a sGLSS of the form (1) we refer to the dimension n_x of the state-space as *dimension* of sGLSS.

Corollary 2 (Minimality): Assume that $\mathbf{S}$ is a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$, and assume that $\mathbf{S}_s$ from Theorem 1 is observable and reachable in the terminology of [10], if viewed as a stationary **GBS**. Then it is minimal dimensional among all the sGLSSs of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$.

Proof: Let $\hat{\mathbf{S}}$ be a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ of smaller dimension than $\mathbf{S}$. Then by Theorem 1, $\hat{\mathbf{S}}_s$ is an asGLSS of $(\mathbf{y}, \boldsymbol{\mu})$ of the same dimension as $\hat{\mathbf{S}}$. However, from [10, Theorem 2], $\mathbf{S}_s$ is a minimal dimensional asGLSS of $(\mathbf{y}, \boldsymbol{\mu})$ and it is of the same dimension as $\mathbf{S}$, i.e. of dimension larger than $\hat{\mathbf{S}}_s$, which a contradiction. ■

Note that observability and reachability in the sense of [10] can be characterized by rank conditions of suitable matrices, which can be constructed from the matrices of $\mathbf{S}_s$. We also get the following sufficient condition for isomorphism of minimal sGLSSs in innovation form.

Corollary 3 (Isomorphism): Assume $\mathbf{S}$ is of the form (1) and $\hat{\mathbf{S}} = (\{\hat{A}_\sigma, \hat{B}_\sigma, \hat{K}_\sigma\}_{\sigma \in \Sigma}, \hat{C}, \hat{D}, I, \mathbf{e}^s)$ and they are both sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form and $\mathbf{S}_s$ and $\hat{\mathbf{S}}_s$ are both reachable and observable as stationary **GBS** in the terminology of [10]. Assume that the covariance matrix $E[\mathbf{e}^s(t)(\mathbf{e}^s(t))^T \boldsymbol{\mu}_\sigma^2(t)]$ is nonsingular and $\mathrm{Im}[B_\sigma^T, \hat{B}_\sigma^T]^T \subseteq \mathrm{Im}[K_\sigma^T, \hat{K}_\sigma^T]^T$, $\sigma \in \Sigma$ and $\hat{D} = D$. Then there exists a non-singular matrix T such that for all $\sigma \in \Sigma$,

$$TA_\sigma T^{-1}{=}\hat{A}_\sigma,\ T[K_\sigma, B_\sigma] = [\hat{K}_\sigma, \hat{B}_\sigma],\ CT^{-1}{=}\hat{C} \quad (9)$$

Proof: Since both $\mathbf{S}_s$ and $\hat{\mathbf{S}}_s$ are both minimal asGLSS of $(\mathbf{y}, \boldsymbol{\mu})$ in innovation form, and by [10, Theorem 2], they are isomorphic, i.e., there exists a non-singular matrix T such that $TA_\sigma T^{-1} = \hat{A}_\sigma$, $TK_\sigma = \hat{K}_\sigma$, $CT^{-1} = \hat{C}$ holds. Since $\mathrm{Im}[B_\sigma^T, \hat{B}_\sigma^T]^T \subseteq \mathrm{Im}[K_\sigma^T, \hat{K}_\sigma^T]^T$, for some matrix Z_σ, $B_\sigma = K_\sigma Z_\sigma$, $\hat{B}_\sigma = \hat{K}_\sigma Z_\sigma$, from which (9) follows. ■

Corollary 3 provides sufficient conditions for two sGLSSs in innovation form to be isomorphic, and hence have the same deterministic behavior, as defined in Section I.

IV. Conclusion

We have shown that outputs of stochastic generalized linear switched systems can be decomposed into two parts, deterministic and stochastic one, and we used it to derive existence of representation in innovation form and to formulate sufficient conditions for minimality and uniqueness of such representations up to isomorphism. Future work will be directed towards extending these results for a larger class of inputs and switching signals.

V. Appendix: Proofs of Theorems 1 and 2

A. Proof of Theorem 1

The proof of is an extension of [9, proof of Lemma 1]. Let $\mathcal{H}_{t,+}^{\mathbf{u}}$ be the closed subspace of $\mathcal{H}_1$ (see Notation 2) generated by the components of $\{\mathbf{z}_w^{\mathbf{u}}(t)\}_{w \in \Sigma^+} \cup \{\mathbf{u}(t)\}$.

