Realization Theory of Stochastic Jump-Markov Linear Systems

Mihály Petreczky, René Vidal

Abstract

In this paper, we present a complete stochastic realization theory for stochastic jump-linear systems. We present necessary and sufficient conditions for the existence of a realization, along with a characterization of minimality in terms of reachability and observability. We also formulate a realization algorithm and argue that minimality can be checked algorithmically. The main tool for solving the stochastic realization problem for jump-linear systems is the formulation and solution of a stochastic realization problem for a general class of bilinear systems with non-white-noise inputs. The solution to this generalized stochastic bilinear realization problem is based on the theory of formal power series. Stochastic jump-linear systems represent a special case of generalized stochastic bilinear systems.

I. Introduction

Hybrid systems are dynamical systems that exhibit both continuous and discrete behaviors. Such systems have a wide range of applications, including systems biology, computer vision, flight control systems, etc. While there is a vast amount of literature on stability, reachability, observability, identification, and controller design for hybrid systems, there are relatively fewer results available on realization theory of hybrid systems.

Realization theory is one of the central topics of control and systems theory. Its goals are to study the conditions under which the observed behavior of a system can be represented by a state-space representation of a certain type and to develop algorithms for finding a (preferably

Mihály Petreczky is with Centre de Recherche en Informatique, Signal et Automatique de Lille (CRIStAL), UMR CNRS 9189, Ecole Centrale de Lille, 59651 Villeneuve dAscq, France (e-mail: mihaly.petreczky@ec-lille.fr).

René Vidal is with the Center for Imaging Science, Department of Biomedical Engineering, Johns Hopkins University, 3400 N. Charles Street, Baltimore MD 21218, USA (e-mail: rvidal@cis.jhu.edu).

minimal) state-space representation of the observed behavior. The study of these problems is not only of theoretical relevance, but also of practical importance in several applications such as model reduction and systems identification. In fact, one can argue that stochastic realization theory is indispensable for the understanding of systems identification.

A. State-of-the-art

For the class of linear systems, the realization problem is relatively well understood thanks to the work of Kalman et al. in the sixties [1], [2]. For instance, it is well known that all minimal representations, i.e. representations such that the dimension of the state-space is minimal, are related by a change of basis of the state-space. Also, it is well known that the rank of a Hankel matrix H formed from the output measurements is related the dimension of all minimal representations and that a realization of the system can be obtained from the factorization of H. Such results have lead to a huge literature on identification of linear systems [3], including the well-known subspace identification methods [4].

For the class of bilinear systems, the realization problem is also relatively well studied thanks to the works of Brockett [5], Fliess [6], Isidori et al. [7], [8], [9], Sontag [10] and Sussman [11], [12] in the 1970's. However, realization of stochastic bilinear systems is relatively unstudied, except the case when input is white noise [13], [14]. On the other hand, there are a number of papers on identification of bilinear systems with inputs which are not white noise, see e.g., [15], [16], [17], [18]. However, all these papers require a number of conditions on the underlying system in order to operate correctly.

For more general nonlinear systems, the realization problem is not as well understood. There exists a complete realization theory for analytic nonlinear systems [19], [20], [21], [22], [23], [24] and for general smooth systems [25], [26]. However, the algorithmic aspects of this theory are not that well developed. There is a substantial amount of work on realization theory of polynomial systems [27], [28], and rational systems [29], [30], [31] both in continuous and discrete time. However, the issue of minimality for polynomial systems is not that well understood.

One of the earliest attempts to characterize realization of deterministic hybrid systems can be found in [32], though a formal theory is not presented. Since then, most of the work has concentrated on switched linear systems [33], [34], switched bilinear systems [35], linear and bilinear hybrid systems without guards and partially observed discrete states [36], [37], and nonlinear analytic hybrid systems without guards [38]. The main assumptions made are that the continuous dynamics evolve in continuous-time and the discrete events which initiate the change of the discrete states are part of the input. Hence, the discrete states may (switched systems) or may not (hybrid systems without guards) be fully observed. For the classes of hybrid systems mentioned above, with the exception of nonlinear hybrid systems without guards, a complete realization theory and realization algorithms are available. [39] contains partial results on realization theory of piecewise-affine autonomous hybrid systems with guards. In that paper necessary and sufficient conditions for existence of a realization were presented, but the problem of minimality was not dealt with. As far as the authors know, the only paper dealing with realization theory of stochastic hybrid systems is [40], where only necessary conditions for the existence of a realization were presented.

B. Paper contributions

In this paper we will present a complete stochastic realization theory of discrete-time stochastic jump-linear systems. Stochastic jump-linear systems have a vast literature and numerous applications (see for example [41] and the references therein). For simplicity, we will consider only stochastic jump-linear systems with fully observed discrete state. In addition, we will assume that the continuous state-transition depends not only on the current, but also on the next discrete state and that the continuous state at each time instant lives in a state-space that depends on the current discrete state. In this way we will obtain a more general model, which we will call *generalized stochastic jump Markov linear systems*. It turns out that the class of classical stochastic jump-linear systems generates the same class of output processes as the new more general class. However, by looking at more general systems we are able to obtain a neat characterization of minimality as well as necessary and sufficient conditions for the existence of a realization. We will also formulate a realization algorithm and argue that minimality can be checked algorithmically.

The main tool for solving the stochastic realization problem is the solution of a general bilinear realization problem, whose formulation and solution can be described as follows. Consider an output and an input process and imagine you would like to compute recursively the linear projection of the future outputs onto the space of products of past outputs and inputs. Under the assumption that the mixed covariances of the future outputs with the products of past outputs and inputs form a *rational formal power series*, we show that one can construct a bilinear state-space representation of the output process in the forward innovation form. The matrices of this state-space representation are determined by the parameters of the rational representation of the covariance sequence of future and past outputs and inputs. The results on realization theory of stochastic jump-linear systems are then obtained by viewing the discrete state process as an input process.

To the best of our knowledge, both the solution of the realization problem for stochastic jump-linear systems, and the formulation and solution of the general bilinear realization problem are new. In comparison the work of [40] on stochastic realization of jump-linear systems, the main contribution of this paper is that it presents both necessary and sufficient conditions for the existence of a realization as well as a characterization of minimality. In comparison to the work of [13] on stochastic realization of bilinear systems with observed white-noise input process, the main contribution of this paper is to solve the realization problem for a more general class of bilinear systems, without requiring the input process to be white. In comparison with the works of [15], [16], [17], [18] on identification of bilinear systems with inputs that are not necessarily white noise, there are two main contributions. First, the aforementioned papers aim to identify the parameters of the system from the measurements. In contrast, the goal of realization theory is to understand the conditions, under which a (not necessarily identifiable) state-space representation exists. Hence, establishing algorithms for finding the parameters of the system that generate the process answers the realization problem only partially. Second, all the aforementioned papers assume that the system to be identified is already in the forward innovation form and impose a number of observability and stability conditions on the underlying system, which are more restrictive than the conditions assumed here.

C. Paper outline

The outline of the paper is as follows. Section II presents the background material on the theory of rational formal power series. These results will be instrumental for solving the generalized bilinear realization problem, which will be formulated and solved in Section III. Section V formulates the realization problem for stochastic jump Markov linear systems and presents a solution to it based on the results in Section III.

II. RATIONAL POWER SERIES

In this section, we present several results on formal power series, which will be used for solving a general bilinear filtering/realization problem to be presented in Section III. In turn, the solution to this bilinear filtering/realization problem will yield a solution to the realization problem for stochastic jump-linear systems, as we will show in Section V.

The material and results in Subsections II-A and II-B can be found in [42] and [40], respectively. For more details on the classical theory of rational formal power series, the reader is referred to [43], [44], [28] and the references therein.

A. Definition and Basic Theory

Let Σ be a finite set. We will refer to Σ as the *alphabet*. The elements of Σ will be called *letters*, and every finite sequence of letters will be called a *word* or *string* over Σ. Denote by Σ^* the set of all finite words from elements in Σ. An element $w \in \Sigma^*$ of length $|w| = k \geq 0$ is a finite sequence $w = \sigma_1\sigma_2 \cdots \sigma_k$ with $\sigma_1, \ldots, \sigma_k \in \Sigma$. The empty word is denoted by ϵ and its length is zero, i.e. $|\epsilon| = 0$. Denote by Σ^+ the set of all non-empty words over Σ, i.e. $\Sigma^+ = \Sigma^* \setminus \{\epsilon\}$. The concatenation of two words $v = \nu_1 \cdots \nu_m$ and $w = \sigma_1 \cdots \sigma_k \in \Sigma^*$ is the word $vw = \nu_1 \cdots \nu_m \sigma_1 \cdots \sigma_k$.

Definition 1 (Lexicographic ordering): Let $<$ be an ordering on Σ so that $\Sigma = \{\sigma_1, \ldots, \sigma_{|\Sigma|}\}$ with $\sigma_1 < \sigma_2 < \ldots < \sigma_{|\Sigma|}$. We define a lexicographic ordering $\prec$ on Σ^* as follows. For any $v = \nu_1 \cdots \nu_m$ and $w = \sigma_1 \cdots \sigma_k \in \Sigma^*$, $v \prec w$ if either $|v| < |w|$ or $0 < |v| = |w|$, $v \neq w$ and for some $l \leq |w|$, $\nu_l < \sigma_l$ with the ordering $<$ on Σ an $\nu_i = \sigma_i$ for $i = 1, \ldots, l-1$.

Notice that $\prec$ is a complete ordering and that $\Sigma^* = \{v_0, v_1, \ldots\}$ with $v_0 \prec v_1 \prec \ldots$. Therefore, we will call the set $\{v_0, v_1, \ldots\}$ an *ordered enumeration* of Σ^*. Notice also that $v_0 = \epsilon$ and that for all $i \in \mathbb{N}$ and $\sigma \in \Sigma$, we have $\nu_i \prec \nu_i\sigma$. Moreover, denote by $M(N)$ the number of all non-empty words over Σ whose length is at most N, i.e. $M(N) = |\{w \in \Sigma^+ \mid |w| \leq N\}|$. It then follows that with the lexicographic ordering defined above, the set $\{v_0, v_1, \ldots, v_{M(N)}\}$ equals to the set of all words of length at most $M(N)$, including the empty word.

A *formal power series* S with coefficients in $\mathbb{R}^p$ is a map $S : \Sigma^* \to \mathbb{R}^p$. We will call the values $S(w) \in \mathbb{R}^p$, $w \in \Sigma^*$, the *coefficients* of S. We will denote by $\mathbb{R}^p \ll \Sigma^* \gg$ the set of all formal power series with coefficients in $\mathbb{R}^p$. Consider a family of formal power series

$\Psi = \{S_j \in \mathbb{R}^p \ll \Sigma^* \gg \mid j \in J\}$ indexed with a finite index set J. We will call such an indexed set of formal power series a *family of formal power series*.

A family of formal power series Ψ will be called *rational* if there exists an integer $n \in \mathbb{N}$, a matrix $C \in \mathbb{R}^{p \times n}$, a collection of matrices $A_\sigma \in \mathbb{R}^{n \times n}$ indexed by $\sigma \in \Sigma$, and an indexed set $B = \{B_j \in \mathbb{R}^n \mid j \in J\}$ of vectors in $\mathbb{R}^n$, such that for each index $j \in J$ and for all sequences $\sigma_1, \ldots, \sigma_k \in \Sigma$, $k \geq 0$,

$$S_j(\sigma_1 \sigma_2 \cdots \sigma_k) = C A_{\sigma_k} A_{\sigma_{k-1}} \cdots A_{\sigma_1} B_j. \tag{1}$$

The 4-tuple $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma \in \Sigma}, B, C)$ will be called a *representation* of Ψ and the number $n = \dim R$ will be called the *dimension* of R. A representation R_{min} of Ψ will be called *minimal* if all representations R of Ψ satisfy $\dim R_{min} \leq \dim R$. Two representations of Ψ, $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma \in \Sigma}, B, C)$ and $\widetilde{R} = (\mathbb{R}^n, \{\widetilde{A}_\sigma\}_{\sigma \in \Sigma}, \widetilde{B}, \widetilde{C})$, will be called *isomorphic*, if there exists a nonsingular matrix $T \in \mathbb{R}^{n \times n}$ such that$T\widetilde{A}_\sigma = A_\sigma T$ for all $\sigma \in \Sigma$, $T\widetilde{B}_j = B_j$ for all $j \in J$, and $\widetilde{C} = CT$.

Let $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma \in \Sigma}, B, C)$ be a representation of Ψ. In order to characterize whether this representation is reachable and observable, let us define the following short-hand notation

Notation 1: $A_w \doteq A_{\sigma_k} A_{\sigma_{k-1}} \cdots A_{\sigma_1}$ for $w = \sigma_1 \cdots \sigma_k \in \Sigma^*$ and $\sigma_1, \ldots, \sigma_k \in \Sigma$, $k \geq 0$. The map A_ϵ will be identified with the identity map.

— Recall the ordered enumeration of Σ^*, $\{v_0, v_1, \ldots\}$, fix an enumeration of $J = \{j_1, \ldots, j_K\}$ and let $\widetilde{B} = \begin{bmatrix} B_{j_1} & \cdots & B_{j_K} \end{bmatrix}$. Define the following matrices.

$$W_R = \begin{bmatrix} A_{v_0}\widetilde{B} & , \ldots , & A_{v_{M(n-1)}}\widetilde{B} \end{bmatrix} \tag{2}$$

$$O_R = \begin{bmatrix} (CA_{v_0})^T & \ldots & (CA_{v_{M(n-1)}})^T \end{bmatrix}^T. \tag{3}$$

We will call the representation R *observable* if $\ker O_R = \{0\}$ and *reachable* if $\dim R = \operatorname{rank} W_R$. Observability and reachability of representations can be checked numerically. For instance, one can formulate an algorithm for transforming any representation to a minimal representation of the same family of formal power series (see [42] and the references therein for details).

Let $\Psi = \{S_j \in \mathbb{R}^p \ll \Sigma^* \gg \mid j \in J\}$ be a family of formal power series and define $I = \{1, \ldots, p\}$. We define the Hankel-matrix H_Ψ of Ψ as the matrix such that the following holds. The rows of H_Ψ are indexed by pairs (u, i) where $u \in \Sigma^*$ is a word over Σ and i is and integer in $I = \{1, 2, \dots, p\}$. Likewise, the columns of H_Ψ are indexed by pairs (v, j), where $v \in \Sigma^*$ and j is an element of the index set J. Thus, the element of H_Ψ whose row index is (u, i) and whose column index is (v, j) is simply the ith row of the vector $S_j(vu) \in \mathbb{R}^p$, i.e. $(H_\Psi)_{(u,i)(v,j)} = (S_j(vu))_i$.

The following result on realization of formal power series can be found in [44], [28], [42].

Theorem 1 (Realization of formal power series): Let $\Psi = \{S_j \in \mathbb{R}^p \ll \Sigma^* \gg \mid j \in J\}$ be a set of formal power series indexed by J. Then the following holds.

(i) Ψ is rational $\iff$ $\mathrm{rank}\ H_\Psi < +\infty$.

(ii) R is a minimal representation of Ψ $\iff$ R is reachable and observable $\iff$ $\dim R = \mathrm{rank}\ H_\Psi$.

(iii) All minimal representations of Ψ are isomorphic.

It is possible to compute a minimal representation of Ψ from finitely many data. The procedure resembles very much the partial realization algorithms for linear systems. One defines the finite matrix $H_{\Psi,M,N}$ as the finite upper-left block of the infinite Hankel matrix H_Ψ obtained by taking all the rows of H_Ψ indexed by words over Σ of length at most M, and all the columns of H_Ψ indexed by words of length at most N. If $\mathrm{rank}\ H_{\Psi,N,N} = \mathrm{rank}\ H_\Psi$ holds, then there exists an algorithm for computing a minimal representation R_N of Ψ. The algorithm is essentially a generalization of the well-known Kalman-Ho algorithm [1] for linear systems. The condition $\mathrm{rank}\ H_{\Psi,N,N} = \mathrm{rank}\ H_\Psi$ holds, if, for example, N is chosen to be bigger than the dimension of some representation of Ψ. In practice, this means that N has to be an upper bound on the estimated dimension of a potential representation of Ψ. More details on the computation of a minimal representation from a Hankel-matrix can be found in [42] and the references therein.

For the purposes of this paper we will use a specific version of the realization algorithm. In order to present the algorithm, we define the notion of r, N*-selection*: an r, N*-selection* is a pair $(\alpha, beta)$ such that

1) $\alpha \subseteq \Sigma^N \times \{1, \dots, p\}$, $\beta \subseteq \Sigma^N \times J$, $\Sigma^N = \{v \in \Sigma^* \mid |v| \leq N\}$,
2) $|\alpha| = |\beta| = r$.

Intuitively, α represents a selection of r rows of $H_{\Psi,N,N}$ and β represents a selection of r columns of $H_{\Psi,N,N}$. Let (α, β) be an r, N-selection. The proposed algorithm takes as parameter the matrix $H_{\Psi,N+1,N}$ and an r, N-selection (α, β). In addition, we assume that the r, N-selection

(α, β) is such that the following holds. Let $H_{\Psi,\alpha,\beta}$ be the matrix formed by the intersection of the columns of $H_{\Psi,N,N}$ indexed by elements of β with the rows of $H_{\Psi,N,N}$ indexed by the elements of α. We then assume that $\mathrm{rank}\ H_{\Psi,\alpha,\beta} = \mathrm{rank}\ H_{\Psi,N,N+1}$.

Algorithm 1

Inputs: matrix $H_{\Psi,N+1,N}$ and r, N-selection (α, β)

Output: representation $\widetilde{R}_N$.

For each symbol $\sigma \in \Sigma$ let $A_\sigma \in \mathbb{R}^{r \times r}$ be such that

$$A_\sigma H_{\Psi,\alpha,\beta} = Z_\sigma$$

where Z_σ is $r \times r$ matrix with row indices from α and column indices from β such that its entry indexed by $z \in \alpha$, $(v, j) \in \beta$ equals the entry of $H_{\Psi,N,N+1}$ indexed by $(z, (v\sigma, j))$.

Let $B = \{B_j \mid j \in J\}$, where for each index $j \in J$, the vector $B_j \in \mathbb{R}^r$ is formed by those entries of the column (ϵ, j) of H_Ψ which are indexed by the elements of α.

Let $C \in \mathbb{R}^{p \times r}$ whose ith row is the interesection of the row indexed by (ϵ, i) with the columns of H_Ψ indexed by the elements of β, $i = 1, 2, \ldots, p$.

Return $\widetilde{R}_N = (\mathbb{R}^r, \{A_\sigma\}_{\sigma \in \Sigma}, B, C)$.

Theorem 2 ([42], [44], [45]): If $r = \mathrm{rank}\ H_{\Psi,N,N} = \mathrm{rank}\ H_\Psi$, then there exists an r, N-selection (α, β) such that $\mathrm{rank}\ H_{\Psi,\alpha,\beta} = r$ and the the representation $\widetilde{R}_N$ returned by Algorithm 1 when applied to $H_{\Psi,N+1,N}$ and (α, β) is minimal representation of Ψ. Furthermore, if $\mathrm{rank}\ H_\Psi \leq N$, or, equivalently, there exists a representation R of Ψ, such that $\dim R \leq N$, then $\mathrm{rank}\ H_\Psi = \mathrm{rank}\ H_{\Psi,N,N}$, hence $\widetilde{R}_N$ is a minimal representation of Ψ.

B. A Notion of Stability for Formal Power Series

Since our goal is to use formal power series to build a stochastic realization theory for jump-linear systems, we will need to restrict our attention to formal power series that are stable in some sense, similarly to the case of linear systems. In this subsection, we consider the notion of square summability for formal power series, and translate the requirement of square summability into algebraic properties of their representations.

More specifically, consider a formal power series $S \in \mathbb{R}^p \ll \Sigma^* \gg$ and define the sequence

$$L_n = \sum_{k=0}^{n} \sum_{\sigma_1 \in \Sigma} \cdots \sum_{\sigma_k \in \Sigma} ||S(\sigma_1 \sigma_2 \cdots \sigma_k)||_2^2. \tag{4}$$

where $|| \cdot ||_2$ is the Euclidean norm in $\mathbb{R}^p$. The series S will be called *square summable*, if the limit $\lim_{n \to +\infty} L_n$ exists and it is finite. The family $\Psi = \{S_j \in \mathbb{R}^p \ll \Sigma^* \gg | \ j \in J\}$ will be called *square summable*, if for each $j \in J$, the formal power series S_j is square summable.

We now characterize square summability of a family of formal power series in terms of the stability of its representation. Let $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma \in \Sigma}, B, C)$ be an arbitrary representation of $\Psi = \{S_j \in \mathbb{R}^p \ll \Sigma^* \gg | \ j \in J\}$. Assume that $\Sigma = \{\sigma_1, \ldots, \sigma_d\}$, where d is the number of elements of Σ, and consider the matrix $\widetilde{A} = \sum_{i=1}^{d} A_{\sigma_i}^T \otimes A_{\sigma_i}^T$, where $\otimes$ denotes the Kronecker product. We will call R *stable*, if the matrix $\widetilde{A}$ is stable, i.e. if all its eigenvalues λ lie inside the unit disk ($|\lambda| < 1$). We have the following.

Theorem 3: Consider a family of formal power series Ψ. If Ψ admits a stable representation, then Ψ is square summable. If Ψ is square summable, then any minimal representation of Ψ is stable.

Notice the analogy with the case of linear systems, where the minimal realization of a stable transfer matrix is also stable.

Proof of Theorem 3: Assume that Ψ has a stable representation $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma \in \Sigma}, C, B)$. Then all the eigenvalues of the matrix $\widetilde{A} = \sum_{\sigma \in \Sigma} A_\sigma^T \otimes A_\sigma^T$ are inside the unit circle. One can easily see that the matrix $\widetilde{A}$ is in fact a matrix representation of the linear map $\mathcal{Z} : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ defined as

$$\mathcal{Z}(V) = \sum_{\sigma \in \Sigma} A_\sigma^T V A_\sigma.$$

This result is obtained by identifying $\mathbb{R}^{n \times n}$ with $\mathbb{R}^{n^2}$, as it is done in [41, Section 2.1]. As a consequence, the eigenvalues of $\mathcal{Z}$ and $\widetilde{A}$ coincide. Since the eigenvalues of $\mathcal{Z}$ are inside the unit circle, it follows from [41, Proposition 2.5] that for each positive semi-definite matrix $V \geq 0$, the infinite sum $\sum_{k=0}^{\infty} \|\mathcal{Z}^k(V)\|$ is convergent. By noticing that

$$\forall x \in \mathbb{R}^n \quad x^T \mathcal{Z}^k(V) x \leq \|x\|_2^2 \cdot \|\mathcal{Z}^k(V)\|,$$

we conclude that $\sum_{k=0}^{\infty} x^T \mathcal{Z}^k(V) x$ is convergent for all x. It can be shown by induction that

$$\mathcal{Z}^k(V) = \sum_{w \in \Sigma^*, |w| = k} A_w^T V A_w. \tag{5}$$

Thus, letting $V = C^TC$ in $\sum_{k=0}^{\infty} x^T \mathcal{Z}^k(V)x$, we conclude that $\sum_{w\in\Sigma^*} \|CA_w x\|_2^2$ is convergent for all x. If we set $x = B_j$, $j \in J$, we then obtain that $\sum_{w\in\Sigma^*} \|S_j(w)\|_2^2$ is convergent for all $j \in J$, i.e. the family Ψ is square summable.

Assume now that Ψ is square summable and let $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma\in\Sigma}, C, B)$ be a minimal representation of Ψ. Also, let $Q = O_R^T O_R > 0$, where O_R is the observability matrix of R, which is full rank because R is observable. First we show that

$$\sum_{k=0}^{\infty} x^T \mathcal{Z}^k(Q)x = \sum_{w\in\Sigma^*} x^T A_w^T Q A_w x \tag{6}$$

is convergent for all $x \in \mathbb{R}^n$. To see this, notice from the reachability of R that any $x \in \mathbb{R}^n$ is a linear combination of vectors of the form $A_v B_j$, $j \in J$, $v \in \Sigma^*$. Hence, it is sufficient to prove the convergence of (6) for $x = A_v B_j$. But the latter follow from the fact that

$$\sum_{w\in\Sigma^*} (B_j A_v)^T A_w^T Q A_w A_v B_j = \sum_{w\in\Sigma^*} \sum_{i=0}^{M(n-1)} ||S_j(vwv_i)||_2^2$$

and that $\sum_{w\in\Sigma^*} \|S_j(w)\|_2^2$, hence $\sum_{w\in\Sigma^*} \sum_{i=0}^{M(n-1)} \|S_j(vwv_i)\|_2^2$ is convergent. Next we show that

$$\sum_{k=0}^{\infty} x^T \mathcal{Z}^k(V)x = \sum_{w\in\Sigma^*} x^T A_w^T V A_w x \tag{7}$$

is convergent for all $x \in \mathbb{R}^n$ and for all positive semi-definite $n \times n$ matrices $V \geq 0$. To see this, notice that for all $V \geq 0$ and $Q > 0$, there exists $M > 0$ such that $x^T V x \leq M x^T Q x$ for all $x \in \mathbb{R}^n$. Indeed, we can choose $M = \frac{\|V\|}{m}$, where $0 < m = \inf_{\|x\|=1} x^T Q x$, so that $m\|x\|^2 \leq x^T Q x$ and hence $x^T V x \leq \|x\|^2 \|V\| \leq M x^T Q x$. Therefore, for any $V \geq 0$,

$$\sum_{k=0}^{\infty} x^T \mathcal{Z}^k(V)x = \sum_{w\in\Sigma^*} x^T A_w^T V A_w x \leq M \sum_{w\in\Sigma^*} x^T A_w^T Q A_w x = M \sum_{k=0}^{\infty} x^T \mathcal{Z}^k(Q)x,$$

and so $\sum_{k=0}^{\infty} x^T \mathcal{Z}^k(V)x$ is convergent for all $x \in \mathbb{R}^n$ and $V \geq 0$. This implies that

$$\lim_{k\to\infty} x^T \mathcal{Z}^k(V)x = 0$$

for all $x \in \mathbb{R}^n$. Therefore, $\lim_{k\to\infty} \mathcal{Z}^k(V) = 0$ for all $V \geq 0$, which by [41, Proposition 2.5] implies that all the eigenvalues of $\mathcal{Z}$ (and hence of $\widetilde{A}$) have modulus strictly smaller than 1, i.e. R is stable.

■

III. Stochastic Realization of Generalized Bilinear Systems

In this section we formulate and solve the realization problem for generalized stochastic bilinear systems (abbreviated by **GBS**). A **GBS** is stochastic system which is bilinear in state and inputs and where the inputs is an observed stochastic process. Informally, the realization problem can be formulated as follows: given an output process and input process, find a **GBS**which is driven by the input process, and whose output process coincides with the given one. Unlike in [13], we will not require the input to be white. In particular, we will allow finite-state Markov processes as inputs, which will allow us to apply the framework to the realization of stochastic jump-linear systems. Particular cases of this generalized bilinear realization problem include realization of classical linear and bilinear systems, as well as the Kalman filter. In addition, the solution to this general problem provides a solution to the realization of stochastic jump-linear systems, as we will show in Section V.

The motivation of the realization problem stems from system identification and filtering. The link with system identification is quite clear: the realization problem can be viewed as a idealized system identification problem. The link with filtering is less direct. Recall that filtering one is interested in computing the conditional expectation (or the linear projection) of the current output onto the past outputs. The Kalman filter is an algorithm that computes such a projection recursively. If one considers stationary linear systems, then the Kalman filter yields a linear stochastic realization in the forward innovation form. That is, there is a correspondence between recursive filters and stochastic realizations in forward innovation form.

In the case of bilinear situation, the situation is similar. The main difference is that the filtering occurs based not only on past outputs but on past inputs too. In particular, the correspondence between filters and stochastic realizations carries over to bilinear systems. Similarly to the linear case, the construction of the recursive filter (i.e. stochastic realization in forward innovation form) relies on the fact that the covariances of the outputs can be represented as rational formal power series.

The section is organized as follows. In §III-A we define the class of generalized bilinear systems and the corresponding realization problem. In §III-C we present the solution of the realization problem. In §III-E we present a realization algorithm. The proofs of the results of §III-C–III-E are presented in §IV.

In what follows, we will work with random variables and stochastic processes. We will use the standard terminology and notation of probability theory [46]. Throughout the paper, we fix a probability space $(\Omega, \mathcal{F}, P)$ and all the random variables and stochastic processes should be understood with respect to this probability space. Here $\mathcal{F}$ is a σ-algebra over the set Ω, P is a probability measure on $\mathcal{F}$. With a slight abuse of notation, when we want to indicate that a random variable $\mathbf{z}$ takes its values in a set X (i.e. $\mathbf{z}$ is a measurable function $\mathbf{z} : \Omega \to X$), we will write $\mathbf{z} \in X$. We denote the expectation of a random variable $\mathbf{z}$ by $E[\mathbf{z}]$. Let $\mathbb{Z}$ be the set of integers. Recall that a discrete-time stochastic process (in the sequel to be referred to as process or stochastic process) taking values in a set X is just a collection $\{\mathbf{z}(t)\}_{t\in\mathbb{Z}}$ where $\mathbf{z}(t) \in X$ is a random variable for all $t \in \mathbb{Z}$; $\mathbf{z}(t)$ is referred to as the value of the stochastic process $\{\mathbf{z}(t)\}_{t\in\mathbb{Z}}$ at time $t \in \mathbb{Z}$. In the sequel, by abuse of notation, the stochastic process $\{\mathbf{z}(t)\}_{t\in\mathbb{Z}}$ will be denoted by $\mathbf{z}(t)$: whether $\mathbf{z}(t)$ means a stochastic process or its value at time t will be clear from the context. A stochastic process $\mathbf{z}(t) \in \mathbb{R}^k$ is called zero mean and square integrable, if the expectations $E[\mathbf{z}(t)]$ and $E[\mathbf{z}^T(t)\mathbf{z}(t)]$ exist, and $E[\mathbf{z}(t)] = 0$ and $E[\mathbf{z}^T(t)\mathbf{z}(t)] < +\infty$. Furthermore, recall that a process $\mathbf{z}(t) \in \mathbb{R}^k$ is wide sense stationary, if for every $s, t, k \in \mathbb{Z}$, the expectation $E[\mathbf{z}(t+k)\mathbf{z}^T(s+k)]$ exists and its value is independent of k.

A. *Stochastic Realization Problem for Generalized Bilinear Systems*

Let the *input process* be a collection of $\mathbb{R}$ valued random processes $\{\mathbf{u}_\sigma(t)\}_{\sigma\in\Sigma}$ indexed by the elements of a finite alphabet Σ.

Definition 2 (Generalized Bilinear System): A generalized bilinear system (*abbreviated by* ***GBS***) of is a system of the form

$$\mathrm{B} \begin{cases} \mathbf{x}(t+1) = \sum_{\sigma\in\Sigma} (A_\sigma \mathbf{x}(t) + K_\sigma \mathbf{v}(t))\mathbf{u}_\sigma(t) \\ \\ \quad\quad \mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{v}(t), \end{cases} \tag{8}$$

where $A_\sigma \in \mathbb{R}^n$, $K_\sigma \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, $\mathbf{y}(t)$ is a stochastic process with values in $\mathbb{R}^p$, called the *state process*, $\mathbf{x}(t)$ is a stochastic process with values in $\mathbb{R}^n$, called the *state process* and $\mathbf{v}(t()$ is a stochastic process with values in $\mathbb{R}^m$, called the *noise process*. The *dimension* of B is defined as the number n of state variables. The system B is said to be a *realization* of the process $\widetilde{\mathbf{y}}(t)$ if $\widetilde{\mathbf{y}}(t) = \mathbf{y}(t)$ for all $t \in \mathbb{Z}$. The **GBS** B is said to be a *minimal realization* of $\mathbf{y}(t)$ if B is a realization of $\mathbf{y}(t)$ and it has the minimal dimension among all possible **GBS** of $\mathbf{y}(t)$.

Now we are ready to state the realization problem for **GBS**s.

Definition 3 (Realization problem for generalized bilinear systems): Given an output process $\mathbf{y}(t)$ and find conditions for existence of a **GBS** which is a realization of $\mathbf{y}(t)$ and characterize minimality for **GBS**s which are realizations of $\mathbf{y}(t)$.

Notice that by choosing $\mathbf{u}_\sigma(t)$ in an appropriate way, **GBS**s include linear, bilinear, and as we shall see later, even jump-linear systems.

Example 1 (Realization of Linear Systems): Notice that if $\Sigma = \{\sigma\}$ and $\mathbf{u}_\sigma(t) = 1$, then the generalized bilinear stochastic realization problem reduces to the classical stochastic linear realization problem.

Example 2 (Realization of Bilinear Systems): Notice that if $\Sigma = \{1, 2\}$, $\mathbf{u}_1(t) = 1$ and $\mathbf{u}_2(t)$ is white noise, then the generalized bilinear stochastic realization problem reduces to the classical bilinear realization problem [13], [14].

Example 3 (Linear Jump-Markov systems with i.i.d discrete-state): Assume that $\boldsymbol{\theta}(t) \in \Sigma$ are independent and identically distributed random variables, $P(\boldsymbol{\theta}(t) = \sigma) = p_\sigma > 0$. Consider the generalized bilinear system with $\mathbf{u}_\sigma(t) = \chi(\boldsymbol{\theta} = \sigma)$, where χ is the indicator function. In this case the realization problem for **GBS**s yields the realization of Jump-Markov linear systems where is Markov process is observable and i.i.d. In fact, it can be shown that the realization problem of more general type jump-linear systems can also be reduced to that of **GBS**s.

Example 4 (Stochastic LPV systems): Let $\Sigma = \{1, \ldots, d\}$ and let $\mathbf{u}(t) = (\mathbf{u}_1(t), \ldots, \mathbf{u}_d(t))$ be a stochastic process such that $\mathbf{u}$ and $\mathbf{v}$ are independent. The resulting **GBS** can be viewed as a stochastic linear parameter-varying system (LPV), where $\mathbf{u}$ plays the role of the scheduling variable. LPV systems represent a widely applied and popular system class. Identification of LPV systems is a subject of active research. The results of this paper are potentially useful for system identification of LPV systems.

Example 5 (jump-bilinear systems with i.i.d discrete-state): Let Q be a finite set and fix an integer m. Assume that $\boldsymbol{\theta}(t) \in Q$ are i.i.d random variables, $P(\boldsymbol{\theta}(t) = q) = p_q \geq 0$ for all $q \in Q$. Define $\Sigma = Q \times \{0, \ldots, m\}$ and let $\mathbf{u}(t) \in \mathbb{R}^m$ be a colored noise process. Define $\mathbf{u}_{(q,j)}(t) = \mathbf{u}_j(t)\chi(\boldsymbol{\theta} = q)$, where $\mathbf{u}_j(t)$ denotes the jth entry of $\mathbf{u}(t)$ for $j = 1, \ldots, m$ and $\mathbf{u}_0(t) = 1$. With this choice of the input process, we immediately obtain the following jump-bilinear system $\mathbf{x}(t+1) = \sum_{j=0}^{m}(A_{\boldsymbol{\theta}(t),j}\mathbf{x}(t) + K_{\boldsymbol{\theta}(t),j}\mathbf{v}(t))\mathbf{u}_j(t)$ and $\mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{v}(t)$. That is **GBS**s do not only describe known system classes, but they also yield new system classes.

The examples above are intended to demonstrate the versatility of **GBS**s. **GBS**s can be used not only to describe well known system classes, but also system classes which have not been studied in the literature so far.

B. Hilbert-space of square integrable random variables

In order to make the realization problem tractable, we need to make additional assumption on **GBS**s. In particular, in the sequel, the outputs and inputs at any time instance are mean-square integrable random variables. Such random variables form a Hilbert-space $\mathcal{H}$ with covariance playing the role of scalar product. Since $\mathcal{H}$ is an Hilbert-space, we can speak of orthogonal projection of a random variable onto a closed subspace of $\mathcal{H}$. Below we recall the framework of the Hilbert-space of random variables in more detail.

In the sequel, we will identify random variables which differ only on a set of probability zero. A scalar random variable $\mathbf{z} \in \mathbb{R}$ is said to be mean-square integrable, if the expectation $E[\mathbf{z}^2]$ exists and it is finite. The space of scalar mean-square random variables forms an Hilbert-space $\mathcal{H}$ with the scalar product $< \mathbf{z}, \mathbf{x} >= E[\mathbf{z}\mathbf{x}]$ and the corresponding norm $||\mathbf{z}|| = \sqrt{E[\mathbf{z}^2]}$. A sequence of random variables $\mathbf{z}_n$ is said to converge to in mean-square sense to $\mathbf{z}$, if $\lim_{n\rightarrow\infty} E[(\mathbf{z}-\mathbf{z}_n)^2] = 0$, or, in other words, if $\lim_{n\rightarrow\infty} ||\mathbf{z}_n - \mathbf{z}|| = 0$ with the norm $||.||$ defined above. As it is customary in Hilbert-spaces, the scalar product and the norm are continuous operators with respect to the topology induced by mean-square convergence. That is, if $\lim_{n\rightarrow\infty} \mathbf{z}_n = \mathbf{z}$ and $\lim_{n\rightarrow\infty} \mathbf{x}_n = \mathbf{x}$ in the mean-square sense, then $\lim_{n\rightarrow\infty} E[\mathbf{x}_n\mathbf{z}_n] = E[\mathbf{x}\mathbf{z}]$ and $\lim_{n\rightarrow\infty} ||\mathbf{x}_n|| = ||\mathbf{x}||$.

Suppose that M is a closed linear subset of $\mathcal{H}$. The orthogonal projection of a variable $\mathbf{z}$ onto M the unique element $\mathbf{z}^*$ of M which satisfies the following two equivalent conditions: **(a)** $||\mathbf{z}^* - \mathbf{z}|| \le ||\mathbf{x} - \mathbf{z}||$ for all $\mathbf{x} \in M$, **(a)** $\mathbf{z} - \mathbf{z}^*$ is orthogonal to M, i.e. $E[(\mathbf{z} - \mathbf{z}^*)\mathbf{x}] = 0$ for all $\mathbf{x} \in M$. Note that if M is the linear span of finitely many elements, then it is automatically closed.

Consider now a *vector valued* random variable $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_p)^T \in \mathbb{R}^p$. We will call $\mathbf{z}$ mean-square integrable, if the coordinates $\mathbf{z}_i$, $i = 1, \dots, p$ are mean-square integrable scalar random variables. Note that if we denote by $||.||_2$ the Euclidean norm in $\mathbb{R}^p$, then mean-square integrability of $\mathbf{z}$ is equivalent to existence and finiteness of $E[||z||_2^2]$. If $\mathbf{z}_n = (\mathbf{z}_1^n, \mathbf{z}_2^n, \dots, \mathbf{z}_p^n) \in \mathbb{R}^p$, $n \in \mathbb{N}$ and $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_p) \in \mathbb{R}^p$ are mean-square integrable random variables, then we say that $\mathbf{z}_n$ converges to $\mathbf{z}$ in a mean square sense, if for all $i = 1, \dots, p$, the sequence $\mathbf{z}_i^n \in \mathbb{R}$ of i

coordinates of $\mathbf{z}_n$ converges to the ith coordinate $\mathbf{z}_i \in \mathbb{R}$ of $\mathbf{z}$ in the mean-square sense.