Lemma 1: The entries of the variables $\mathbf{v}(t)$ and $\{\mathbf{z}_w^{\mathbf{v}}(t)\}_{w \in \Sigma^+}$ are orthogonal to $\mathcal{H}_{t,+}^{\mathbf{u}}$.

Proof: Define $\mathbf{r}(t) := \begin{bmatrix}\mathbf{v}^T(t) & \mathbf{u}^T(t)\end{bmatrix}^T$. By the definition of a sGLSS, $\mathbf{r}$ is ZMWSII and a white noise process w.r.t. $\boldsymbol{\mu}$. Moreover, $\mathbf{v}(t)$ is the upper n_n block of $\mathbf{r}(t)$. Since $\mathbf{r}$ is a white noise w.r.t. $\boldsymbol{\mu}$, $E[\mathbf{r}(t)(\mathbf{z}_w^{\mathbf{r}}(t))^T] = 0$, and $\frac{1}{\sqrt{p_\sigma}} E[\mathbf{v}(t)(\mathbf{z}_w^{\mathbf{u}}(t))^T]$ is the lower-left block of that latter matrix, and hence it is also zero. Since $E[\mathbf{v}(t)(\mathbf{u}(t))^T \boldsymbol{\mu}_i^2(t)] = 0$ and $E[\mathbf{v}(t)(\mathbf{u}(t))^T \boldsymbol{\mu}_i(t)\boldsymbol{\mu}_j(t)] = 0$ for $i \neq j$ due to $\mathbf{r}$ being ZMWSII ([10, Lemma 7]), and $\sum_{i=1}^{n_p} \alpha_i \boldsymbol{\mu}_i = 1$ for some $\{\alpha_i\}_{i=1}^{n_\mu}$, it follows $E[\mathbf{v}(t)(\mathbf{u}(t))^T] = \sum_{i,j=1}^{n_\mu} \alpha_i\alpha_j E[\mathbf{v}(t)(\mathbf{u}(t))^T \boldsymbol{\mu}_i(t)\boldsymbol{\mu}_j(t)] = 0$. That is, $\mathbf{v}(t)$ is orthogonal to $\mathcal{H}_{t,+}^{\mathbf{u}}$. Since $\mathbf{r}(t)$ is a ZMWSII, from [10, Lemma 7] it follows that $E[\mathbf{z}_w^{\mathbf{r}}(t)(\mathbf{z}_v^{\mathbf{r}}(t))^T] = 0$ for all $v \in \Sigma^+$, $v \neq w$ or $v \notin L$ or $w \notin L$, and if $v = w \in L$ and σ is the first letter of w, then $E[\mathbf{z}_w^{\mathbf{r}}(t)(\mathbf{z}_w^{\mathbf{r}}(t))^T] = E[\mathbf{z}_\sigma^{\mathbf{r}}(t)(\mathbf{z}_\sigma^{\mathbf{r}}(t))^T]$.

Since $E[\mathbf{z}_w^{\mathbf{v}}(t)(\mathbf{z}_v^{\mathbf{u}}(t))^T]$ is the upper right block of $E[\mathbf{z}_w^{\mathbf{r}}(t)(\mathbf{z}_v^{\mathbf{r}}(t))^T]$, it follows that $E[\mathbf{z}_w^{\mathbf{v}}(t)(\mathbf{z}_v^{\mathbf{u}}(t))^T] = 0$ if $v \neq w$ and $E[\mathbf{z}_w^{\mathbf{v}}(t)(\mathbf{z}_w^{\mathbf{u}}(t))^T] = E[\mathbf{z}_\sigma^{\mathbf{v}}(t)(\mathbf{z}_\sigma^{\mathbf{u}}(t))^T] = \frac{1}{p_\sigma} E[\mathbf{u}(t-1)\mathbf{v}(t-1)\boldsymbol{\mu}_\sigma^2(t-1)]$, where σ is the first letter of w, and from Definition 5, it follows that the latter expectation is zero. That is, $E[\mathbf{z}_w^{\mathbf{v}}(t)(\mathbf{z}_v^{\mathbf{u}}(t))^T] = 0$ for all $v \in \Sigma^+$.