Let M be a closed linear subspace of mean-square integrable *scalar* random variables. Let $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_p) \in \mathbb{R}^p$ be a *vector valued* mean-square integrable random variable. By the orthogonal projection of $\mathbf{z}$ onto M we mean the vector valued random variable $\mathbf{z}^* = (\mathbf{z}_1^*, \ldots, \mathbf{z}_p^*)$ such that $\mathbf{z}_i^* \in M$ is the orthogonal projection of the ith coordinate $\mathbf{z}_i$ of $\mathbf{z}$ onto M, as defined for the scalar case. The orthogonal projection $\mathbf{z}^*$ has the following property: $E[(\mathbf{z} - \mathbf{z}^*)\mathbf{x}] = 0$ for all $\mathbf{x} \in M$. If M is generated by closure of the linear span of the coordinates of a subset S of $\mathbb{R}^k$ valued mean-square integrable random variables, then $\mathbf{z}^*$ is uniquely determined by the following property: $E[(\mathbf{z} - \mathbf{z}^*)\mathbf{x}^T] = 0$ for all $\mathbf{x} \in S$ and all the coordinates of $\mathbf{z}^*$ belong to M.

In fact, by abuse of terminology, we will say *that* $\mathbf{z}$ *belongs to* M, if all its coordinates $\mathbf{z}_1, \ldots, \mathbf{z}_p$ belong to M. Similarly, let $\mathbf{x}_i \in \mathbb{R}^k$, $i \in I$ be a family of vector valued mean-square integrable random variables and assume that I is an arbitrary set. Then the Hilbert-space generated by $\{x_i\}_{i \in I}$ is understood to be the smallest closed subspace M of the Hilbert-space of all square integrable random variables such that for $\mathbf{x}_i$, $i \in I$ belongs to M in the above sense (i.e. the components of $\mathbf{x}_i$ belongs to M).

Assume that $\mathbf{z}$ belongs to M and assume that M is the Hilbert-space generated by the components some vector values variables $\{\mathbf{x}_i\}_{i \in I}$. In the sequel, we will often use the following simple result.

Lemma 1: If the $\mathbb{R}^p$-valued random variable $\mathbf{z}$ belongs to M, then $\mathbf{z}$ is measurable with respect to the σ-algebra generated by $\mathcal{F} = \{\mathbf{x}_i\}_{i \in I}$.

Indeed, by [46, Exercise 34.13], the conditional expectation $E[\mathbf{z} \mid \mathcal{F}]$ equals the orthogonal projection of z to the close subspace $\mathcal{H}_{\mathcal{F}}$ generated by all the $\mathcal{F}$ measurable mean square integrable random variables. But M is a subspace of $\mathcal{H}_{\mathcal{F}}$ and hence $\mathbf{z}$ already belongs to $\mathcal{H}_{\mathcal{F}}$. Hence, the orthogonal projection of z to $\mathcal{H}_{\mathcal{F}}$ equals $\mathbf{z}$ itself. Thus, $\mathbf{z} = E[\mathbf{z} \mid \mathcal{F}]$ and since $E[\mathbf{z} \mid \mathcal{F}]$ is $\mathcal{F}$ measurable by definition, Lemma 1 follows.

C. Solution of the realization problem for ***GBS***

Below we present the solution of the realization problem for **GBS**s. We will only state the results, their proofs will be presented in §IV. In order to state the results, will introduce the following notation and terminology.

Notation 2: We fix a collection $\{p_\sigma > 0\}_{\sigma\in\Sigma}$ of real numbers. For each $w \in \Sigma^*$ define the number p_w as follows: $p_\epsilon = 1$ and if $w = v\sigma$ for some $v \in \Sigma^*$ and $\sigma \in \Sigma$, then let $p_w = p_v p_\sigma$. The roles of $\{p_\sigma\}_{\sigma\in\Sigma}$ will become clear later on. For each word $w = \sigma_1\sigma_2\cdots\sigma_k \in \Sigma^+$, $k \geq 1$, $\sigma_1, \ldots, \sigma_k \in \Sigma$, define the random variables

$$\mathbf{u}_w(t) = \mathbf{u}_{\sigma_1}(t-k+1)\mathbf{u}_{\sigma_2}(t-k+1)\cdots\mathbf{u}_{\sigma_k}(t) \tag{9}$$

Using the notation defined above, we formulate the following assumptions which will be valid for the rest of the section.

Assumption 1 (Input process): 1) $\sum_{\sigma\in\Sigma}\alpha_\sigma\mathbf{u}_\sigma(t) = 1$ for some numbers $\{\alpha_\sigma \in \mathbb{R}\}_{\sigma\in\Sigma}$.

2) For each $w \in \Sigma^+$, all the first and second order moments of the process $\mathbf{u}_w(t)$ are finite.

We mention a number of examples of $\mathbf{u}_\sigma(t)$ which satisfies the assumptions above.

Example 6 (Bilinear systems [13]): $\Sigma = \{0,1\}$, $\mathbf{u}_0(t) = 1$, $\mathbf{u}_1(t)$ is a white noise Gaussian process. In this case, $\alpha_0 = 1$, $\alpha_1 = 0$.

Example 7 (Discrete valued input): Assume there exists a process $\boldsymbol{\theta}(t)$ takes its values from a finite alphabet Σ and let $\mathbf{u}_\sigma(t) = \chi(\boldsymbol{\theta}(t) = \sigma)$. Then $E[|\mathbf{u}_w(t)|^k] = E[\mathbf{u}_w(t)] = P(\boldsymbol{\theta}(t-k) = \sigma_1\cdots\boldsymbol{\theta}(t-1) = \sigma_k)$ and with $\alpha_\sigma = 1$, $\sum_{\sigma\in\Sigma}\mathbf{u}_\sigma(t) = 1$.

Next, we define a class of stochastic processes which will play an important role in the rest of the paper. Let $\mathbf{r}(t) \in \mathbb{R}^k$ be a stochastic process and define for each $w \in \Sigma^+$

$$\mathbf{z}^{\mathbf{r}}_w(t) = \mathbf{r}(t-|w|)\mathbf{u}_w(t-1)\frac{1}{\sqrt{p_w}}. \tag{10}$$

In the sequel, the process $\mathbf{z}^{\mathbf{y}}_w(t)$, obtained from (10) by choosing $\mathbf{r}(t) = \mathbf{y}(t)$ will play a central role. For this reason, we introduce the following notation

Notation 3: In the sequel we denote by $\mathbf{z}_w(t)$ the process $\mathbf{z}^{\mathbf{y}}_w(t)$.

Below, we will define a number of properties of $\mathbf{z}^{\mathbf{r}}_w(t)$ and we will require that the noise, state, and output processes $\mathbf{x}(t)$, $\mathbf{v}(t)$ and $\mathbf{y}(t)$ of a **GBS** are such that $\mathbf{z}^{\mathbf{x}}_w(t)$, $\mathbf{z}^{\mathbf{v}}_w(t)$ and $\mathbf{z}^{\mathbf{y}}_w(t)$ satisfy those properties. Intuitively, these properties say that $\mathbf{z}^{\mathbf{r}}_w(t)$ is a wide-sense stationary stochastic process if w is also viewed as multidimensional time. To this, we introduce the following definitions.

Definition 4 (Admissible words): A set $L \subseteq \Sigma^+$ is a *set of admissible words*, if the following conditions hold.

1) $\Sigma \subseteq L$ and for all $w \in \Sigma^+ \setminus L$, $\mathbf{u}_w(t) = 0$ almost surely.

2) There exists a set $S \subseteq \Sigma \times \Sigma$, such that the word $w = \sigma_1 \cdots \sigma_k \in \Sigma^+$, $\sigma_1, \ldots, \sigma_k \in \Sigma$, $k > 1$ belongs L if and only if $(\sigma_i, \sigma_{i+1}) \in S$ for all $i = 1, \ldots, k-1$.

For the rest of the paper L will denote a fixed set of admissible words. The motivation behind introducing the set L is that for certain $w \in \Sigma^+$, we might wish to set $\mathbf{z}_w(t)$ to zero. This will be the case when we try to use realization theory for **GBS**s for jump-markov systems. A simpler motivating example is presented below.

Example 8 (Jump-markov systems with restricted switching): Consider the system described in Example 3 but with the following modification. We no longer assume that $\boldsymbol{\theta}$ is an i.i.d process. Instead we assume that there exists a set $S \subseteq Q \times Q$ describing the admissible discrete state transitions, and $\boldsymbol{\theta}(t)$ is a stationary Markov process such that $P(\boldsymbol{\theta}(t+1) = q_2 \mid \boldsymbol{\theta}(t) = q_1) = p_{q_2}$ if $(q_1, q_2) \in S$ and $P(\boldsymbol{\theta}(t+1) = q_2 \mid \boldsymbol{\theta}(t) = q_1) = 0$ if $(q_1, q_2) \in S$. In this case, the $\mathbf{u}_w(t) = 0$ almost surely for $w \notin L$, where L is as defined in 4

Definition 5 (Recursive covariance property): A process $\mathbf{r}(t)$ is said to have *recursive covariance property (abbreviated by **RC**)* if it satisfies the following conditions.

1) The processes $(\mathbf{r}(t), \{\mathbf{z}^{\mathbf{r}}_w(t) \mid w \in \Sigma^+\})$ are jointly wide-sense stationary, that is, for all $t, k \in \mathbb{Z}$, and for all $w, v \in \Sigma^+$ we have that $E[\mathbf{r}(t)] = 0$, $E[\mathbf{z}^{\mathbf{r}}_w(t)] = 0$, and

$$E[\mathbf{r}(t+k)(\mathbf{z}^{\mathbf{r}}_w(t+k))^r] = E[\mathbf{r}(t)(\mathbf{z}^{\mathbf{r}}_w(t))^T] \quad \text{and} \quad E[\mathbf{z}^{\mathbf{r}}_w(t+k)(\mathbf{z}^{\mathbf{r}}_v(t+k))^T] = E[\mathbf{z}^{\mathbf{r}}_w(t)(\mathbf{z}^{\mathbf{r}}_v(t))^T].$$

2) Denote by

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

Then for any $w, v \in \Sigma^+$, $\sigma, \sigma^{'} \in \Sigma$, $T_{\sigma,\sigma^{'}} = 0$ for $\sigma \neq \sigma^{'}$ and

$$T^{\mathbf{r}}_{w\sigma,v\sigma^{'}} = \begin{cases} T^{\mathbf{r}}_{w,v} & \text{if } \sigma = \sigma^{'} \text{ and } w\sigma \in L \text{ or } v\sigma \in L \\ 0 & \text{if } \sigma \neq \sigma^{'} \end{cases} \quad \text{and} \tag{11}$$

$$T^{\mathbf{r}}_{w\sigma,\sigma^{'}} = \begin{cases} (\Lambda^{\mathbf{r}}_w)^T & \text{if } \sigma = \sigma^{'} \\ 0 & \text{if } \sigma \neq \sigma^{'} \end{cases}. \tag{12}$$

3) In addition, $T^{\mathbf{r}}_{w,v} = 0$ if $w \notin L$ or $v \notin L$. If $w\sigma \in L$ then for all $v\sigma \notin L$, $T^{\mathbf{r}}_{v,w} = 0$, and similarly, if $v\sigma \in L$, then for all $w\sigma \notin L$, $T^{\mathbf{r}}_{v,w} = 0$.

Remark 1: It can be shown that Part 3 of Definition 5 is by the other conditions.

Remark 2: It is clear that if $\mathbf{r}(t) \in \mathbb{R}^r$ is an **RC** process, then for any matrix $F \in \mathbb{R}^{l \times r}$, $l > 0$, the process $\mathbf{s}(t) = F\mathbf{r}(t)$, $t \in \mathbb{Z}$, is **RC**.

Intuitively, if $\mathbf{r}$ is **RC**, then the processes $\mathbf{z}^{\mathbf{r}}_w$ obtained by multiplying $\mathbf{r}(t)$ with future inputs $\mathbf{u}_w(t+|w|)$ are zero-mean wide-sense stationary, moreover, the covariances $T^{\mathbf{r}}_{w,v}$ have a specific recursive structure. This recursive structure can be interpreted as wide-sense stationarity, if w is viewed as a time instant on the multidimensional time axis Σ^+. This property coincides with the property required of multidimensional positive kernels in [47] and a special instance of this property was also used in [14], [13]. This property (Part 3 of Definition 5) is crucial for developing stochastic realization theory, especially for the realization algorithm.

*Example 9 (Examples of **RC** processes):* Assume that $L = \Sigma^+$, $\mathbf{r}(t)$ is a zero-mean wide-sense stationary process, $\mathbf{r}(t)$ and $\mathbf{u}_\sigma(t+k)$, $k \geq 0$ are independent, $\mathbf{u}_\sigma(t)$ are i.i.d and $E[\mathbf{u}_\sigma^2] = p_\sigma$, and $\{\mathbf{u}_{\sigma_1}(t)\}_{t \in \mathbb{Z}}$, $\{\mathbf{u}_{\sigma_2}(t)\}_{t \in \mathbb{Z}}$ are uncorrelated for all $\sigma_1 \neq \sigma_2$, i.e. $E[\mathbf{u}_{\sigma_1}(t)\mathbf{u}_{\sigma_2}(l)] = 0$, $l, t \in \mathbb{Z}$. Moreover, assume that $\mathbf{u}_\sigma(t)$ satisfies Assumption 1 and that for all $w \in \Sigma^+$, $E[\mathbf{r}(t - |w|)\mathbf{u}_w(t-1)\mathbf{r}^T(t)]$ is independent of t. Then $\mathbf{r}$ is a **RC** process.

One particular examples of the situation is when $\mathbf{u}_\sigma(t)$ is a zero mean i.i.d Gaussian process. Another example if when $\Sigma = \{0, 1\}$, $\mathbf{u}_0(t) = 1$ and $\mathbf{u}_1(t)$ is an i.i.d zero mean Gaussian process with variance p_σ. This latter example is the one which occurs in bilinear stochastic systems. Finally, consider $\mathbf{u}_\sigma(t)$ is as in Example 7. Assume, moreover that $\boldsymbol{\theta}(t)$ are i.i.d $p_\sigma = P(\boldsymbol{\theta}(t) = \sigma)$. Then with $L = \Sigma^+$, $\mathbf{r}(t)$ is an **RC** process.

Although Example 9 covers a lot important cases, the example below demonstrates that RC processes where $\mathbf{u}_\sigma$ is not an i.i.d process also plays an important role.

Example 10: Consider the process $\boldsymbol{\theta}$ from Example 8 and assume that $\{\boldsymbol{\theta}(t+l) \mid l \geq 0\}$ and $\mathbf{r}(t)$ are conditionally independent w.r.t. to $\{\boldsymbol{\theta}(t-l) \mid l \geq 0\}$. Assume that $\mathbf{r}(t)$ is wide-sense stationary, square integrable and zero-mean. Then $\mathbf{r}(t)$ is a **RC** process with L defined in Example 8.

Now we are ready to formulate the assumptions we are going to make about **GBS**s.

Assumption 2: In the sequel, we will only consider **GBS**s which satisfy the following conditions.

1) The noise process $\mathbf{v}(t)$ have the **RC** property.
2) For every $w, v \in \Sigma^+$, $w \neq v$, $\mathbf{z}^{\mathbf{v}}_v(t)$ and $\mathbf{z}^{\mathbf{v}}_w(t)$ are orthogonal, i.e. $E[\mathbf{z}^{\mathbf{v}}_w(t)(\mathbf{z}^{\mathbf{v}}_v(t))^T] = 0$.
3) The state $\mathbf{x}(t)$ belongs to the Hilbert-space generated by the entries of $\{\mathbf{z}^{\mathbf{v}}_w(t) \mid w \in \Sigma^+\}$.

4) The matrix $\Sigma_{\sigma\in\Sigma} p_\sigma A_\sigma^T \otimes A_\sigma^T$ is stable, i.e. all its eigenvalues are inside the unit circle.

5) For all $\sigma_1, \sigma_2 \in \Sigma$, if $\sigma_1\sigma_2 \notin L$, then $A_{\sigma_2}A_{\sigma_1} = 0$ and $A_{\sigma_2}K_{\sigma_1}T^{\mathbf{v}}_{\sigma_1,\sigma_1} = 0$.

Intuitively, Part 1 of Assumption 2 requires that the state and noise process are stationary and that they are very loosely correlated with future inputs. Parts 2–1 of Assumption 2 say that the noise processes are uncorrelated. Parts 3–4 intuitively express the assumption that $\mathbf{x}(t)$ is the result of starting at zero initial state at $-\infty$ and allowing the system to be driven by the noise process alone. The stability assumption is there to guarantee that this can be done. In fact, Assumption 2 yields the following.

Lemma 2: If $\mathcal{B}$ of the form (8) satisfies Assumption 2, then $\left[\mathbf{v}^T(t), \mathbf{x}^T(t)\right]^T$ is an **RC** process, and hence $\mathbf{x}(t)$ is an **RC** process. Moreover, $w, v \in \Sigma^+$, $|w| \geq |v|$, $\mathbf{z}^{\mathbf{x}}_w(t)$ and $\mathbf{z}^{\mathbf{v}}_v(t)$ are uncorrelated, i.e. $E[\mathbf{z}^{\mathbf{x}}_w(t)(\mathbf{z}^{\mathbf{v}}_v(t))^T] = 0$, and

$$\forall t \in \mathbb{Z} : \mathbf{x}(t) = \sum_{w\in\Sigma^*}\sum_{\sigma\in\Sigma} \sqrt{p_{\sigma w}} A_w B_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t). \tag{13}$$

Here we used Notation 1 for the matrix product A_w, $w \in \Sigma^*$ and convergence is understood in the mean-square sense.

In fact, we can also show that under some mild conditions, the trajectories of $\mathcal{B}$ converge to $\mathbf{x}(t)$ as t goes to infinity.

Lemma 3: With the assumptions of Lemma 2, if $\hat{\mathbf{x}}(t)$ is a process which satisfies the first equation of (8) and for all $w, v \in \Sigma^*$, $\sigma_1, \sigma_2 \in \Sigma$, $|w| = |v|$, $T^{\hat{\mathbf{x}}}_{\sigma_1 w, \sigma_2 v} = E[\hat{\mathbf{x}}(0)\hat{\mathbf{x}}^T(0)\mathbf{u}_w(|w|-1)\mathbf{u}_v(|v|-1)]$ is such that $T^{\hat{\mathbf{x}}}_{\sigma_1 w, \sigma_2 v} = 0$ if $\sigma_1 w \neq \sigma_2 v$ and $T^{\hat{\mathbf{x}}}_{\sigma_1 w, \sigma_2 v} = E[\hat{\mathbf{x}}(0)\hat{\mathbf{x}}^T(0)\mathbf{u}^2_\sigma(0)]p_w$ otherwise, then

$$\lim_{t\to\infty} E[||\mathbf{x}(t) - \hat{\mathbf{x}}(t)||^2] = 0.$$

If $\hat{\mathbf{x}}(0)$ is independent of $\mathbf{u}_\sigma(t)$, $t \geq 0$, $\sigma \in \Sigma$ and $E[\mathbf{u}_w(|w|-1)\mathbf{u}_v(|v|-1)] = 0$ for $w \neq v$ and $p_w = E[\mathbf{u}_w(|w|-1)\mathbf{u}_v(|v|-1)] = 0$, then the assumptions of Lemma 3 are satisfied. In particular, the assumptions of Lemma 3 are the standard ones made for the systems described in Examples 1–3. Finally, note that $\mathbf{x}(t)$ is wide-sense stationary and the following holds.

Lemma 4: Consider a **GBS** $\mathcal{B}$ of the form (8) and assume that $\mathcal{B}$ satisfies Assumptions 2. Consider the equation

$$P_\sigma = p_\sigma\Big(\sum_{\sigma_1\in\Sigma, \sigma_1\sigma\in L} A_{\sigma_1}P_{\sigma_1}A^T_{\sigma_1} + K_{\sigma_1}Q_{\sigma_1}K^T_{\sigma_1}\Big) \tag{14}$$

where $Q_\sigma = E[\mathbf{v}(t)\mathbf{v}^T(t)\mathbf{u}_\sigma^2(t)]$ and $\{P_q\}_{q\in Q}$ is a family of matrix-valued indeterminate. Then (14) has a unique solution determined by $P_\sigma = E[\mathbf{x}(t)\mathbf{x}^T(t)\mathbf{u}_\sigma^2(t)]$, $\sigma \in \Sigma$.

The proofs of Lemma 2–Lema 4 require certain technical results, for this reason we postpone them to §IV-A. The Lemma 2 says that the state of Σ is the one which one would obtain by starting the system at zero at $-\infty$. Lemma 3 says that if we pick any initial state which satisfies some mild conditions, then the resulting state trajectory of Σ will converge to the stationary trajectory $\mathbf{x}(t)$. In fact, the existence of the right-hand side of (13) does not require Part 3 of Assumption 2. Hence, Lemma 2 – 3 can be interpreted as stating that if the system Σ satisfies Assumption 2, except Part 3, then it has a state trajectory which satisfies Part 3, moreover any state-trajectory of Σ converges to that particular one. The situation is similar to that of for stable linear systems: asymptotically, a the state-trajectory of a stable linear system is stationary. Finally, Lemma 4 provides a formula for the state covariance as a solution of a Lyapunov-like equation. Note that similar formulas are well-known for the linear [48] and even bilinear case [13], [14]. The formula of Lemma 4 represents a generalization of those well-known results.

We present a number of examples of systems which satisfy Assumptions 2.

Example 11 (Linear systems): A stationary stable Gaussian linear system with the standard assumption can be viewed as a **GBS** which satisfies Assumption 2. In this case, $\Sigma = \{0\}$, $\mathbf{u}_0(t) = 1$, A_0 is stable, $L = \Sigma^+$, $\mathbf{v}(t)$ is an i.i.d process which is Gaussian and zero mean. If we assume that the initial state of the system at time ∞ was zero, then it is easy to see that the resulting **GBS** satisfies Assumption 2.

Example 12 (Bilinear systems): The bilinear systems from [13], [14] satisfy Assumption 2. In that case, $\Sigma = \{0, 1\}$, $\mathbf{u}_0(t) = 1$, $\mathbf{u}_\sigma(t)$ is a white noise Gaussian process, $\mathbf{v}(t)$ is also a white noise Gaussian process, $B_1 = 0$ and the random variables $\mathbf{v}(t)$ and $\mathbf{u}_1(t + l)$, $l \in Z$ are assumed to be independent (the σ-algebra generated by them is independent). Moreover, it is assumed that $\mathbf{x}(t)$ is zero-mean, wide-sense stationary and satisfies (13). In fact in [13], [14] it was not explicitly assumed that $\mathbf{x}(t)$ satisfies (13), but from the discussion after Lemma 2 it follows that this can be assumed without loss of generality. Moreover, the state process of the realization constructed by the algorithm [13], [14] does satisfy the assumptions of Lemma 2.

Example 13 (Jump-linear systems driven by i.i.d.): Consider jump-linear systems driven by an i.i.d process as described in Example 3. In this case, $L = \Sigma^+$. Assume that $\{\boldsymbol{\theta}(t)\}_{t\in Z}$ are independent, identically distributed, $p_\sigma = P(\boldsymbol{\theta}(t) = \sigma) > 0$, $\sigma \in \Sigma$. Assume that the noise process is $\{\mathbf{v}(t)\}_{t\in\mathbb{Z}}$ is independent of $\{\boldsymbol{\theta}(t)\}_{t\in\mathbb{Z}}$ and that $\mathbf{v}(t)$ is a wide-sense colored noise process i.e. $E[\mathbf{v}(t)\mathbf{u}_w(l-1)\mathbf{v}^T(l)] = 0$, $l > t$, $w \in \Sigma^+$, $|w| = l - t + 1$, $E[\mathbf{v}(t)] = 0$, $E[\mathbf{v}(t)\mathbf{v}^T(t)] = Q > 0$. Assume that Part 3, Part 4 and Part 5 of Assumption 2 holds. Then the system satisfies Assumption 2. Note that the assumptions we made are quite mild, they are similar to the ones of [49].

The examples above represent a special case of the following class of **GBS**s.

*Example 14 (**GBS** with independent inputs):* Consider a **GBS**$\mathcal{B}$ such that $\mathbf{v}$ and $\mathbf{u}_\sigma$, $\sigma \in \Sigma$ satisfy Example 9. That is, $\mathbf{u}_\sigma$ is an i.i.d process, $E[\mathbf{u}_\sigma^2(t)] = p_\sigma$, and the σ-algebras generated by $\{\mathbf{v}(t-l)\}_{l=0}^{\infty}$ and $\{\mathbf{u}_\sigma(t+l) \mid \sigma \in \Sigma, l \geq 0\}$ are independent for any t. Assume moreover that $\mathbf{v}$ is a zero mean wide sense stationary process and $E[\mathbf{v}(t-l)\mathbf{v}^T(t)] = 0$, $l > 0$, $w \in \Sigma^+$, $|w| = l$, $t \in \mathbb{Z}$. Let $L = \Sigma^+$ and assume $\sum_{\sigma\in\Sigma} p_\sigma A_\sigma^T \otimes A_\sigma^T$ is a stable matrix. Assume that the state $\mathbf{x}(t)$ is obtained by starting the system in zero initial state at time $-\infty$. Then $\mathcal{B}$ satisfies Assumption 2.

Examples 11–13 represent special cases of Example 14. Example 14 can also be used to obtain bilinear jump-markov systems as described in Example 5. Unfortunately, Example 14 does not cover the case of jump-markov linear systems where the discrete state process is not i.i.d. Below we show that even such cases can be cast into our framework. Here we only present a special class of jump-markov linear systems, the general case is dealt with in §refsect:real.

Example 15 (Jump-markov linear systems with restricted switching): Consider the input process $\mathbf{u}_\sigma$, $\sigma \in \Sigma$ described in Example 8. Consider a **GBS** with this input process, such that the following holds. Denote by $\mathcal{D}_t$ the σ-algebra generated by $\{\boldsymbol{\theta}(l) \mid l < t\}$. Assume that $\mathbf{v}(t)$ is a wide-sense stationary zero mean process, such that $\mathbf{v}(t)$ and $\mathbf{v}(l)$, $l \neq t$ are $\mathbf{v}(t)$ and $\mathbf{v}(l)$, $l \leq t$ are conditionally uncorrelated with respect to the σ-algebra $\mathcal{D}_{l,t-1}$ generated by $\{\boldsymbol{\theta}(t)\}_{t=l}^{t_1-1}$, i.e. $E[\mathbf{v}(t)\mathbf{v}^T(l) \mid \mathcal{D}_{l,t-1}] = 0$. Moreover, assume that the σ-algebras generated by $\{\boldsymbol{\theta}(t+l)\}_{l=0}^{\infty}$ and $\{\mathbf{v}(t-l)\}_{l=0}^{\infty}$ are conditionally independent with respect to $\mathcal{D}_t$. In addition, assume that Part 3, Part 4 and Part 5 of Assumption 2 holds. Then the resulting system will again satisfy Assumption 2. The **GBS**s described above can be thought of as a special class of jump-markov linear systems, where the transition probabilities of the discrete state process are either zero or depend only on the final state.

Next, we state a number of assumptions on the output process $\mathbf{y}(t)$ which will guarantee existence of a **GBS** realization of $\mathbf{y}$. To this end, recall that $\mathbf{z}_w(t)$ denotes the process $\mathbf{z}_w^{\mathbf{y}}(t)$.

When constructing a **GBS** realization of $\mathbf{y}$, we will compute the orthogonal projection of the future outputs onto the Hilbert space formed by the past outputs and inputs. In order to simplify the discussion about orthogonal projections, we will use the following notation.

Notation 4 (Orthogonal projection E_l): Let Z be a set of $\mathbb{R}^p$-valued mean-square integrable random variables. Let $\mathbf{z} \in \mathbb{R}^k$, $k > 0$, be another mean-square integrable random variable. We denote by $E_l[\mathbf{z} \mid Z]$ the orthogonal projection of $\mathbf{z}$ onto the subspace M, where M is the closure of the linear space spanned by the coordinates of the elements of Z.
One can interpret $E_l[\mathbf{z} \mid Z]$ as the best approximation (prediction) of $\mathbf{z}$ in terms of (infinite) linear combination of elements of Z. Next, we define the forward innovation process for $\mathbf{y}$.

Definition 6 (Forward innovation): The *forward innovation process* $\mathbf{e}$ of $\mathbf{y}$ is defined as

$$\mathbf{e}(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w(t) \mid w \in \Sigma^+\}]. \tag{15}$$

That is, the forward innovation is the difference between the predicted output and the actual one, if the prediction is based on linear extrapolation of past outputs. The forward innovation process has all the properties required of the noise of a **GBS**. Below we define a class of **GBS**s where $\mathbf{e}$ is the noise.

*Definition 7 (**GBS** in forward innovation form):* Let $\mathcal{B}$ be **GBS** of the form (8). Then $\mathcal{B}$ is in *forward innovation form*, if $D = I_p$, $\mathbf{v}(t) = \mathbf{e}(t)$ for all $t \in \mathbb{Z}$, and $\mathcal{B}$ satisfies Assumption 2.
That is, if Σ is in forward innovation form, then the noise equals $\mathbf{e}$ and $C\mathbf{x}(t)$ equals the linear projection of $\mathbf{y}(t)$ to the space $\{\mathbf{z}_w(t) \mid w \in \Sigma^+\}$, i.e. $C\mathbf{x}(t)$ is the best linear estimate of $\mathbf{y}(t)$ in terms of $\{\mathbf{z}_w(t) \mid w \in \Sigma^+\}$. Moreover, due to Part 3 of Assumption 3, the state $\mathbf{x}(t)$ of Σ belongs to the Hilber-space generated by the variables $\{\mathbf{z}_w(t) \mid w \in \Sigma^+\}$. Hence, a realization in forward innovation form is its own Kalman-filter, and it can be viewed as a system which is driven by the past outputs and inputs.

As we have mentioned before, for realizability by **GBS**, the covariances of the outputs and inputs should form a rational formal power series. Below, we define these formal power series.

Definition 8 (Family of formal power series $\Psi_{\mathbf{y}}$): For each $j \in I = \{1, \dots, p\}$, $\sigma \in \Sigma$, define the formal power series $S_{(j,\sigma)} \in \mathbb{R}^p \ll \Sigma^* \gg$ as $S_{(j,\sigma)}(w) = (\Lambda^{\mathbf{y}}_{\sigma w})_{.,j}$, where $(\Lambda^{\mathbf{y}}_{w\sigma})_{.,j}$ denotes the jth column of the $p \times p$ covariance matrix $\Lambda^{\mathbf{y}}_{w\sigma} = E[\mathbf{y}(t)\mathbf{z}^T_{w\sigma}(t)]$. Define the family of formal power series

$$\Psi_{\mathbf{y}} = \{S_{(j,\sigma)} \mid j \in I, \sigma \in \Sigma\}. \tag{16}$$

We can now state the following assumptions which guarantee existence of a **GBS** realization.

Assumption 3: The process $\mathbf{y}$ is **RC** and the family of $\Psi_{\mathbf{y}}$ is is square summable and rational.
In addition, we will use the following assumption. Define the random variables $\mathbf{z}^f_w(t)$, $w \in \Sigma^+$,

$$\mathbf{z}^f_w(t) = \mathbf{y}^T(t+|w|)\mathbf{u}_w(t+|w|-1)\frac{1}{\sqrt{p_w}}. \tag{17}$$

Assumption 4: For each $w \in \Sigma^+$, assume that the variable $\mathbf{z}^f_w(t)$ is square integrable.

Remark 3: In many important cases, Assumption 4 is automatically satisfied if $\mathbf{y}$ satisfies Assumption 3. We present below a number such cases.

1) $\mathbf{u}_\sigma(t)$ is essentially bounded for all $\sigma \in \Sigma$, $t \in Z_m$, i.e. there exists a constant $K > 0$ such that $|\mathbf{u}_\sigma(t)| \leq K$ almost everywhere. This is the case when for example $\mathbf{u}_\sigma$ arises from a discrete valued process, as described in Example 7. Then $E[(\mathbf{z}^f_w(t))^T\mathbf{z}^f_w(t)] \leq E[\mathbf{y}^T(t+k)\mathbf{y}(t+k)]K^2\frac{1}{p_w} < +\infty$, $k = |w|$.
2) If $\mathbf{y}(t)$, $\mathbf{u}_w(t)$ have finite fourth order moments, then by Hölders inequality, $E[(\mathbf{z}^f_w(t))^T\mathbf{z}^f_w(t)] \leq (E[(\mathbf{y}^T(t+k)\mathbf{y}(t+k))^2]E[\mathbf{u}^4_w(t+k)])^{1/2} < +\infty$, $k = |w|$. In particular, this assumption was made in [13].

Now we can state the main result on existence of a **GBS** realization.

*Theorem 4 (Stochastic realization of **GBS**s: existence):* Assume that $\mathbf{y}$ satisfies Assumption 4. Then $\mathbf{y}$ has a realization by a **GBS** which satisfies Assumption 2 if and only if $\mathbf{y}$ satisfies Assumption 3. Moreover, if $\mathbf{y}$ has a realization by a **GBS** which satisfies Assumption 2, then it has a realization by a **GBS** in forward innovation form.

Recall that by Remark 3, in many cases Assumption 4 follows from Assumption 2.

Corollary 1: Assume $\mathbf{u}_\sigma(t) = \chi(\boldsymbol{\theta}(t) = \sigma)$ where $\boldsymbol{\theta}(t)$ is a Σ-valued process with $P(\boldsymbol{\theta}(t) = \sigma) = p_\sigma > 0$. Then $\mathbf{y}$ has a realization by a **GBS** which satisfies Assumption 2 if and only if $\mathbf{y}$ satisfies Assumption 3.

Theorem 4 is an easy consequence of Theorem 5 – 6 which will be stated below. Theorem 5 implies that the condition of Theorem 4 is necessary for existence of a realization, and Theorem 6 implies that this condition is sufficient. In order to state Theorems 5 – 6, we need the following definition.

Definition 9 (Full rank process): We will say that $\mathbf{y}$ is a full rank, for each $\sigma \in \Sigma$ the covariance $E[\mathbf{e}(t)\mathbf{e}^T(t)\mathbf{u}_\sigma^2(t)]$ is of rank p, hence strictly positive definite. [1]

Strictly speaking, the concept of a full rank process is not necessary for Theorem 4. However, it plays an important role in formulating a realization algorithm. For this reason, we prefer to state Theorems 5–6 in such a way, that the concept of a full rank process is already used.

Next, we relate **GBS**s and rational representations.

*Definition 10 (Representation associated with **GBS**):* Consider the unique collection of $n \times n$ matrices $\{P_\sigma\}_{\sigma\in\Sigma}$ which satisfy (14). Define, the matrices

$$B_\sigma = \frac{1}{\sqrt{p_\sigma}}(A_\sigma P_\sigma C^T + K_\sigma Q_\sigma D^T). \tag{18}$$

Define the *representation associated with* B as $R_\mathrm{B} = (\mathbb{R}^n, \{\sqrt{p_\sigma}A_\sigma\}_{\sigma\in\Sigma}, B, C)$, where $B = \{B_{\sigma,j} \mid \sigma \in \Sigma, j = 1, \ldots, p\}$ and $B_{\sigma,j}$ denotes the jth column of B_σ.

Theorem 5 (Necessary condition for existence): If B is a realization of $\mathbf{y}$ and B satisfies Assumption 2, then the following holds.

- The process $\mathbf{y}$ is **RC**.
- The representation R_B well defined, stable, and R_B is a representation of $\Psi_\mathbf{y}$.
- $\mathbf{y}$ satisfies Assumption 3.
- If, in addition, for all $\sigma \in \Sigma$, $DE[\mathbf{v}(t)\mathbf{v}^T(t)\mathbf{u}_\sigma^2(t)]D^T > 0$, then $\mathbf{y}$ is full rank.

Remark 4: The definition of R_B implies that it is completely determined by the matrices $(C, D, \{A_\sigma, K_\sigma, Q_\sigma)\}_{\sigma\in\Sigma})$.

The first two statements of Theorem 5 state that if $\mathbf{y}$ has a realization by a **GBS** B, then $\mathbf{y}$ is **RC** and R_B is a stable representation of $\Psi_\mathbf{y}$. The third statement, i.e. that $\mathbf{y}$ satisfies Assumption 3, is an easy corollary of the previous ones and Theorem 3 represent necessary conditions for realizability. Theorem 5 not only shows that Assumption 3 represent a sufficient condition, but it also described how to obtain a stable representation of the family of formal power series $\Psi_\mathbf{y}$. The last statement of Theorem 5 says that under some mild assumptions the output of a **GBS** is full rank. This is important, because it shows that the requirement that $\mathbf{y}$ is a full rank process is not an unnatural one. In turn, this assumption allows us to propose a realization algorithm.

Next, we present the result stating the sufficient condition for existence of a realization.

[1]Note that the concept of a full rank process already has an established definition [48], which is slightly different from the one used in this paper. In the linear case, i.e. when $\Sigma = \{\mathtt{z}\}$ and $\mathbf{u}_z = 1$, the two definitions coincide. Hence, our definition represents a slight abuse of terminology.

Theorem 6 (Sufficient condition for existence): If $\mathbf{y}(t)$ satisfies Assumption 3, then it has a **GBS** B in forward innovation and this **GBS** B can be obtained from a minimal rational representation of $\Psi_{\mathbf{y}}$ as follows.

$$\mathrm{B}\begin{cases} \mathbf{x}(t+1) = \sum_{\sigma\in\Sigma}(\frac{1}{\sqrt{p_\sigma}}A_\sigma \mathbf{x}(t) + K_\sigma \mathbf{e}(t))\mathbf{u}_\sigma(t) \\ \mathbf{y}(t) = C\mathbf{x}(t) + \mathbf{e}(t) \end{cases} \tag{19}$$

where

- $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma\in\Sigma}, B, C)$, $B = \{B_{(i,\sigma)} \in \mathbb{R}^n \mid \sigma \in \Sigma, i = 1, \ldots, p\}$ is a minimal representation of Ψ.
- Let O_R the observability matrix of R. Define the random variable $Y_n(t)$ as

$$Y_n(t) = \begin{bmatrix} \mathbf{z}^f_{v_0}(t) & \ldots & \mathbf{z}^f_{v_{M(n-1)}} \end{bmatrix}^T \tag{20}$$

  where $\mathbf{z}^f_\epsilon(t) = \mathbf{y}^T(t)$ and for all $w \in \Sigma^+$, $\mathbf{z}^f$ is as defined in (24). The variable $Y_n(t)$ can be thought of as the products of future outputs and inputs. Notice that R is observable, hence the matrix O_R is has a left inverse, which we will denote by O_R^{-1}. Then the state $\mathbf{x}(t)$ is define as
- For each $\sigma \in \Sigma$,

$$K_\sigma(p_\sigma T_{\sigma,\sigma} - CP_\sigma C^T) = (B_\sigma\sqrt{p_\sigma} - \frac{1}{\sqrt{p_\sigma}}A_\sigma P_\sigma C^T) \tag{21}$$

  where $P_\sigma = E[\mathbf{x}(t)\mathbf{x}^T(t)\mathbf{u}_\sigma(t)\mathbf{u}_\sigma(t)]$, and

$$B_\sigma = \begin{bmatrix} B_{(1,\sigma)}, & B_{(2,\sigma)}, & \ldots, & B_{(p,\sigma)} \end{bmatrix} \in \mathbb{R}^{n\times p}. \tag{22}$$

If, in addition, $\mathbf{y}$ is a full-rank process, then $(p_\sigma T_{\sigma,\sigma} - CP_\sigma C^T)$ is invertible and

$$K_\sigma = (B_\sigma\sqrt{p_\sigma} - \frac{1}{\sqrt{p_\sigma}}A_\sigma P_\sigma C^T)(p_\sigma T_{\sigma,\sigma} - CP_\sigma C^T)^{-1}. \tag{23}$$

Moreover, the **GBS** B constructed above satisfies Assumption 2.