Since $\mathbf{r}(t)$ is a white noise w.r.t. $\boldsymbol{\mu}$, by [10, Lemma 7] $E[\mathbf{z}_w^{\mathbf{r}}(t)(\mathbf{r}(t))^T] = E[\mathbf{z}_{ws}\mathbf{r}(t)(\mathbf{z}_s^{\mathbf{r}}(t))^T] = 0$ for any $s \in \Sigma^+$, and since $E[\mathbf{z}_w^{\mathbf{v}}(t)(\mathbf{u}(t))^T]$ is the upper right block of $E[\mathbf{z}_w^{\mathbf{r}}(t)(\mathbf{r}(t))^T]$, $E[\mathbf{z}_w^{\mathbf{v}}(t)(\mathbf{u}(t))^T] = 0$. Since $\mathbf{z}_w^{\mathbf{v}}(t)$ is uncorrelated with random variable which generate $\mathcal{H}_{t,+}^{\mathbf{u}}$, the statement of the lemma follows. ■

Let us denote by $\mathcal{H}_t^{\mathbf{u}}$, the closed subspace generated by the components of $\{\mathbf{z}_w^{\mathbf{u}}(t)\}_{w \in \Sigma^+}$. It is clear that $\mathcal{H}_t^{\mathbf{u}} \subseteq \mathcal{H}_{t,+}^{\mathbf{u}}$. Define $\mathbf{x}^d(t) = E_l[\mathbf{x}(t) \mid \{\mathbf{z}_w^{\mathbf{u}}(t)\}_{w \in \Sigma^+} \cup \{\mathbf{u}(t)\}]$.

Lemma 2: The entries of $\mathbf{x}^d(t)$ belong to $\mathcal{H}_t^{\mathbf{u}}$ and

$$\mathbf{x}^d(t) = \sum_{w \in \Sigma^*, \sigma \in \Sigma, \sigma w \in L} \sqrt{p_{\sigma w}} A_w B_\sigma \mathbf{z}_{\sigma w}^{\mathbf{u}}(t), \quad (10)$$

where the convergence is in the mean square sense.

Proof: It is clear from the definition that the components of $\mathbf{x}^d(t)$ belong to $\mathcal{H}_{t,+}^{\mathbf{u}}$. From Lemma 1 it follows that, $E_l[\mathbf{z}_{\sigma w}^{\mathbf{v}}(t) \mid H_{t,+}^{\mathbf{u}}] = 0$, and since the components of $\mathbf{z}_{\sigma w}^{\mathbf{u}}(t)$ belong to $\mathcal{H}_{t,+}^{\mathbf{u}}$, it follows that $E_l[\mathbf{z}_{\sigma w}^{\mathbf{u}}(t) \mid H_{t,+}^{\mathbf{u}}] = \mathbf{z}_{\sigma w}^{\mathbf{u}}(t)$, Since (II) holds and the map $z \mapsto E_l[z \mid M]$ (where $z \in \mathcal{H}_1$) is a continuous linear operator for any closed subspace M, it follows that $\mathbf{x}^d(t)$ will be the infinite sum of the elements $\sqrt{p_{\sigma w}} A_w \left(K_\sigma E_l[\mathbf{z}_{\sigma w}^{\mathbf{v}}(t) \mid H_{t,+}^{\mathbf{u}}] + B_\sigma E_l[\mathbf{z}_{\sigma w}^{\mathbf{u}}(t) \mid H_{t,+}^{\mathbf{u}}]\right)$, i.e., (10) holds. Since the components of $\mathbf{z}_{\sigma w}^{\mathbf{u}}(t)$ belong to $\mathcal{H}_t^{\mathbf{u}}$, the components of the right-hand side of (10) belong to $\mathcal{H}_t^{\mathbf{u}}$ and hence the components of $\mathbf{x}^d(t)$ belong to $\mathcal{H}_t^{\mathbf{u}}$. The convergence of the right-hand side of (10) in the mean square sense follows from that of (II). ■

Lemma 3: Define $\mathbf{x}^s(t) = \mathbf{x}(t) - \mathbf{x}^d(t)$. The entries of $\mathbf{x}^s(t)$ belong to $\mathcal{H}_t^{\mathbf{v}}$, they are orthogonal to $\mathcal{H}_{t,+}^{\mathbf{u}}$ and

$$\mathbf{x}^s(t) = \sum_{w \in \Sigma^*, \sigma \in \Sigma, \sigma w \in L} \sqrt{p_{\sigma w}} A_w K_\sigma \mathbf{z}_{\sigma w}^{\mathbf{v}}(t), \quad (11)$$

where the sum converges in the mean-square sense.

Proof: From (10), $\mathbf{x}^s(t) = \mathbf{x}(t) - \mathbf{x}^d(t)$ and (II), it follows that (11) holds. By Lemma 1, $\{\mathbf{z}_w^{\mathbf{v}}(t)\}_{w \in \Sigma^+}$ are orthogonal to $\mathcal{H}_{t,+}^{\mathbf{u}}$, hence all the summands in the right-hand side of (11) are orthogonal to $\mathcal{H}_{t,+}^{\mathbf{u}}$. ■

Proof: [Proof of Theorem 1] Since $\sum_{\sigma \in \Sigma} p_\sigma A_\sigma \otimes A_\sigma$ is stable and $\mathbf{u}$ and $\mathbf{v}$ are both white noise processes w.r.t. $\boldsymbol{\mu}$, and from (10)-(11) and [10, Lemma 3] it follows that $\mathbf{x}^d$ is the unique state process of $\mathbf{S}_d$ and $\mathbf{x}^s$ is the unique state process of $\mathbf{S}_s$. Notice that

$$\mathbf{y}^d(t) = E_l[\mathbf{y}(t) \mid \mathcal{H}_{t,+}^{\mathbf{u}}] =$$
$$= CE_l[\mathbf{x}(t) \mid \mathcal{H}_{t,+}^{\mathbf{u}}] + DE_l[\mathbf{u}(t) \mid \mathcal{H}_{t,+}^{\mathbf{u}}] + E_l[\mathbf{v}(t) \mid \mathcal{H}_{t,+}^{\mathbf{u}}].$$

By Lemma 1, $\mathbf{v}(t)$ is orthogonal to $\mathcal{H}_{t,+}^{\mathbf{u}}$, $E_l[\mathbf{v}(t) \mid \mathcal{H}_{t,+}^{\mathbf{u}}] = 0$ and as the components $\mathbf{u}(t)$ belong to $\mathcal{H}_{t,+}^{\mathbf{u}}$, $E_l[\mathbf{u}(t) \mid \mathcal{H}_{t,+}^{\mathbf{u}}] = \mathbf{u}(t)$. Hence, $\mathbf{y}^d(t) = C\mathbf{x}^d(t) + D\mathbf{u}(t)$ and $\mathbf{y}^s(t) = C\mathbf{x}^s(t) + F\mathbf{v}(t)$. That is, $\mathbf{S}_d$ is an asGLSS of $(\mathbf{y}^d, \boldsymbol{\mu})$ and $\mathbf{S}_s$ is an asGLSS of $(\mathbf{y}^s, \boldsymbol{\mu})$ respectively. ■

B. Proof of Theorem 2

The proof of Theorem 2 is an adaptation of [9, proof of Lemma 2]. Assume that $\mathbf{S}$ of the form (1) is a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. Denote by $\mathcal{H}_{t,+}^{\mathbf{v}}$ and $\mathcal{H}_t^{\mathbf{v}}$ the closed subspaces of $\mathcal{H}_1$ (see Notation 2) generated by the components $\{\mathbf{z}_w^{\mathbf{v}}(t)\}_{w \in \Sigma^+} \cup \{\mathbf{v}(t)\}$ and $\{\mathbf{z}_w^{\mathbf{v}}(t)\}_{w \in \Sigma^+}$ respectively.

Lemma 4: The entries of $\{\mathbf{z}_v^{\mathbf{y}^s}(t), \mathbf{z}_v^{\mathbf{e}^s}(t)\}_{v \in \Sigma^+}$, $\mathbf{y}^s(t), \mathbf{e}^s(t)$ belong to $\mathcal{H}_{t,+}^{\mathbf{v}}$.