Remark 5 (Algebraic Ricccati equation): By Theorem 6, if $\mathbf{y}$ is full rank, then the combination of (23) and (14) yields an equation of which $\{P_\sigma\}_{\sigma\in\Sigma}$ is a unique solution. This equation is analogous to the well-known algebraic Riccati equation for linear systems.

Theorem 6 not only gives a sufficient condition for existence of a **GBS** realization, but it serves as a starting point of a realization algorithm. Moreover, it makes the relationship between realization theory and filtering more precise. In particular, Remark 5 and Theorem 6 imply that the data contained in a rational representation of $\Psi_{\mathbf{y}}$ (i.e. of covariances of outputs and inputs) contains all the necessary information for constructing a **GBS** realization of $\mathbf{y}$ in forward innovation form. As it was mentioned before, such a **GBS** can be viewed as recursive filter for computing the best linear estimates of future outputs based on past outputs. Together with Theorem 5 and Theorem 1 it yields an algorithm for computing such a filter from an arbitrary **GBS** realization of $\mathbf{y}$: we first compute the representation R_{B} associated with a **GBS** realization B of $\mathbf{y}$, then we use Theorem 6 to obtain a **GBS** in forward innovation form.

Theorem 5 – 6 imply the following characterization of minimality.

Definition 11 (Minimality): A **GBS** B which satisfies Assumption 2 is said to be a *minimal realization* of $\mathbf{y}(t)$ if it realizes $\mathbf{y}(t)$ and it has the minimal dimension among all possible**GBS** realizations of $\mathbf{y}(t)$ which satisfy Assumption 2.

*Theorem 7 (Minimality of **GBS**s):* Assume B is a **GBS** which satisfies Assumption 2 and which is a realization of $\mathbf{y}$. The **GBS** B is minimal if and only if R_{B} is minimal. If the **GBS**s B_1 and B_2 are both minimal realizations of $\mathbf{y}$ and they both satisfy Assumption 2, then R_{B_1} and R_{B_2} are isomorphic.

Remark 6: The isomorphism of R_{B_1} and R_{B_2} can be directly translated into a relationship between the matrices of B_1 and B_2. If B_1 is of the form (8) and the corresponding matrices of B_2 are $\hat{A}_\sigma$, $\hat{K}_\sigma$ and $\hat{C}$ and $\hat{D}$, then isomorphism of R_{B_1} and R_{B_2} implies that there exists a non-singular matrix $\mathrm{S} \in \mathbb{R}^{n \times n}$ such that $C\mathrm{S}^{-1} = \hat{C}, \forall \sigma \in \Sigma : \mathrm{S}A_\sigma \mathrm{S}^{-1} = \hat{A}_\sigma$. Note that we do not claim that $SK_\sigma = \hat{K}_\sigma$, $\sigma \in \Sigma$ or that $D = \hat{D}$. In fact, in general it will not be true.

Remark 7 (Checking minimality): From Theorem 5 it follows that R_{B} can be computed bases solely on the matrices of B and the covariance of the noise. From Theorem 1 it follows that minimality of R_{B} can be checked effectively, by checking if R_{B} is reachable and observable. Hence, minimality of a **GBS** can be checked effectively, based on the knowledge of the matrices $(C, D, \{A_\sigma, K_\sigma, Q_\sigma)\}_{\sigma \in \Sigma})$

*D. Realization theory for subclasses of **GBS**s*

We have argued before that **GBS**s include a large number of system classes such as linear, bilinear stochastic systems and even jump-markov linear systems. However, the solution of the realization problem for **GBS**s does not directly yield solutions to the realization problems for those system classes. The reason for this is quite obvious: while the necessary conditions remain valid for subclasses of **GBS**s, the sufficient conditions need not remain valid. After all, it could easily happen that even if $\mathbf{y}$ has a realization by a **GBS** belonging to a certain subclass, the realization prescribed by Theorem 6 does not fall into that subclass. Nevertheless, the results obtained for general **GBS**s can be used to solve the realization for the various sub-classes of **GBS**s described above. Below we will discuss this topic in more detail.

We start with specializing the results to **GBS**s described in Example 14. We will call such **GBS**s ***GBS**s with i.i.d. inputs*. We will show that the following conditions are necessary and sufficient for existence of a **GBS** realization with i.i.d inputs.

Assumption 5: 1) $\{\mathbf{y}(t), \mathbf{z}_w(t) \mid w \in \Sigma^+\}$ is zero-mean, wide-sense stationary.

2) The σ-algebras generated by respectively $\{\mathbf{y}(t-l)\}_{l=0}^{\infty}$ and $\{\mathbf{u}_\sigma(t+l)\}_{l=0}^{\infty}$, $\sigma \in \Sigma$ are independent.

3) The family $\Psi_{\mathbf{y}}$ is square summable and rational.

We obtain the following corollary of Theorem 4.

Corollary 2: A process $\mathbf{y}$ has a realization by a **GBS** with i.i.d input if and only if $\mathbf{y}$ satisfies Assumption 5. If $\mathbf{y}$ satisfies Assumption 5, then the **GBS** realization of $\mathbf{y}$ described in Theorem 6 is a **GBS** with i.i.d input.

Indeed, if $\mathbf{y}$ satisfies Assumption 5, then $\mathbf{y}$ is **RC** and hence it satisfies Assumption 3. Hence, by Theorem 6, Assumption 5 implies existence of a **GBS** realization B of $\mathbf{y}$ in forward innovation form. The noise process of this **GBS** is then the innovation process $\mathbf{e}(t)$. By Lemma 1, since the coordinates of $\mathbf{e}(t)$ belong to the Hilbert space generated by $\{\mathbf{y}(t), \mathbf{z}_w(t) \mid w \in \Sigma^+\}$, it is measurable w.r.t to the σ-algebra generated by $\{\mathbf{y}(t-l), \mathbf{u}(t-l-1)\}_{l=0}^{\infty}$. The latter σ-algebra is independent of the σ-algebra generated by $\{\mathbf{u}_\sigma(t+l)\}_{l=0}^{\infty}$, since $\mathbf{u}$ is an i.i.d process and $\mathbf{y}$ satisfies Assumption 5. Hence, the σ-algebras generated by $\{\mathbf{e}(t-l)\}_{l=0}^{\infty}$ and $\{\mathbf{u}_\sigma(t+l)\}_{l=0}^{\infty}$ are independent. Hence, B is a **GBS** with i.i.d inputs. Conversely, if $\mathbf{y}$ has a realization by a **GBS** with i.i.d inputs, then by Theorem 5 $\mathbf{y}$ satisfies Assumption 3. Moreover, since $\mathbf{x}(t)$ and hence $\mathbf{y}(t)$ belongs to the Hilbert-space generated by $\{\mathbf{v}(t), \mathbf{z}_w^{\mathbf{v}}(t) \mid w \in \Sigma^+\}$ and the latter variables are independent of $\mathbf{u}_\sigma(t+l)$, $l \geq 0$, from Lemma 1 it follows that the σ-algebras $\{\mathbf{y}(t-l)\}_{l=0}^{\infty}$ and $\{\mathbf{u}_\sigma(t+l)\}_{l=0}^{\infty}$, $\sigma \in \Sigma$ are independent. Hence, $\mathbf{y}$ satisfies Assumption 5.

Corollary 3: Theorem 7 remains valid if we replace **GBS**s by **GBS**s with i.i.d inputs.

Indeed, from Theorem 7 it follows that a reachable and observable **GBS** with i.i.d inputs is minimal. Conversely, by Theorem 7 the **GBS** realization of $\mathbf{y}$ described by Theorem 6 is minimal, and by Corollary 2 it implies that if a **GBS** with i.i.d inputs which has the minimal dimension among all the **GBS**s with i.i.d. inputs, then it has the smallest possible dimension among all the **GBS**s realizations of $\mathbf{y}$. Hence, minimal **GBS**s with i.i.d inputs are also reachable and observable. Moreover, there is a minimal **GBS** realization of $\mathbf{y}$ with i.i.d inputs. Finally, isomorphism of minimal **GBS** realizations with i.i.d. inputs follows directly from Theorem 7.

Recall the linear systems (Example 11), bilinear stochastic systems (Example 12) and jump-markov linear systems with i.i.d discrete state (Example 13) arise from **GBS** with i.i.d inputs by a specific choice of the input process $\mathbf{u}_\sigma$, σ. If we apply Assumption 5 to the case of linear Gaussian systems, then we obtain the classical results on realization theory of linear systems. Notice that the last part of Assumption 5, when applied to the linear case, reduces to requiring that the power spectrum is stable and rational. If we apply Assumption 5 to bilinear stochastic systems, then we obtain the conditions of [13], [14]. Note that in [13] only the sufficiency of the condition was shown, not the necessity. Furthermore, [14] deals with weak realization (see Definition 12) and it assumes that the output equation does not contain a noise term. If we specialize Corollary 2–3 to jump-markov linear systems with i.i.d state process we obtain the following results. We will call the **GBS** of the type described in Example 13 *jump-markov linear systems with i.i.d switching (abbreviated by JMLSIID)*.

Corollary 4 (Realization of JMLSIID): The process $\mathbf{y}$ has a realization by JMLSIID if and only if the following conditions hold:

1) $\{\mathbf{y}(t), \mathbf{z}_w(t) \mid w \in \Sigma^+\}$ is zero-mean, wide-sense stationary,
2) the σ-algebras generated by $\{\mathbf{y}(t-l)\}_{l=0}^{\infty}$ and $\{\boldsymbol{\theta}(t+l)\}_{l=0}^{\infty}$ are independent,
3) the family $\Psi_{\mathbf{y}}$ is square summable and rational.

If $\mathbf{y}$ satisfies the conditions above, then it has a minimal JMLSIID realization in forward innovation form described in Theorem 6. Moreover, Theorem 7 holds if we replace **GBS**s by JMLSIID.

To the best of our knowledge, Corollary 4 represents a new result. That is, the framework of **GBS**s not only extends existing results on bilinear stochastic systems, but also yields, as a special case, new results on a completely different system class.

Finally, we show how the results above specialize to the case of *jump-markov linear systems with restricted switching (abbreviated by JMLSRS)*, described in Example 15.

Assumption 6: 1) $\{\mathbf{y}(t), \mathbf{z}_w(t) \mid w \in \Sigma^+\}$ is jointly zero-mean, wide-sense stationary,

2) the σ-algebras generated by $\{\mathbf{y}(t-l)\}_{l=0}^{\infty}$ and $\{\boldsymbol{\theta}(t+l)\}_{l=0}^{\infty}$ are conditionally independent w.r.t to the σ-algebra $\mathcal{D}_t$ generated by $\{\boldsymbol{\theta}(t-l-1)\}_{l=0}^{\infty}$
3) The family $\Psi_{\mathbf{y}}$ is square summable and rational.

Corollary 5 (Realization of JMLSRS): The process $\mathbf{y}$ has a realization by a JMLSRS if and only if it satisfies Assumption 6. If $\mathbf{y}$ satisfies Assumption 6, then it has a minimal JMLSRS realization in forward innovation form described in Theorem 6. Moreover, Theorem 7 holds if we replace **GBS**s by JMLSRS.

The proof of this corollary is similar to the proof of Corollary 2. First, if $\mathbf{y}$ has a realization by a JMLSR, then, since a JMLSR is a **GBS** satisfying Assumption 2, $\mathbf{y}$ satisfies Assumption 3. Moreover, $\mathbf{y}(t)$ belongs to the Hilbert-space generated by $\{\mathbf{v}(t), \mathbf{z}_w^{\mathbf{v}}(t) \mid v \in \Sigma^+\}$, where $\mathbf{v}$ is the noise process of a JMLSR realization. From Lemma 1 it then follows that $\mathbf{y}(t)$ is measurable with respect to the σ-algebra generated by $\{\mathbf{v}(t-l), \boldsymbol{\theta}(t-l-1)\}_{l=0}^{\infty}$. By the definition of JMLSRS and the well-known properties of conditional independence, σ-algebras generated by $\{\mathbf{y}(t-l)\}_{l=0}^{\infty}$ and $\{\boldsymbol{\theta}(t+l)\}_{l=0}^{\infty}$ are conditionally independent w.r.t. $\mathcal{D}_t$. This, together with Assumption 3 implies that $\mathbf{y}$ satisfies Assumption 6. Conversely, Assumption 6 implies Assumption 3. Then there exists a minimal **GBS** realization B of $\mathbf{y}$ in forward innovation form. The noise process is then the innovation process $\mathbf{e}$ and $\mathbf{e}(t)$ belongs to the Hilbert-space generated by $\{\mathbf{y}(t), \mathbf{z}_w(t) \mid w \in \Sigma^+\}$. Using Lemma 1 it then follows that $\mathbf{e}(t)$ is measurable w.r.t. to the σ-algebra generated by $\{\mathbf{y}(t-l), \boldsymbol{\theta}(t-l-1)\}_{l=0}^{\infty}$. The latter σ-algebra and the σ-algebra generated by $\{\boldsymbol{\theta}(t+l)\}_{l=0}^{\infty}$ are conditionally independent w.r.t to $\mathcal{D}_t$ by Assumption 6. Hence, the σ-algebras generated by $\{\mathbf{e}(t-l)\}_{l=0}^{\infty}$ and $\{\boldsymbol{\theta}(t+l)\}_{l=0}^{\infty}$ are conditionally independent w.r.t. $\mathcal{D}_t$. That is, B is a JMLSRS and it is a minimal one among all the **GBS** realizations. Hence, if a JMSRS is minimal among all the JMLSRS realizations of $\mathbf{y}$, then it is minimal among all the **GBS** realizations of $\mathbf{y}$. Then the last part of Corollary 5 is a direct consequence of Theorem 7.

The result of Corollary 5 is new, to the best of our knowledge. This result is another proof of versatility of the **GBS** framework.

E. Weak realization and realization algorithms

Below we present a realization algorithm for **GBS**s. We only state the algorithm and the related results, the proofs are presented in §IV. Theorem 6 proposes a procedure for construction a **GBS** realization of $\mathbf{y}$ using the knowledge of $\mathbf{y}$ and $\{\mathbf{u}_\sigma\}_{\sigma \in \Sigma}$. In this construction, the noise and the state processes are constructed explicitly using $\mathbf{y}$ and $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. Unfortunately, this procedure is not effective. In fact, it cannot be made effective, since it presumes the knowledge of stochastic processes $\mathbf{y}$ and $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$. The latter objects cannot be represented by finite number of data points. Note however, that for many application the knowledge of the state or noise process is not required, instead it is sufficient to know the matrices of the **GBS** and covariance of the state process. These matrices can be approximated from finitely many data points. This prompts us to introduce the notion of a *weak realization*.

Definition 12 (Weak realization): A collection $(\{A_\sigma, K_\sigma, P_\sigma, Q_\sigma\}_{\sigma\in\Sigma}, C, D)$, where $A_\sigma, P_\sigma \in \mathbb{R}^{n\times n}$, $K_\sigma \in \mathbb{R}^{n\times m}, Q_\sigma \in \mathbb{R}^{m\times m}$, $\sigma \in \Sigma$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$, is called a *weak realization* of $\mathbf{y}$, if there exists a **GBS** Σ of the form (8), such that Σ is a realization of $\mathbf{y}$ and Σ satisfies Assumption 2 and $E[\mathbf{x}(t)\mathbf{x}^T(t)\mathbf{u}_\sigma^2(t)] = P_\sigma$, $E[\mathbf{v}(t)\mathbf{v}^T(t)\mathbf{u}_\sigma^2(t)] = Q_\sigma$, $\sigma \in \Sigma$. The data $(\{A_\sigma, K_\sigma, P_\sigma\}_{\sigma\in\Sigma}, C, D)$ is said to be a weak realization of $\mathbf{y}$ in forward innovation form, if the **GBS** Σ above is a realization of $\mathbf{y}$ in forward innovation form.

By a slight abuse of notation, we will identify Σ with the data $(\{A_\sigma, K_\sigma, P_\sigma, Q_\sigma\}_{\sigma\in\Sigma}, C, D)$ and write $\Sigma = (\{A_\sigma, K_\sigma, P_\sigma, Q_\sigma\}_{\sigma\in\Sigma}, C, D)$.
That is, a **GBS** Σ is said to be a weak realization of $\mathbf{y}$, if there exists a **GBS** realization of $\mathbf{y}$ with the same matrices, state and noise covariance as those of Σ. It turns out that the construction of Theorem 6 can be used to compute a weak realization of $\mathbf{y}$ from finite data.

As the first step, we construct approximations of the state and noise processes from Theorem 6 based on finitely many random variables. More precisely, we define a sequence of candidate state-variables $\mathbf{x}_N(t)$ and noise variables $\mathbf{e}_N(t)$ as

$$\mathbf{x}_N(t) = E_l[O_R^{-1}(Y_n(t)) \mid \{\mathbf{z}_w(t) \mid w \in \Sigma^N\}]$$

$$\mathbf{e}_N(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w(t) \mid w \in \Sigma^N\}$$

Recall that $\Sigma^N = \{w \in \Sigma^+ \mid |w| \le N\}$. Recall that the original construction of $\mathbf{x}(t)$ and $\mathbf{e}(t)$ the projection of future outputs to the space generated by infinitely many past outputs and inputs. In contrast, $\mathbf{x}_N(t)$ and $\mathbf{e}_N(t)$ determined by projections of future outputs to finitely many past outputs and inputs. Intuitively, $\mathbf{x}_N(t)$ and $\mathbf{e}_N(t)$ are approximations of $\mathbf{x}(t)$ and $\mathbf{e}(t)$ respectively. In fact, the following result holds.

Lemma 5: $\lim_{N\to\infty} \mathbf{x}_N(t) = \mathbf{x}(t)$, $\lim_{N\to\infty} \mathbf{e}_N(t) = \mathbf{e}(t)$, $\lim_{N\to\infty} \mathbf{x}_N(t)\mathbf{u}_\sigma(t) = \mathbf{x}(t)\mathbf{u}_\sigma(t)$, and $\lim_{N\to\infty} \mathbf{e}_N(t)\mathbf{u}_\sigma(t) = \mathbf{e}(t)\mathbf{u}_\sigma(t)$.

It turns out that an analog of (19) holds for $\mathbf{x}_N$.

Lemma 6: There exist $n \times p$ matrices K^N_σ, $\sigma \in \Sigma$ such that

$$\begin{aligned} \mathbf{x}_{N+1}(t+1) &= \sum_{\sigma\in\Sigma}(\frac{1}{\sqrt{p_\sigma}}A_\sigma \mathbf{x}_N(t) + K^N_\sigma \mathbf{e}_N(t))\mathbf{u}_\sigma(t) \\ \mathbf{y}(t) &= C\mathbf{x}_N(t) + \mathbf{e}_N(t). \end{aligned} \tag{24}$$

If $P^N_\sigma = E[\mathbf{x}_N(t)\mathbf{x}_N(t)\mathbf{u}^2_\sigma(t)]$, and $(p_\sigma T_{\sigma,\sigma} - CP^N_\sigma C^T)$ is invertible, then K^N_σ

$$K_N = (\sqrt{p_\sigma}B_\sigma - \frac{1}{\sqrt{p_\sigma}}A_\sigma P^N_\sigma C^T)(p_\sigma T_{\sigma,\sigma} - CP^N_\sigma C^T)^{-1}. \tag{25}$$

In fact, we will show later on that $P_\sigma = \lim_{N\to\infty} P^N_\sigma$ and $K_\sigma = \lim_{N\to\infty} K^N_\sigma$. Hence, if we know P^N_σ and K^N_σ, then Lemma 6 yields an approximation of the weak **GBS** realization described in Theorem 6.