Proof: Recall from the proof of Theorem 1 that $\mathbf{y}^s(t) = C\mathbf{x}^s(t) + \mathbf{v}(t)$. Then by (11), the components of $\mathbf{y}^s(t)$ belong to $\mathcal{H}^{\mathbf{v}}_{t,+}$. Then by [10, Lemma 11], the components of $\mathbf{z}^{\mathbf{y}^s}_v(t)$ belong to $\mathcal{H}^{\mathbf{v}}_{t,+}$ and hence, $\mathcal{H}^{\mathbf{y}^s}_t \subseteq \mathcal{H}^{\mathbf{v}}_{t,+}$. Since $\mathbf{e}^s(t) = \mathbf{y}^s(t) - E_l[\mathbf{y}^s(t) \mid \mathcal{H}^{\mathbf{y}^s}_t]$, this then implies that the components of $\mathbf{e}^s(t)$ belong to $\mathcal{H}^{\mathbf{v}}_{t,+}$. Since $\mathbf{z}^{\mathbf{v}}_v(t) = \sum_{i=1}^{n_\mu} \alpha_i \mathbf{z}^{\mathbf{v}}_{vi}(t+1)$, $\mathbf{v}(t) = \sum_{i=1}^{n_\mu} \alpha_i \mathbf{z}^{\mathbf{v}}_i(t+1)$, as $\sum_{i=1}^{n_\mu} \alpha_i \boldsymbol{\mu}_i = 1$, it follows that $\mathcal{H}^{\mathbf{v}}_{t,+} \subseteq \mathcal{H}^{\mathbf{v}}_{t+1}$ and from [10, Lemma 11] it follows that the components of $\mathbf{z}^{\mathbf{e}^s}_v(t)$ belong to $\mathcal{H}^{\mathbf{v}}_t \subseteq \mathcal{H}^{\mathbf{v}}_{t,+}$. ∎

Lemma 5: The entries of $\{\mathbf{z}^{\mathbf{y}^s}_v(t), \mathbf{z}^{\mathbf{e}^s}_v(t)\}_{v\in\Sigma^+}$, $\mathbf{y}^s(t)$ and $\mathbf{e}^s(t)$ are orthogonal to $\mathcal{H}^{\mathbf{u}}_{t,+}$.

Proof: Using Lemma 1 and it follows that the elements of $\mathcal{H}^{\mathbf{v}}_{t,+}$ are orthogonal to $\mathcal{H}^{\mathbf{u}}_{t,+}$. Since the coordinates of $\mathbf{y}^s(t), \mathbf{e}^s(t), \{\mathbf{z}^{\mathbf{y}^s}_v(t), \mathbf{z}^{\mathbf{e}^s}_v(t)\}_{v\in\Sigma^+}$ belong to $\mathcal{H}^{\mathbf{v}}_{t,+}$, the statement follows. ∎

Lemma 6: $\mathbf{r} = \begin{bmatrix}(\mathbf{e}^s)^T & \mathbf{u}^T\end{bmatrix}^T$ is a white noise process w.r.t. $\boldsymbol{\mu}$ and $E[\mathbf{e}^s(t)\mathbf{u}^T(t)\boldsymbol{\mu}^2_\sigma(t)] = 0$ for all $\sigma \in \Sigma$.

Proof: We first show that $\mathbf{r}$ is a ZMWSII, by showing that $\mathbf{r}$ satisfies the conditions of Definition 2 one by one. First, we show that the processes $\mathbf{r}(t), \mathbf{z}^{\mathbf{r}}_w(t), w \in \Sigma^+$ is zero mean, square integrable.

By assumption $\mathbf{u}$ is a ZMWSII and white noise process w.r.t. $\boldsymbol{\mu}$. From the fact that $\hat{\mathbf{S}}_s$ is an asGLSS of $(\mathbf{y}^s, \boldsymbol{\mu})$ it follows that $\mathbf{e}^s$ is also a ZMWSII. Thus $\mathbf{e}^s(t), \mathbf{u}(t), \{\mathbf{z}^{\mathbf{e}^s}_w(t), \mathbf{z}^{\mathbf{u}}_w(t)\}_{w\in\Sigma^+}$ is zero mean, square integrable. From this it follows that $\mathbf{r}(t)$ and $\mathbf{z}^{\mathbf{r}}_w(t)$ are zero mean and square integrable.