The computation of P^N_σ and K^N_σ requires the knowledge of the random variables $\{\mathbf{z}_w(t) \mid w \in \Sigma^N\}$. In practice, however, one has only data, i.e. samples of the random variables $\{\mathbf{z}_w(t) \mid w \in \Sigma^N\}$. Below we present a formula on approximating P^N_σ (and hence K^N_σ) from such a sample. To this end, notice that $\mathbf{x}_N(t)$ belongs to the space spanned by the entries of $\{\mathbf{z}_w(t) \mid w \in \Sigma^N\}$. Recall that $M(N) = |\Sigma^N|$. and $v_1 \prec v_2 \prec \cdots \prec v_{M(N)}$ is an enumeration of Σ^N based on lexicographic ordering. Then there exists $\alpha^N \in \mathbb{R}^{n\times pM(N)}$, such that

$$\mathbf{x}_N(t) = \alpha^N \mathbf{Z}_N(t), \tag{26}$$

where $\mathbf{Z}_N(t) = \begin{bmatrix}\mathbf{z}^T_{v_1}(t) & \dots & \mathbf{z}^T_{v_{M(N)}}(t)\end{bmatrix}^T \in \mathbb{R}^{pM(N)\times 1}$. If we define

$$T_N = E[\mathbf{Z}_N(t)\mathbf{Z}^T_N(t)] \text{ and } \widetilde{\Lambda}_N = E[O^{-1}_R(Y_n(t+1))\mathbf{Z}^T_N(t)],$$

then by the well-known properties of orthogonal projection, α_N is determined by $\widetilde{\Lambda}_N$ and T_N. In fact, if T_N is invertable, then $\alpha_N = \widetilde{\Lambda}_N T^{-1}_N$. From (26) and the assumption that $\mathbf{y}$ is **RC** it then follows that

$$P^N_\sigma = p_\sigma \alpha_N T_N D \alpha^T_N. \tag{27}$$

where D is a diagonal matrix such that the ith diagonal entry $D_i i$ is 1 if $u_i\sigma \in L$ or it is zero otherwise. It then follows that the knowledge of $\widetilde{\Lambda}_N$ and T_N yields α_N and P^N_σ. Note that $\widetilde{\Lambda}_N$ can be computed from a minimal representation R of $\Psi_\mathbf{y}$ as follows: $\widetilde{\Lambda}_N = \begin{bmatrix}\widetilde{\Lambda}_{u1} & \dots & \widetilde{\Lambda}_{u_{M(N)}}\end{bmatrix}$, where $\widetilde{\Lambda}_{\sigma v} = A_v B_\sigma$ with $B_\sigma = \begin{bmatrix}B_{1,\sigma} & \dots & B_{p,\sigma}\end{bmatrix}$, for all $v \in \Sigma^*$, $\sigma \in \Sigma$.

The discussion above yields the realization algorithm presented in Algorithm 2. In Algorithm 2 we assume that we measure the finite time series $\{y(t), u_\sigma(t) \mid \sigma \in \Sigma, t = 0, \ldots, N+M\}$ for some $N, M \geq 0$ and that we have a (n,n)-selection (α, β) at our disposal.

Theorem 8 (Correctness of Algorithm 1): Assume that the following holds:

1) The process $(\mathbf{y}, \{\mathbf{u}_w \mid w \in \Sigma^+\})$ is ergodic and the time series $\{y(t), u_\sigma(t) \mid \sigma \in \Sigma, t = 0, 1, \ldots, \}$ are such that for all $v, w \in \Sigma^+$.

$$\begin{aligned} E[\mathbf{y}(t)\mathbf{z}_w^T(t)] &= \lim_{N\to\infty} \frac{1}{N} \sum_{r=|w|}^{N} y(r) z_w^T(r) \\ E[\mathbf{z}_v(t)\mathbf{z}_w^T(t)] &= \lim_{N\to\infty} \frac{1}{N} \sum_{r=\max|w|,|v|}^{N} z_v(r) z_w^T(r) \end{aligned} \tag{28}$$

2) The n,n–selection (α, β) is such that $\operatorname{rank} H_{\Psi_{\mathbf{y}},\alpha,\beta} = \operatorname{rank} H_{\Psi_{\mathbf{y}}} \leq n$.
3) The representation returned by Algorithm 1 when applied to $H_{\Psi_{\mathbf{y}},n,n+1}$ and (α, β) is of the form $R = (\mathbb{R}^r, \{A_\sigma\}_{\sigma\in\Sigma}, B, C)$.
4) The process $\mathbf{y}$ satisfies Assumption 3, Assumption 4 and it is full rank.

Let Σ be the **GBS** realization of $\mathbf{y}$ from (19) and let $Q_\sigma = E[\mathbf{e}(t)\mathbf{e}^T(t)\mathbf{u}_\sigma^2$. Identify Σ with the corresponding weak realization $\Sigma = (\{A_\sigma, K_\sigma, P_\sigma, Q_\sigma\}_{\sigma\in\Sigma}, C, I_p)$. Then the following holds

1) For large enough N, M, $T_{N,M}$ and $Q_\sigma^{N,M}$ are invertable and Algorithm 1 is well posed.
2) $\lim_{M\to\infty} \Sigma_{N,M} = (\{A_\sigma, K_\sigma^N, P_\sigma^N, Q_\sigma^N\}_{\sigma\in\Sigma}, C, I_p)$, where $Q_\sigma^N = E[\mathbf{e}_N(t)\mathbf{e}_N(t)\mathbf{u}_\sigma^2(t)]$ and P_σ^N and K_σ^N are defined as Lemma 6.
3) $\lim_{N\to\infty} \lim_{M\to\infty} \Sigma_{N,M} = \Sigma$

Informally, Theorem 8 says the following. If we let M go to infinity, then the weak realization $\Sigma_{N,M}$ returned by Algorithm 8 corresponds to the approximate realization described in Lemma 6. In that realization, the state process $\mathbf{x}(t)$ is approximated by $\mathbf{x}_N(t)$, the latter being the (linear combination of) projection of future outputs to finitely many past outputs and inputs. If we let N go to infinity too, then $\Sigma_{N,M}$ will converge (as a tuple of matrices) to the weak realization which corresponds to the **GBS** described in Theorem 6. Theorem 8 and Algorithm 2 open up the possibility of formulating subspace-like realization algorithms for **GBS**s and for analyzing existing ones [50], [18], [15], [16]. Pursuing this direction remains future work.

Algorithm 2

Input: data $\{y_t, u_\sigma(t) \mid t = 0, \ldots, N+M, \sigma \in \Sigma\}$ and $(n.n)$–selection (α, β).

Output: weak realization $\Sigma_{N,M} = (\{{}^M F_\sigma, K_\sigma^{N,M}, P_\sigma^{N,M}, Q_\sigma^{N,M}\}_{\sigma\in\Sigma}, {}^M H, I_p)$.

1: Approximate the covariances Λ_w, $w \in \Sigma^{2n-1}$, and the covariances T_{v_1,v_2} for $v_1, v_2 \in \Sigma^N$ from the time-series using the formula:

$$\Lambda_w \approx \Lambda_w^M \stackrel{\text{def}}{=} \frac{1}{M} \sum_{t=2n-1}^{M+2n-1} y(t) z_w(t)$$

$$T_{v_1,v_2} \approx T_{v_1,v_2}^M \stackrel{\text{def}}{=} \frac{1}{M} \sum_{t=N}^{N+M+1} z_{v_1}(t) z_{v_2}^T(t)$$

where for any $w = \sigma_1 \cdots \sigma_k \in \Sigma^{2n-1}$, $\sigma_1, \ldots, \sigma_k \in \Sigma$, $k \geq 2n-1$, $z_w(t) = y(t-k)u_{\sigma_1}(t-k) \cdots u_{\sigma_k}(t-1)$.

2: Construct the finite Hankel matrix $H_{\Psi_{\mathbf{y}},n+1,n}^M$ by replacing the covariances $\Lambda_w^{\mathbf{y}}$, $w \in \Sigma^{2n-1}$ in the definition of $H_{\Psi_{\mathbf{y}},n,n+1}$ by the estimates Λ_w^M, $w \in \Sigma^{2n-1}$.

3: Choose a n,n-selection (α, β) such that $\operatorname{rank} H_{\Psi_{\mathbf{y}},\alpha,\beta} = \operatorname{rank} H_{\Psi,N,N}$. Apply Algorithm 1 Section II to $H_{\Psi_{\mathbf{y}},n+1,n}^M$ and the n,n-selection (α, β) to obtain a representation $R_M = (\mathbb{R}^n, \{{}^M F_\sigma\}_{\sigma\in\Sigma}, {}^M G, {}^M H)$.

4: Use the estimates T_{v_1,v_2}^M, $v_1, v_2 \in \Sigma^N$ to construct the matrix $T_{N,M}$: the matrix $T_{N,M}$ has the same structure as T_N, but instead of the covariances T_{v_1,v_2} we use the approximations T_{v_1,v_2}^M.

5: Define $\widetilde{\Lambda}_{N,M}$ in the same way $\widetilde{\Lambda}_N$, but using ${}^M F_v {}^M G_\sigma$ instead of $\widetilde{\Lambda}_{\sigma v}$, where ${}^M G_\sigma = \begin{bmatrix} {}^M G_{1,\sigma} & \ldots & {}^M G_{p,\sigma} \end{bmatrix}$.

6: Assume that $T_{N,M}$ is invertable, and find

$$\begin{aligned}
\alpha_{N,M} &= \widetilde{\Lambda}_{N,M} T_{N,M}^{-1} \\
P_\sigma^{N,M} &= p_\sigma \alpha_{N,M} T_{N,M} D \alpha_{N,M}^T \\
Q_\sigma^{N,M} &= p_\sigma (T_{\sigma,\sigma}^M - \frac{1}{p_\sigma} {}^M H P_\sigma^{N,M M} H^T) \\
K_\sigma^{N,M} &= (\sqrt{p_\sigma} {}^M G_\sigma - \frac{1}{\sqrt{p_\sigma}} {}^M F_\sigma P_\sigma^{N,M M} H^T)(p_\sigma T_{\sigma,\sigma}^M - {}^M H P_\sigma^{N,M M} H^T)^{-1}
\end{aligned}$$

Here D is a diagonal matrix such that the ith diagonal entry D_{ii} is 1 if $v_i \sigma \in L$ or it is zero otherwise.

7: Return the weak realization $\Sigma_{N,M} = (\{{}^M F_\sigma, K_\sigma^{N,M}, P_\sigma^{N,M}, Q_\sigma^{N,M}\}_{\sigma\in\Sigma}, {}^M H, I_p)$.

IV. Proof of the results on realization theory of GBSs

A. Technical preliminaries and the proofs of Lemma 2–4

Below we will present a number of technical results on **RC** processes. These results will allow us to prove Lemmas 2–4 and the main theorems.

Notation 5: Let I_k denote the $k \times k$ identity matrix.

Let $\mathbf{r}(t) \in \mathbb{R}^r$ be an **RC** process.

Notation 6: Denote by $H_t^{\mathbf{r}}$ the Hilbert-space generated by the entries of $\{\mathbf{z}_w^{\mathbf{r}}(t) \mid w \in \Sigma^+\}$.

Lemma 7: With the notation above, if $H_t^{\mathbf{r}} \subseteq H_{t+1}^{\mathbf{r}}$, $\mathbf{r}(t) \in H_{t+1}^{\mathbf{r}}$.

Proof of Lemma 7: From Assumption 1 it follows that $\sum_{\sigma\in\Sigma} \alpha_\sigma \mathbf{u}_\sigma(t) = 1$ for any $t \in Z$, and hence $\mathbf{r}(t) = \sum_{\sigma\in\Sigma} \alpha_\sigma \mathbf{r}(t)\mathbf{u}_\sigma(t) = \sum_{\sigma\in\Sigma} \mathbf{z}_\sigma^{\mathbf{r}}(t+1) \in H_{t+1}^{\mathbf{r}}$. Similarly, $\mathbf{z}_w^{\mathbf{r}}(t) = \sum_{\sigma\in\Sigma} \alpha_\sigma \mathbf{z}_w^{\mathbf{r}}(t)\mathbf{u}_\sigma(t) = \sum_{\sigma\in\Sigma} \alpha_\sigma \mathbf{z}_{w\sigma}^{\mathbf{r}}(t+1) \in H_{t+1}^{\mathbf{r}}$. ■

Lemma 8: Let $\mathbf{z}(t) \in \mathbb{R}^d$ be a process such that the entries of $\mathbf{z}(t)$ belong to $H_t^{\mathbf{r}}$ for any $t \in Z$ and that $E[\mathbf{z}(t+k)(\mathbf{z}_w^{\mathbf{r}}(t+k))^T] = E[\mathbf{z}(t)(\mathbf{z}_w^{\mathbf{r}}(t))^T]$. Then the process $\begin{bmatrix} \mathbf{r}(t) \\ \mathbf{z}(t) \end{bmatrix}$ is **RC**.

For the proof of this lemma we will need the following results.

Lemma 9: If $\mathbf{z} \in \mathbb{R}$ is a mean-square integrable random variable and it belongs to the linear span of the components of $\mathbf{z}_v(t)$, $v \in L$, then $E[\mathbf{z}^2\mathbf{u}_\sigma^2(t)] \leq p_\sigma E[\mathbf{z}^2]$.

Proof of Lemma 9: Assume that for some finite subset $S \subseteq L$, $\mathbf{z} = \sum_{v\in S} \alpha_v \mathbf{z}_v(t)$ for some $\alpha_v \in \mathbb{R}^{1\times p}$. Define $S_1 = \{v \in S, v\sigma \in L\}$, $S_2 = \{v \in S, v\sigma \notin L\}$. Then, by noticing that $E[\mathbf{z}_v(t)\mathbf{z}_w^T(t)\mathbf{u}_\sigma^2(t)] = E[\mathbf{z}_{v\sigma}(t+1)\mathbf{z}_{w\sigma}^T(t+1)] = T_{v\sigma,w\sigma}$ and taking into account Part 3 of Definition 5 and that $T_{v\sigma,s}^T = T_{s,v\sigma} = 0$ for $v\sigma \notin L$, we obtain

$$E[\mathbf{z}^2\mathbf{u}_\sigma^2(t)] = \sum_{v,w\in S_1} p_\sigma \alpha_v^T T_{v\sigma,w\sigma} \alpha_w^T = p_\sigma \sum_{v,w\in S_1} \alpha_v T_{v,w} \alpha_w^T \tag{29}$$

On the other hand, by Part 3 of Definition 5, if $v \in S_1$ and $w \in S_2$ or other way around, then $T_{v,w} = 0$. Moreover, $(T_{v,w})_{v,w\in S_2}$ is positive definite, i.e. $\sum_{v,w\in S_2} \alpha_v T_{v,w} \alpha_w^T \geq 0$. Hence, by noticing that $S = S_1 \cup S_2$,

$$E[\mathbf{z}^2] = \sum_{v,w\in S} \alpha_v T_{w,v} \alpha_w^T = \sum_{v,w\in S_1} \alpha_v T_{v,w} \alpha_w^T + \sum_{v,w\in S_2} \alpha_v T_{v,w} \alpha_w^T \geq \sum_{v,w\in S_1} \alpha_v T_{v,w} \alpha_w^T \tag{30}$$

Combining (29) and (30) yields the statement of the lemma. ■

Lemma 10: Assume that $\mathbf{z}_N \in \mathbb{R}$ is a sequence such that $\mathbf{z}_N$ is a finite linear combination of $\mathbf{z}^{\mathbf{r}}_w(t)$, $w \in \Sigma^+$ and $\mathbf{z} = \lim_{N\to\infty} \mathbf{z}_N$ in the mean-square sense. Then for each $\sigma \in \Sigma$, $\mathbf{z}\mathbf{u}_\sigma(t) = \lim_{N\to\infty} \mathbf{z}_N \mathbf{u}_\sigma(t)$. in the mean-square sense.

Proof of Lemma 10: If $\mathbf{z} = \lim_{N\to\infty} \mathbf{z}_N$ in the mean-square sense, then it $\mathbf{z}_N\mathbf{u}_\sigma(t)$ converges to $\mathbf{z}_N\mathbf{u}_\sigma(t)$ in mean sense. Indeed, from Hölders inequality it follows that $E[|\mathbf{z}_N\mathbf{u}_\sigma(t) - z\mathbf{u}_{\sigma(t)}|] = E[|(\mathbf{z}_N - \mathbf{z})||\mathbf{u}_\sigma(t)|] \le \sqrt{E[|z-\mathbf{z}_N|^2]}\sqrt{E[\mathbf{u}_\sigma(t)^2]}$. On the other hand, it can be shown that $\mathbf{z}_N\mathbf{u}_\sigma(t)$ is a Cauchy-sequence in the mean-square sense. Notice that by Lemma 9, $\mathbf{z}_N\mathbf{u}_\sigma(t)$ is in fact mean-square integrable. Consider $\mathbf{z}_{N+K} - \mathbf{z}_N$ for any $K > 0$. Since $\mathbf{z}_{N+K} - \mathbf{z}_N$ belongs to the closed linear space M_{N+K} generated by the entries of $\{\mathbf{z}_k\}_{k\le N}$, by Lemma 9, $E[|\mathbf{z}_{N+K}\mathbf{u}_\sigma(t) - \mathbf{z}_N\mathbf{u}_\sigma(t)|^2] = E[|\mathbf{z}_{N+K} - \mathbf{z}_N|^2\mathbf{u}^2_\sigma(t)] \le p_\sigma E[|\mathbf{z}_{N+K} - \mathbf{z}_N|^2]$. Since $\mathbf{z}_N$ is convergent, it is then a Cauchy sequence and hence by the inequality above so is $\mathbf{z}_N\mathbf{u}_\sigma(t)$. But by Jensen's inequality, $E[|\mathbf{z}_N(t)\mathbf{u}_\sigma(t) - h|] \le \sqrt{E[|z_N(t)\mathbf{u}_\sigma(t) - h|^2]}$, and hence h is the limit of $\mathbf{z}_N\mathbf{u}_\sigma(t)$ in the mean sense as well. It then follows from the uniqueness of the limit in L_1 sense that $h = \mathbf{z}\mathbf{u}_\sigma(t)$ almost surely. Hence, $\mathbf{z}\mathbf{u}_\sigma(t) = h$ is indeed the limit of $\mathbf{z}_N\mathbf{u}_\sigma(t)$ in the mean-square sense. ■

Proof of Lemma 8: From $\mathbf{z}(t) \in H_t$ it follows that $\mathbf{z}(t) = \lim_{N\to\infty} \mathbf{z}_N$ where $\mathbf{z}_N = \sum_{w\in\Sigma^+,|w|\le N} \alpha_w \mathbf{z}^{\mathbf{r}}_w(t)$ for some $\alpha_w \in \mathbb{R}^{d\times r}$. Define $\mathbf{z}_N(k) = \sum_{s\in\Sigma^+,|s|\le N} \alpha_s \mathbf{z}^{\mathbf{r}}_s(k)$ for all $k \in \mathbb{Z}$. From $E[\mathbf{z}(t)\mathbf{z}^T_N(t)] = E[\mathbf{z}(t+k)\mathbf{z}^T_N(t+k)]$, it follows that $E[||\mathbf{z}(t+k) - \mathbf{z}_N(t+k)||^2] = E[||\mathbf{z}(t) - \mathbf{z}_N(t)||^2]$ and hence $\mathbf{z}(k) = \lim_{N\to\infty} \mathbf{z}_N(k)$, $t,k \in \mathbb{Z}$. For every $v \in \Sigma^+$, denote by $\mathbf{z}^v_N(t)$ the finite sum $\mathbf{z}^v_N(t) = \sum_{s\in\Sigma^+,|s|\le N} \alpha_s \mathbf{z}^{\mathbf{r}}_{sv}(t)$. By repeated application of Lemma 10 we obtain that

$$\mathbf{z}^{\mathbf{z}}_v(t) = \lim_{N\to\infty} \mathbf{z}^v_N(t)$$

Since $E[\mathbf{z}^{\mathbf{z}}_v(t)(\mathbf{z}^{\mathbf{z}}_w(t))^T]$ and $E[\mathbf{z}^{\mathbf{z}}(t)(\mathbf{z}^{\mathbf{z}}_w(t))^T]$ are the limits of $E[\mathbf{z}^v_N(t)(\mathbf{z}^w_N(t))^T]$ and $E[\mathbf{z}_N(t)(\mathbf{z}^w_N(t))^T]$ respectively. Hence, if $\mathbf{r}(t)$ satisfies Part 1 of Definition 5, i.e. $\{\mathbf{r}(t), \mathbf{z}^{\mathbf{r}}_w(t) \mid w \in \Sigma^+\}$ is zero mean wide-sense stationary, then so is $\{\mathbf{z}(t), \mathbf{z}^{\mathbf{z}}_w(t) \mid w \in \Sigma^+\}$. That is $\mathbf{z}$ satisfies Part 1 of Definition 5. Finally, in order to prove that $\mathbf{z}(t)$ satisfies Part 2 of Definition 5, notice that $T^{\mathbf{z}}_{w\sigma,v\sigma'}$ is is the limit of linear combinations of $T^{\mathbf{r}}_{sw\sigma,hv\sigma'}$ for $s,h \in \Sigma^+$. If $\sigma \ne \sigma'$, then by virtue of $\mathbf{r}$ satifying Part 2 of Definition 5, $T^{\mathbf{r}}_{w\sigma,v\sigma'} = 0$. If $\sigma = \sigma'$ and $w\sigma, v\sigma \in L$, Finally, if $w\sigma \notin L$ (respectively $v\sigma \notin L$), then for all $s \in \Sigma^+$, $sw\sigma \notin L$ (respectively $hv\sigma \notin L$ for all $v \in \Sigma^+$) and hence $T^{\mathbf{r}}_{sw\sigma,hv\sigma} = 0$. By combining the results above and taking limits we readily conclude that

$\mathbf{z}(t)$ satisfies Part 2 of Definition 5. Finally, as it was remarked in Remark 1, Part 3 of Definition 5 follows from Parts 1–2 of Definition 5. ■

Notation 7: For every $w \in \Sigma^+$, denote by $H^{\mathbf{r}}_{t,w}$ the Hilbert-space generated by the entries of $\{\mathbf{z}^{\mathbf{r}}_{vw}(t) \mid v \in \Sigma^+\}$ and denote $H^{\mathbf{r},*}_{t,w}$ the Hilbert-space generated by the entries of $\{\mathbf{z}_{vw}(t)^{\mathbf{r}} \mid v \in \Sigma^*\}$. Clearly, $H^{\mathbf{r}}_{t,w} \subseteq H^{\mathbf{r},*}_{t,w}$.

Lemma 11: With the notation above, for every $\sigma_1, \sigma_2 \in \Sigma$, $\sigma_1 \neq \sigma_2$, $H^{\mathbf{r},*}_{t,\sigma_1}$ and $H^{\mathbf{r},*}_{t,\sigma_2}$ are orthogonal and hence $H^{\mathbf{r}}_{t,\sigma_1}$ and $H^{\mathbf{r}}_{t,\sigma_2}$ are orthogonal. Moreover, if $\mathbf{z} \in H^{\mathbf{r}}_t$, then $\mathbf{z}\mathbf{u}_\sigma(t) \in H^{\mathbf{r}}_{t+1,\sigma}$.

Proof of Lemma 11: The first statement of the lemma is an immediate consequence of the fact that $E[\mathbf{z}^{\mathbf{r}}_{v\sigma_1}(t)(\mathbf{z}^{\mathbf{r}}_{w\sigma_2}(t))^T] = 0$ for all $w, v \in \Sigma^+$, $\sigma_1 \neq \sigma_2 \in \Sigma$. The second statement follows by noticing that $\mathbf{z}^{\mathbf{r}}_w(t)\mathbf{u}_\sigma(t) \in H^{\mathbf{r}}_{t+1,\sigma}$. If $\mathbf{z} \in H^{\mathbf{r}}_t$, then $\mathbf{z} = \lim_{N\to\infty} r_N$, where r_N is a finite linear combination of $\mathbf{z}^{\mathbf{r}}_w(t)$, $w \in \Sigma^+$. It then follows that $r_N\mathbf{u}_\sigma(t) \in H^{\mathbf{r}}_{t+1,\sigma}$. From Lemma 10 it follows that $\mathbf{z} = \lim_{N\to\infty} r_N\mathbf{u}_\sigma(t)$ and hence $\mathbf{z} \in H^{\mathbf{r}}_{t+1,\sigma}$. ■

Lemma 12: Let $\mathbf{h}(t) \in \mathbb{R}^l$, $\mathbf{z}(t) \in \mathbb{R}^p$ be processes such that $\mathbf{s}(t) = (\mathbf{z}^T(t), \mathbf{h}^T(t))^T$ is **RC** and the coordinates of $\mathbf{z}(t)$ are orthogonal to $H^{\mathbf{h}}_t$ for all $t \in \mathbb{Z}$. Then for all $w \in \Sigma^+$, the coordinates of $\mathbf{z}^{\mathbf{z}}_w(t)$ are orthogonal to $H^{\mathbf{h}}_{t,w}$ for all $t \in \mathbb{Z}$.

Proof: It then follows that $\mathbf{z}(t) = C_1\mathbf{s}(t)$ and $\mathbf{h}(t) = C_1\mathbf{s}(t)$ for suitable matrices C_1, C_2. Note $E[\mathbf{z}(t)(\mathbf{z}^{\mathbf{h}}_v(t))^T] = C_1\Lambda^{\mathbf{s}}_v C_2^T$ and $E[\mathbf{z}^{\mathbf{z}}_w(t)(\mathbf{z}^{\mathbf{h}}_{vw}(t))^T] = C_1 T^{\mathbf{s}}_{w,vw} C_2^T$, $t \in \mathbb{Z}$. Since $\mathbf{s}(t)$ is **RC**, $T^{\mathbf{s}}_{w,vw} = \Lambda^{\mathbf{s}}_v$ if $vw \in L$ and $T^{\mathbf{s}}_{w,vw} = 0$ otherwise. Hence, $E[\mathbf{z}^{\mathbf{z}}_w(t)(\mathbf{z}^{\mathbf{h}}_{vw}(t))^T] = E[\mathbf{z}(t)(\mathbf{z}^{\mathbf{h}}_v(t))^T]$ if $vw \in L$ and $E[\mathbf{z}^{\mathbf{z}}_w(t)(\mathbf{z}^{\mathbf{h}}_{vw}(t))^T] = 0$ otherwise. Since by the orthogonality assumption $E[\mathbf{z}(t)(\mathbf{z}^{\mathbf{h}}_v(t))^T] = 0$, it then follows that $E[\mathbf{z}^{\mathbf{z}}_w(t)(\mathbf{z}^{\mathbf{h}}_{vw}(t))] = 0$ for all $v \in \Sigma^+$. ■

Proof of Lemma 2: It is clear that if $\mathbf{s}(t) = \begin{bmatrix}\mathbf{v}^T(t), & \mathbf{x}^T(t)\end{bmatrix}^T$ is **RC**, then $\mathbf{x}(t) = \begin{bmatrix}0, & I_n\end{bmatrix}\mathbf{s}(t)$ is **RC** too. The claim that $\begin{bmatrix}\mathbf{v}^T(t), & \mathbf{x}^T(t)\end{bmatrix}^T$ is **RC** follows directly from Part 3, Assumption 2, and Lemma 8, if we can show that $E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{v}}_v(t))^T]$ does not depend on t for any $v \in \Sigma^+$. For any $k \geq 0$,

$$\mathbf{x}(t) = \sum_{w\in\Sigma^+, |w|=k} \sqrt{p_w} A_w \mathbf{z}^{\mathbf{x}}_w(t) + \sum_{w\in\Sigma^*, |w|\leq k-1} \sum_{\sigma\in\Sigma} \sqrt{p_{\sigma w}} A_w B_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t). \tag{31}$$

If $k = |v|$, then it follows that $E[\mathbf{z}^{\mathbf{x}}_w(t)(\mathbf{z}^{\mathbf{v}}_v(t)^T] = 0$. Indeed, by Part 3, Assumption 2 and Lemma 10, $\mathbf{z}^{\mathbf{x}}_w(t)$ belongs to the Hilbert-space generated by the components of $\mathbf{z}^{\mathbf{v}}_{sw}(t)$, $s \in \Sigma^+$. Since, $|w| = |v| = k$, $sw \neq v$ for all $s \in \Sigma^+$. Hence, $\mathbf{z}^{\mathbf{v}}_v(t)$ is orthogonal to the latter Hilbert-space.

Hence, for $k = |v|$,

$$E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{v}}_v(t))^T] = \sum_{w \in \Sigma^*, |w| \le k-1} \sum_{\sigma \in \Sigma} \sqrt{p_{\sigma w}} A_w B_\sigma T^{\mathbf{v}}_{\sigma w, v},$$

and the latter expression does not depend on t.

Using Part 3, Assumption 2 and Lemma 11, the coordinates of $\mathbf{z}^{\mathbf{x}}_w(t)$ belong to the Hilbert-space $H^{\mathbf{v}}_{t,w}$ generated by the coordinates of $\mathbf{z}^{\mathbf{v}}_{sw}(t)$, $s \in \Sigma^+$. Since, $|v| \le |w| = k$, $sw \ne v$ for all $s \in \Sigma^+$, and hence $E[\mathbf{z}^{\mathbf{v}}_{sw}(t)(\mathbf{z}^{\mathbf{v}}_v(t))^T] = 0$. That is, the coordinates of $\mathbf{z}^{\mathbf{v}}_v(t)$ are orthogonal to $H^{\mathbf{v}}_{t,w}$ and hence to $\mathbf{z}^{\mathbf{x}}_w(t)$.

In order to show (13), we go back to (31). We will show that $r_k(t) = \sum_{w \in \Sigma^+, |w|=k} \sqrt{p_w} A_w \mathbf{z}^{\mathbf{x}}_w(t)$ converges to zero as $k \to \infty$. Since $\mathbf{x}(t)$ is **RC**, $E[\mathbf{z}^{\mathbf{x}}_w(t)(\mathbf{z}^{\mathbf{x}}_v(t))^T] = 0$ for any $w \ne v$ or $w = v \notin L$, $|w| = |v| = k$, and for all $w \in L$, $E[\mathbf{z}^{\mathbf{x}}_w(t)(\mathbf{z}^{\mathbf{x}}_w(t))^T] = \frac{1}{p_\sigma} E[\mathbf{x}(t-k)\mathbf{x}^T(t-k)\mathbf{u}^2_\sigma(t-k)]$, where $w = \sigma s$ for $\sigma \in \Sigma$ and $s \in \Sigma^*$. Denote by $P_\sigma = E[\mathbf{x}(t-k)\mathbf{x}^T(t-k)\mathbf{u}^2_\sigma(t-k)]$. Note that by virtue of $\mathbf{x}(t)$ being **RC**, the definition of P_σ does not depend on t and k. Moreover, from Part 4 it follows that $A_s A_\sigma = 0$ if $s \in \Sigma^*, \sigma \in \Sigma$, $\sigma s \notin L$. In then follows that

$$E[r_k(t) r_k^T(t)] = \sum_{s \in \Sigma^*} \sum_{\sigma \in \Sigma} p_s A_s A_\sigma P_\sigma A_\sigma^T A_s^T. \tag{32}$$

Define $S = \sum_{\sigma \in \Sigma} A_\sigma P_\sigma A_\sigma^T$ and define the linear map $\mathcal{R}$ on the space of matrices $\mathbb{R}^{n \times n}$ as

$$\mathcal{R}(V) = \sum_{\sigma \in \Sigma} p_\sigma A_\sigma V A_\sigma^T.$$

Then $E[r_k(t) r_k^T(t)] = \mathcal{R}^{k-1}(S)$. Notice that $\sum_{\sigma \in \Sigma} p_\sigma A_\sigma^T \otimes A_\sigma^T$ is just the matrix representation of $\mathcal{R}(V)$ in the basis described in [41, Chapter 2]. Hence, by Part 4 of Definition 2 and [41, Proposition 2.5], $\lim_{k \to \infty} \mathcal{R}^k(S) = 0$. Hence, it follows that the limit of $E[||r_k(t)||^2] = \mathrm{trace} E[r_k(t) r_k^T(t)]$ equals zero as $k \to \infty$, which is a equivalent to saying that the mean-square limit of $r_k(t)$ is zero as k goes to ∞. ■

Proof of Lemma 3: In order to prove the statement of the lemma, we use the proof of Lemma 2. Notice that (31) remains valid for $t = k$, if we replace $\mathbf{x}$ by $\hat{\mathbf{x}}$. The assumptions of the lemma ensure that (32) remains valid for $t = k$, where $r_k(k) = \sum_{w \in \Sigma^+, |w|=k} \sqrt{p_w} A_w \mathbf{z}^{\hat{\mathbf{x}}}_w(k)$ and $P_\sigma = E[\hat{\mathbf{x}}(0)\hat{\mathbf{x}}(0)^T \mathbf{u}^2_\sigma(0)]$. With the argument as above, it then follows that $\lim_{k \to \infty} r_k(k) = 0$ in the mean-square sense. Notice that $\mathbf{x}(t) - \hat{\mathbf{x}}(t) = \sum_{v \in \Sigma^*, |v| \ge t} \Sigma_{\sigma \in \Sigma} \sqrt{p_{\sigma v}} A_v B_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma v}(t) - r_t(t)$. The first terms converges to zero in the mean-square sense as $t \to +\infty$, since the series on the right-hand side of (13) is convergent in the mean-square sense. It was shown that the second term $r_t(t)$ converges to zero as $t \to \infty$. Hence, $\mathbf{x}(t) - \hat{\mathbf{x}}(t)$ converges to 0 in mean-square sense. ■

Proof of Lemma 4: First, we show that there exists at most one solution to (14). To this end, assume that there are two solutions $\{P_\sigma\}_{\sigma\in\Sigma}$ and $\{P'_\sigma\}_{\sigma\in\Sigma}$ to (14). Define $\hat{P}_\sigma = P_\sigma - P'_\sigma$. By subtracting th equation (14) for P_σ and P'_σ,

$$\hat{P}_\sigma = \sum_{\sigma_1\in\Sigma, \sigma_1\sigma\in L} p_\sigma A_{\sigma_1} \hat{P}_{\sigma_1} A_{\sigma_1}^T. \tag{33}$$

Using the equation above and the fact that $A_\sigma A_{\sigma_1} = 0$ for $\sigma_1\sigma \notin L$, we obtain

$$A_\sigma \hat{P}_\sigma A_\sigma^T = \sum_{\sigma_1\in\Sigma} p_\sigma A_\sigma A_{\sigma_1} \hat{P}_{\sigma_1} A_{\sigma_1}^T A_\sigma \tag{34}$$

Consider the map $\mathcal{Z} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ defined as $\mathcal{Z}(V) = \sum_{\sigma\in\Sigma} p_\sigma A_\sigma V A_\sigma^T$. It is easy to see that $\sum_{\sigma\in\Sigma} p_\sigma A_\sigma^T \otimes A_\sigma^T$ is a matrix representation of $\mathcal{Z}$. Hence, from Part 4 of Assumption 2 it follows that all the eigenvalues of $\mathcal{Z}$ are inside the unit circle. Define $Q = \sum_{\sigma\in\Sigma} A_\sigma \hat{P}_\sigma A_\sigma^T$ and notice that (34) implies that $Q = \sum_{\sigma\in\Sigma} p_\sigma A_\sigma (\sum_{\sigma_1\in\Sigma} A_{\sigma_1} P_{\sigma_1} A_{\sigma_1}^T) A_\sigma^T = \mathcal{Z}(Q)$. Since 1 is not an eigenvalue of Q, it implies that $Q = 0$. But if $Q = 0$, then (34) implies that $A_\sigma \hat{P}_\sigma A_\sigma^T = p_\sigma A_\sigma Q A_\sigma^T = 0$. Applying (33) yields $\hat{P}_\sigma = 0$, and hence $P_\sigma = P'_\sigma$ for all $\sigma \in \Sigma$.

Next, we show that a solution to (14) exists and it is determined by $P_\sigma = E[\mathbf{x}(t)\mathbf{x}(t)^T \mathbf{u}_\sigma^2(t)] = p_\sigma E[\mathbf{z}_\sigma^{\mathbf{x}}(t)(\mathbf{z}_\sigma^{\mathbf{x}}(t))^T]$. By Lemma 8 $\mathbf{x}(t)$ is **RC**. From Part 3 of Assumption 2 it also follows that for every $w, v \in \Sigma^+$, $|w| \geq |v|$, $\mathbf{z}_w^{\mathbf{x}}(t)$ and $\mathbf{z}_v^{\mathbf{v}}(t)$ are orthogonal. Indeed by Lemma 10, $\mathbf{z}_w^{\mathbf{x}}(t)$ belongs to the Hilbert space generated by $\mathbf{z}_{sw}^{\mathbf{v}}(t)$, $s \in \Sigma^+$ and by Assumption 2, $\mathbf{z}_v^{\mathbf{v}}(t)$ and $\mathbf{z}_{sw}^{\mathbf{v}}(t)$ are orthogonal, since clearly $|sw| > |w| \geq |v|$. Notice the identities $P_\sigma = p_\sigma E[\mathbf{z}_\sigma^{\mathbf{x}}(t+1)(\mathbf{z}_\sigma^{\mathbf{x}}(t+1))^T] = p_\sigma T_{\sigma,\sigma}^{\mathbf{x}}$, $\mathbf{z}_{\sigma_1\sigma}^{\mathbf{x}}(t+1) = \frac{1}{\sqrt{p_{\sigma_1\sigma}}} \mathbf{x}(t-1)\mathbf{u}_{\sigma_1}(t-1)\mathbf{u}_\sigma(t)$, $\mathbf{z}_{\sigma\sigma'}^{\mathbf{v}}(t+1) = \frac{1}{\sqrt{p_{\sigma_1\sigma}}} \mathbf{v}(t-1)\mathbf{u}_{\sigma_1}(t-1)\mathbf{u}_\sigma(t)$ and

$$\mathbf{z}_\sigma^{\mathbf{x}}(t+1) = \sum_{\sigma_1\in\Sigma} \sqrt{p}_{\sigma\sigma_1} (A_{\sigma_1} \mathbf{z}_{\sigma_1\sigma}^{\mathbf{x}}(t+1) + K_{\sigma_1} \mathbf{z}_{\sigma_1\sigma}^{\mathbf{v}}(t+1). \tag{35}$$

Notice that $E[\mathbf{z}_{\sigma_1\sigma}^{\mathbf{x}}(t)(\mathbf{z}_{\sigma_2\sigma}^{\mathbf{x}}(t))^T]$ equals zero, if $\sigma_1 \neq \sigma_2$ or $\sigma_1 = \sigma_2, \sigma_1\sigma \notin L$, and $p_\sigma P_{\sigma_1}$ otherwise. In a similar fashion, $E[\mathbf{z}_{\sigma_1\sigma}^{\mathbf{v}}(t)(\mathbf{z}_{\sigma_2\sigma}^{\mathbf{v}}(t))^T]$ equals zero, if $\sigma_1 \neq \sigma_2$ or $\sigma_1 = \sigma_2, \sigma_1\sigma \notin L$ and $p_\sigma Q_{\sigma_1}$ otherwise. In addition, $E[\mathbf{z}_{\sigma_1\sigma}^{\mathbf{x}}(t+1)(\mathbf{z}_{\sigma_2\sigma}^{\mathbf{v}}(t+1))^T] = 0$, $\sigma_1, \sigma_2, \sigma \in \Sigma$. By noticing $P_\sigma = p_\sigma E[\mathbf{z}_\sigma^{\mathbf{x}}(t+1)(\mathbf{z}_\sigma^{\mathbf{x}}(t+1))^T]$, and applying (35), it follows that $\{P_\sigma\}_{\sigma\in\Sigma}$ satisfies (14). ■

B. Proof of Theorem 5

We prove the claims one by one.