From Lemma 4 it follows that the components $\mathbf{e}^s(t)$ belongs to $\mathcal{H}^{\mathbf{v}}_{t,+}(t)$. Moreover, by definition of sGLSS, $\mathbf{w} = \begin{bmatrix}\mathbf{v}^T & \mathbf{u}^T\end{bmatrix}^T$ is ZMWSII. Hence, with the notation of Definition 2, the σ-algebras $\mathcal{F}^{\mathbf{w}}_t$ and $\mathcal{F}^{\boldsymbol{\mu},+}_t$ are conditionally independent w.r.t. $\mathcal{F}^{\boldsymbol{\mu},-}_t$. From the fact that $\mathbf{e}^s(t)$ belongs to $\mathcal{H}^{\mathbf{v}}_{t,+}$ it follows that $\mathbf{e}^s(t)$ is measurable with respect to the σ-algebra generated by $\{\mathbf{v}(t)\} \cup \{\mathbf{z}^{\mathbf{v}}_v(t)\}_{v\in\Sigma^+}$ and the latter σ-algebra is a subset of $\mathcal{F}^{\mathbf{w}}_t \vee \mathcal{F}^{\boldsymbol{\mu},-}_t$, where for two σ-algebras $\mathcal{F}_i$, $i = 1, 2$, $\mathcal{F}_1 \vee \mathcal{F}_2$ denotes the smallest σ-algebra generated by the σ-algebras $\mathcal{F}_1, \mathcal{F}_2$. That is, $\mathbf{e}^s(t)$ is measurable w.r.t. the σ algebra $\mathcal{F}^{\mathbf{w}}_t \vee \mathcal{F}^{\boldsymbol{\mu},-}_t$. Hence, $\mathcal{F}^{\mathbf{r}}_t \subseteq \mathcal{F}^{\mathbf{w}}_t \vee \mathcal{F}^{\boldsymbol{\mu},-}_t$. Since $\mathcal{F}^{\mathbf{w}}_t$ and $\mathcal{F}^{\boldsymbol{\mu},+}$ are conditionally independent w.r.t. $\mathcal{F}^{\boldsymbol{\mu},-}_t$, from [16, Proposition 2.4] it follows that $\mathcal{F}^{\mathbf{w}}_t \vee \mathcal{F}^{\boldsymbol{\mu},-}_t$ and $\mathcal{F}^{\boldsymbol{\mu},+}_t$ are conditionally independent w.r.t. $\mathcal{F}^{\boldsymbol{\mu},-}_t$, and as $\mathcal{F}^{\mathbf{r}}_t \subseteq \mathcal{F}^{\mathbf{w}}_t \vee \mathcal{F}^{\boldsymbol{\mu},-}_t$, it follows that $\mathcal{F}^{\mathbf{r}}_t$ and $\mathcal{F}^{\boldsymbol{\mu},+}$ are conditionally independent w.r.t. $\mathcal{F}^{\boldsymbol{\mu},-}_t$.

Finally, we show that $\mathbf{r}(t), \mathbf{z}^{\mathbf{r}}_w(t), w \in \Sigma^+$ are jointly wide-sense stationary, i.e., for all $s, t, \in \mathbb{Z}$, $s \leq t$, $v, w \in \Sigma^+$, $E[\mathbf{h}_1(t)(\mathbf{h}_1(t)^T]$, where $\mathbf{h}_1(t), \mathbf{h}_2(t) \in \{\mathbf{r}(t)\} \cup \{\mathbf{z}^{\mathbf{r}}_w(t), w \in \Sigma^+\}$ does not depend on t. We show only the case, $E\left[\mathbf{r}(t)(\mathbf{z}^{\mathbf{r}}_w(s))^T\right] = E\left[\mathbf{r}(t-s)(\mathbf{z}^{\mathbf{r}}_w(0))^T\right]$, the proof of the general case is similar. From Lemma (6) it follows that $E\left[\mathbf{z}^{\mathbf{e}^s}_w(t+k)(\mathbf{z}^{\mathbf{u}}_v(s+k))^T\right] = 0$, and hence the matrix $E\left[\mathbf{z}^{\mathbf{r}}_w(t)(\mathbf{z}^{\mathbf{r}}_v(s))^T\right]$ is a block-diagonal one, where the blocks on the diagonal are $E\left[\mathbf{z}^{\mathbf{e}^s}_w(t)(\mathbf{z}^{\mathbf{e}^s}_v(s))^T\right]$ and $E\left[\mathbf{z}^{\mathbf{u}}_w(t)(\mathbf{z}^{\mathbf{u}}_v(s))^T\right]$ and by $\mathbf{u}$ and $\mathbf{e}^s$ being ZMWSII, it follows that the latter do not depend on t. That is, we have shown that $\mathbf{r}$ satisfies all the conditions of Definition 2.

Next we show that $\mathbf{r}$ is a white noise process w.r.t. $\boldsymbol{\mu}$. To this end, by [10, Lemma 7], it is enough to show that $E[\mathbf{r}(t)(\mathbf{z}^{\mathbf{r}}_w(t))^T] = 0$ for all $w \in \Sigma^+$. From Lemma 6 it follows that $E\left[\mathbf{r}(t)(\mathbf{z}^{\mathbf{r}}_w(t))^T\right]$ is block diagonal, with the block on the diagonal being $E\left[\mathbf{e}^s(t)(\mathbf{z}^{\mathbf{e}^s}_w(t))^T\right]$, $E\left[\mathbf{u}(t)(\mathbf{z}^{\mathbf{u}}_w(t))^T\right]$, and the latter are zero as $\mathbf{e}_s$ and $\mathbf{u}$ are white noise w.r.t $\boldsymbol{\mu}$. Finally, $E[\mathbf{e}^s(t)\mathbf{u}^T(t)\boldsymbol{\mu}^2_\sigma(t)] = 0$ for all $\sigma \in \Sigma$ follows from Lemma 5. ∎

Proof: [Proof of Theorem 2] From Lemma 6 it follows that the noise process $\mathbf{e}^s$ and the input $\mathbf{u}$ satisfy the condition of $E[\mathbf{e}^s(t)(\mathbf{u}(t))^T\boldsymbol{\mu}^2_\sigma(t)] = 0$, $\sigma \in \Sigma$. Since $\hat{\mathbf{S}}_s$ and $\hat{\mathbf{S}}_d$ are both asGLSS, it follows that $\sum_{i=1}^{n_\mu} p_i \hat{A}_i \otimes \hat{A}_i$ is stable. Hence $\hat{\mathbf{S}}$ satisfies the conditions of a sGLSS. Let $\hat{\mathbf{x}}^s$ and $\hat{\mathbf{x}}^d$ be the unique state processes of $\hat{\mathbf{S}}_s$ and $\hat{\mathbf{S}}_d$ respectively. We claim that $\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}^d(t) + \hat{\mathbf{x}}^s(t)$ is the unique state process of $\hat{\mathbf{S}}$. Indeed, $\hat{\mathbf{x}}(t+1) = \sum_{i=1}^{n_\mu}(\hat{A}_i\hat{\mathbf{x}}() + \hat{B}_i\mathbf{u}(t) + \hat{K}_i\mathbf{e}^s(t))\pi_1(t)$ holds and $\hat{\mathbf{x}}(t)$ is a ZMWSII, as it is a sum of two ZMWSII processes. Hence, $\hat{\mathbf{x}}(t)$ is the uniqe state process of $\hat{\mathbf{S}}$. Finally, from $\mathbf{y}^d(t) = \hat{C}\hat{\mathbf{x}}^d(t) + \hat{D}\mathbf{u}(t)$ (as $\hat{\mathbf{S}}_d$ is an asGLSS of $(\mathbf{y}^d, \boldsymbol{\mu})$) and $\mathbf{y}^s(t) = \hat{C}\hat{\mathbf{x}}^s(t) + \mathbf{e}^s(t)$ (as $\hat{\mathcal{S}}_s$ is an asGLSS of $(\mathbf{y}^s, \boldsymbol{\mu})$), it follows that $\mathbf{y}(t) = \hat{C}\hat{\mathbf{x}}(t) + \hat{D}\mathbf{u}(t) + \mathbf{e}^s(t)$, i.e., $\hat{\mathbf{S}}$ is a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. ∎

References

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

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

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

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

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

[6] F. Lauer and G. Bloch. *Hybrid System Identification: Theory and Algorithms for Learning Switching Models*. Springer, 2019.

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

[8] M. Mejari and M. Petreczky. Realization and identification algorithm for stochastic LPV state-space models with exogenous inputs. In *3rd IFAC LPVS*, 2019.

[9] M. Mejari and M/ Petreczky. Realization and identification algorithm for stochastic LPV state-space models with exogenous inputs. 2019. arXiv 1905.10113.

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

[11] D. Piga, P. Cox, R. Tóth, and V. Laurain. LPV system identification under noise corrupted scheduling and output signal observations. *Automatica*, 53:329–338, 2015.

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

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

[14] H. Tanaka and K. Ikeda. State estimation for closed-loop lpv system identification via kernel methods. *IFAC-PapersOnLine*, 56(2):11669–11674, 2023.

[15] R. Tóth. *Modeling and Identification of Linear Parameter-Varying Systems*. Springer, 2010.

[16] C. van Putten and J. H. van Schuppen. Invariance properties of the conditional independence relation. *Ann. Probab.*, 13(3):934–945, 1985.

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

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

[19] V. Verdult and M. Verhaegen. Subspace identification of piecewise linear systems. In *CDC*, 2004.