Proof that $\mathbf{y}$ is RC From Lemma 2 it follows that $\mathbf{s}(t) = \begin{bmatrix}\mathbf{v}^T(t), & \mathbf{x}^T(t)\end{bmatrix}^T$, and as Notice $\mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{v}(t) = \begin{bmatrix}C, & D\end{bmatrix}\mathbf{s}(t)$, it then follows that $\mathbf{y}$ is **RC**.

Proof that R_{B} is well-defined and that it is a representation of $\Psi_{\mathbf{y}}$. From Lemma 4 it follows that (14) has at most one solution.

Next, we show that R_{B} is a representation of $\Psi_{\mathbf{y}}$. By induction on $|w|$ we obtain that for all $w \in \Sigma^*$,

$$E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{x}}_{\sigma w}(t))^T] = \frac{1}{\sqrt{p_\sigma}}\sqrt{p_w}A_w A_\sigma E[\mathbf{x}(t-k)\mathbf{x}^T(t-k)\mathbf{u}^2_\sigma], \tag{36}$$

where p_w is defined as in Notation 2. Indeed, for $w = \epsilon$, $\mathbf{z}^{\mathbf{x}}_\sigma(t) = \frac{1}{\sqrt{p_\sigma}}\mathbf{x}(t-1)\mathbf{u}_\sigma(t-1)$ and using that $\mathbf{x}(t) = \sum_{\sigma\in\Sigma}\sqrt{p_\sigma}(A_\sigma \mathbf{z}^{\mathbf{x}}_\sigma(t) + K_\sigma \mathbf{z}^{\mathbf{v}}_\sigma(t))$ and $E[\mathbf{z}^{\mathbf{v}}_{\sigma_1}(t)(\mathbf{z}^{\mathbf{x}}_\sigma(t))^T] = 0$ for all $\sigma_1, \sigma \in \Sigma$ (see Lemma 2) , we obtain (36). If $w = v\hat{\sigma}$, then using $\mathbf{x}(t) = \sum_{\sigma_1\in\Sigma}\sqrt{p_\sigma}(A_{\sigma_1}\mathbf{z}^{\mathbf{x}}_{\sigma_1}(t) + K_{\sigma_1}\mathbf{z}^{\mathbf{v}}_{\sigma_1}(t))$, the induction hypothesis, and the equalities $E[\mathbf{z}^{\mathbf{x}}_{\sigma_1}(t)(\mathbf{z}^{\mathbf{x}}_{\sigma v\hat{\sigma}}(t))^T] = 0$ if $\sigma_1 \neq \hat{\sigma}$ or $\sigma v\hat{\sigma} \notin L$, and $E[\mathbf{z}^{\mathbf{x}}_{\sigma_1}(t)(\mathbf{z}^{\mathbf{x}}_{\sigma v\hat{\sigma}}(t))^T] = \sqrt{p_{\hat{\sigma}}}E[\mathbf{x}(t-1)(\mathbf{z}^{\mathbf{x}}_{\sigma v}(t))^T]$ for $w = \sigma v\hat{\sigma} \in L$, and $E[\mathbf{z}^{\mathbf{v}}_{\sigma_1}(t)(\mathbf{z}^{\mathbf{x}}_{\sigma v\hat{\sigma}}(t))^T] = 0$ (see Lemma 2), and using $A_w A_\sigma = 0$ if $\sigma w \notin L$, we again readily obtain (36). In a similar fashion, we can show that

$$E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{v}}_{\sigma w}(t))^T] = \begin{cases} \frac{1}{\sqrt{p_\sigma}}\sqrt{p_w}A_w K_\sigma Q_\sigma, & \sigma w \in L \\ 0 & \text{otherwise} \end{cases},$$

where $Q_\sigma = E[\mathbf{v}(t)\mathbf{v}^T(t)\mathbf{u}^2_\sigma(t)]$. Finally, notice that $\mathbf{z}_w(t) = C\mathbf{z}^{\mathbf{x}}_w(t) + D\mathbf{z}^{\mathbf{v}}_w(t)$, and $\mathbf{v}(t)$ is orthogonal to the variables $\mathbf{z}^{\mathbf{x}}_w(t)$ and $\mathbf{z}^{\mathbf{v}}_w(t)$. Using the definition $\Lambda^{\mathbf{y}}_{\sigma w} = E[\mathbf{y}(t)\mathbf{z}^T_{\sigma w}(t)]$, $A_w A_\sigma = 0$, $A_w K_\sigma Q_\sigma = 0$ for $\sigma w \notin L$, and (36), we derive

$$\Lambda^{\mathbf{y}}_{\sigma w} = CE[\mathbf{x}(t)(\mathbf{z}^{\mathbf{x}}_w(t))^T]C^T + CE[\mathbf{x}(t)(\mathbf{z}^{\mathbf{v}}_w(t))^T]D^T = \sqrt{p_w}CA_w\frac{1}{\sqrt{p_\sigma}}(A_\sigma P_\sigma C^T + K_\sigma Q_\sigma D^T)$$

That is, $\Lambda^{\mathbf{y}}_{\sigma w} = CA_w B_\sigma$, i.e. R_{B} is a representation of $\Psi_{\mathbf{y}}$.

Finally, from Part 4 of Definition 2 it follows that R_{B} is a stable representation.

Proof that $\mathbf{y}$ satisfies Assumption 3 From the discussion above it follows that $\mathbf{y}$ is **RC** and R_{B} is a stable representation of $\Psi_{\mathbf{y}}$. Hence, $\Psi_{\mathbf{y}}$ is rational and by Theorem 3 $\Psi_{\mathbf{y}}$ is square-summable too.

Proof that $\mathbf{y}$ is full rank To this end, notice that $\mathbf{z}_w(t) = C\mathbf{z}^{\mathbf{x}}_w(t) + D\mathbf{z}^{\mathbf{v}}_w(t)$. Part 3 of Assumption 2 and repeated application of Lemma 11 implies the coordinates of $\mathbf{z}^{\mathbf{x}}_w(t)$ belong to

$H^{\mathbf{v}}_{t,w} \subseteq H^{\mathbf{v}}_t$ and $H^{\mathbf{y}}_t \subseteq H^{\mathbf{v}}_t$. Let $H^{\perp}_t$ be the orthogonal complement of $H^{\mathbf{y}}_t$ in $H^{\mathbf{v}}_t$. From Definition 2 it follows that $E[\mathbf{v}(t)h] = 0$ for any $h \in H^{\mathbf{v}}_t$. Hence, $\mathbf{v}(t)$ is orthogonal to $H^{\mathbf{y}}_t$. Notice that the entries of $\mathbf{x}(t)$ belong to $H^{\mathbf{v}}_t$ and hence it can be written as $\mathbf{x}(t) = \mathbf{x}_1(t) + \mathbf{x}_2(t)$ such that the entries of $\mathbf{x}_2(t)$ belong to $\mathcal{H}^{\perp}_t$. It then follows that $E_l[\mathbf{x}(t) \mid \{\mathbf{z}_w(t) \mid w \in \Sigma^+\}] = \mathbf{x}_1(t)$ since for all $w \in \Sigma^+$, $E[\mathbf{x}_2(t)\mathbf{z}^T_w(t)] = 0$ and hence $E[\mathbf{x}(t)\mathbf{z}^T_w(t)] = E[\mathbf{x}_1(t)\mathbf{z}^T_w(t)]$. Then $E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w(t) \mid w \in \Sigma^+\}] = C\mathbf{x}_1(t)$, since $E[\mathbf{y}(t)\mathbf{z}^T_w(t)] = CE[\mathbf{x}_1(t)\mathbf{z}^T_w(t)]$ and the entries of $C\mathbf{x}_1(t)$ belong to H_t. Moreover, from Lemma 8 it follows that $(\mathbf{y}^T(t), \mathbf{x}^T_2(t))^T$ is **RC**. Similarly, since $\mathbf{v}(t)$ is orthogonal to $H^{\mathbf{v}}_t$, by Lemma 12 $\mathbf{v}(t)\mathbf{u}_\sigma(t)$ is orthogonal to $H^{\mathbf{v}}_{t+1,\sigma}$. Since by Lemma 11 the entries of $\mathbf{x}_i(t)\mathbf{u}_\sigma(t)$, $i = 1, 2$ belong to $H^{\mathbf{v}}_{t+1,\sigma}$, it then follows that $\mathbf{x}_i(t)\mathbf{u}_\sigma(t)$, $i = 1, 2$ and $\mathbf{v}(t)\mathbf{u}_\sigma(t)$ are orthogonal. Notice that $\mathbf{e}(t) = \mathbf{y}(t) - C\mathbf{x}_1(t) = C\mathbf{x}_2(t) + D\mathbf{v}(t)$. Hence, for all $\sigma \in \Sigma$, $E[\mathbf{e}(t)\mathbf{e}^T(t)\mathbf{u}^2_\sigma(t)] = CE[\mathbf{x}_2(t)\mathbf{x}^T_2(t)\mathbf{u}^2_\sigma(t)]C^T + DE[\mathbf{v}(t)\mathbf{v}^T(t)\mathbf{u}^2_\sigma(t)]D^T > 0$, i.e. $\mathbf{y}$ is full rank.

C. *Proof of Theorem 6*

The proof of the theorem is organized as follows. First, we present a number of properties of the state process $\mathbf{x}(t)$ and the innovation process $\mathbf{e}(t)$. Then we show the existence of the matrix K_σ, $\sigma \in \Sigma$. Finally, we prove (19).

Properties of $\mathbf{x}(t)$ and $\mathbf{e}(t)$ Below we present some important properties of $\mathbf{x}(t)$ and $\mathbf{e}(t)$ constructed above. The exposition is organized as a series of lemmas.

Lemma 13: For each $w \in \Sigma^*$ and $\sigma \in \Sigma$ such

$$E[\mathbf{x}(t)\mathbf{z}^T_{w\sigma}(t)] = E[O_R^{-1}(Y_n(t))\mathbf{z}^T_{w\sigma}(t)] = A_w B_\sigma,$$

where $B_\sigma = \begin{bmatrix} B_{1,\sigma}, & \dots, & B_{p,\sigma} \end{bmatrix}$.

Proof of Lemma 13: Notice that because of the properties of orthogonal projection it holds that $\forall v \in \Sigma^+ : E[\mathbf{x}(t)\mathbf{z}^T_v(t)] = E[O_R^{-1}(Y_n(t))\mathbf{z}^T_v(t)]$. Notice that the entries of $E[Y_n(t)\mathbf{z}^T_v(t)]$ are of the form $E[(\mathbf{z}^f_{v_i})^T(t)\mathbf{z}^T_v(t)]$ $i = 0, \dots, M(n-1)$. By writing out the definition of $\mathbf{z}^f_s$ and $\mathbf{z}_v(t)$, it follows that for any $v, s \in \Sigma^+$,

$$E[(\mathbf{z}^f_s)^T(t)\mathbf{z}^T_v(t)] = E[\mathbf{y}(t+l)\mathbf{z}^T_{vs}(t+l)] = \Lambda^{\mathbf{y}}_{vs}.$$

Hence, applying the result above to $v = \sigma w$ and noticing that $\Lambda^{\mathbf{y}}_{\sigma w v_i} = CA_{v_i}A_w B_\sigma$ we obtain $E[Y_n(t)\mathbf{z}_{\sigma w}(t)] = O_R A_w B_\sigma$. From this, by taking into account that $E[\mathbf{x}(t)\mathbf{z}^T_{\sigma w}(t)] = O_R^{-1}E[Y_n(t)\mathbf{z}_w(t)]$, the statement of the lemma follows. ∎

Lemma 13 explains the relationship between states of the would-be generalized bilinear realization and the states of the rational representation R of $\Psi_{\mathbf{y}}$. In particular, it yields the following corollary.

Corollary 6: With the notation of Lemma 13, $C\mathbf{x}(t) = E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w(t) \mid w \in \Sigma^+\}]$.

Proof of Corollary 6: Indeed, from Lemma 13 it follows that for any $v \in \Sigma^+$ of the form $v = \sigma w$ for some $w \in \Sigma^*$, $\sigma \in \Sigma$: $E[C\mathbf{x}(t)\mathbf{z}^T_{\sigma w}(t)] = CA_wB_\sigma = \Lambda^{\mathbf{y}}_{\sigma w} = E[\mathbf{y}(t)\mathbf{z}^T_{\sigma w}(t)]$, and hence for any $v \in \Sigma^+$, $E[(\mathbf{y}(t)-C\mathbf{x}(t))\mathbf{z}^T_v(t)] = 0$, i.e. the entries of $\mathbf{y}(t)-C\mathbf{x}(t)$ are orthogonal to the Hilbert-space generated by $\{\mathbf{z}_v(t) \mid v \in \Sigma^+\}$. Since the entries $C\mathbf{x}(t)$, obviously belong to that Hilbert-space, the corollary follows. ■

The corollary above says that $C\mathbf{x}(t)$ is the projection of the current output to past outputs and inputs.

Lemma 14: The processes $\mathbf{x}(t)$ and $\mathbf{e}(t) = \mathbf{v}(t)$ satisfy Part 1–2 of Assumption 2. Moreover, $\mathbf{s}(t) = \begin{bmatrix}\mathbf{x}^T(t), \mathbf{y}^T(t), \mathbf{e}^T(t)\end{bmatrix}^T$ is **RC**.

Proof of Lemma 14:

From Lemma 13 it follows that $E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{y}}_v(t))^T]$, $v \in \Sigma^+$ does not depend on t. By noticing that the entries of $\mathbf{x}(t)$ belong to $H^{\mathbf{y}}_t$ and applying Lemma 8, it then follows that $\mathbf{s}_1(t) = \begin{bmatrix}\mathbf{x}^T(t), & \mathbf{y}^T(t)\end{bmatrix}^T$ is **RC**. By noticing that $\mathbf{s}(t) = \begin{bmatrix} I_n & 0 \\ 0 & I_p \\ -C & I_p \end{bmatrix} \mathbf{s}_1(t)$, it follows that $\mathbf{s}(t)$ is **RC**.

Proof of Part 1 of Assumption 2 Since $\mathbf{x}(t)$ and $\mathbf{e}(t)$ are components of $\mathbf{s}(t)$, it follows that $\mathbf{x}(t)$ and $\mathbf{e}(t)$ are **RC**.

Proof of Part 2 of Assumption 2 Assume that $w, v \in \Sigma^+$. Assume first that $|w| > |v|$ and $w = sv$ for some $s \in \Sigma^+$. Since the coordinates of $\mathbf{x}(t)$ belong to H_t, from Lemma 11 it follows that the entries of $\mathbf{z}^{\mathbf{x}}_w(t)$ belong to $H^{\mathbf{y}}_{t,w} \subseteq H^{\mathbf{y}}_{t,v}$. From the construction of $\mathbf{e}(t) = \mathbf{y}(t) - C\mathbf{x}$ it follows that $\mathbf{z}^{\mathbf{e}}_w(t) = \mathbf{z}_w(t) - C\mathbf{z}^{\mathbf{x}}_w(t)$. Hence, as the coordinates of $\mathbf{z}_w(t)$ belong to $H^{\mathbf{y}}_{t,v}$, the coordinates of $\mathbf{z}^{\mathbf{e}}_w(t)$ belong to $H^{\mathbf{y}}_{t,v}$. Note that the coordinates of $\mathbf{e}(t)$ are orthogonal to $H^{\mathbf{y}}_t$. Moreover, recall that $\mathbf{s}(t) = \begin{bmatrix}\mathbf{x}^T(t), & \mathbf{e}^T(t), & \mathbf{y}^T(t)\end{bmatrix}^T$ is **RC**. By applying Lemma 12 to $\mathbf{z}(t) = \mathbf{e}(t)$, $\mathbf{h}(t) = \mathbf{y}(t)$, it follows that $\mathbf{z}^{\mathbf{e}}_v(t)$ is orthogonal to $H^{\mathbf{y}}_{t,v}$. Hence, it follows that $E[\mathbf{z}^{\mathbf{e}}_v(t)(\mathbf{z}^{\mathbf{e}}_w(t))^T] = 0$. If $|w| > |v|$ but w does not end with v, then from the fact that $\mathbf{e}(t)$ is **RC** and Part 3 of Definition 5 it follows that $E[\mathbf{z}^{\mathbf{e}}_v(t)(\mathbf{z}^{\mathbf{e}}_w(t))^T] = 0$. If $|w| < |v|$, then $E[\mathbf{z}^{\mathbf{e}}_v(t)(\mathbf{z}^{\mathbf{e}}_w(t))^T] = 0$ follows from the discussion above by considering the transpose of

$E[\mathbf{z}^{\mathbf{e}}_w(t)(\mathbf{z}^{\mathbf{e}}_v(t))^T]$. If $|w| = |v|$ but $w \neq v$, $E[\mathbf{z}^{\mathbf{e}}_v(t)(\mathbf{z}^{\mathbf{e}}_w(t))^T] = 0$ follows from the fact that $\mathbf{e}(t)$ is **RC**, by repeated application of Part 3 of Definition 5 to $\mathbf{r}(t) = \mathbf{e}(t)$. ■

Lemma 15: For any $w \in \Sigma^+$ and $\sigma \in \Sigma$ such that $w\sigma \in L$, $E[\mathbf{x}(t)\mathbf{z}^T_w(t)\mathbf{u}^2_\sigma(t)] = p_\sigma E[\mathbf{x}(t)\mathbf{z}^T_w(t)]$.

Proof of Lemma 15: From Lemma 8 it follows that $\mathbf{s}(t) = (\mathbf{x}^T(t), \mathbf{y}^T(t))^T$ is a **RC** process and hence $T^{\mathbf{s}}_{\sigma,w\sigma} = \Lambda^{\mathbf{s}}_w$ for $w\sigma \in L$. Since $E[\mathbf{x}(t)\mathbf{z}^T_w(t)\mathbf{u}^2_\sigma(t)]$ and $E[\mathbf{x}(t)\mathbf{z}^T_w(t)]$ are the sub-matrices of $p_\sigma T^{\mathbf{s}}_{\sigma,w\sigma}$ and respectively $\Lambda^{\mathbf{s}}_w$, the statement of the lemma follows. ■

Definition of K_σ In order to define K_σ, we need the following auxiliary result.

Lemma 16:

$$H^{\mathbf{y}}_{t+1} = \bigoplus_{\sigma\in\Sigma} H^{\mathbf{y}}_{t,\sigma} \oplus \bigoplus_{\sigma\in\Sigma} < \mathbf{e}(t)\mathbf{u}_\sigma(t) > \tag{37}$$

where $\bigoplus$ denotes the direct sum and $< \mathbf{e}(t)\mathbf{u}_\sigma(t) >$ denoted the Hilbert-space generated by the entries of $\mathbf{e}(t)\mathbf{u}_\sigma(t)$. Here we used Notation 6–7.

Proof of Lemma 16: Indeed, from the definition of $H^{\mathbf{y}}_{t+1}$ and Lemma 8 it is clear that $H^{\mathbf{y}}_{t+1}$ is the closure of the space $\sum_{\sigma\in\Sigma}(H^{\mathbf{y}}_{t,\sigma} + < \mathbf{e}(t)\mathbf{u}_\sigma(t) >)$. From Lemma 11 it follows that $H^{\mathbf{y}}_{t,\sigma}$ and $H^{\mathbf{y}}_{t,\hat{\sigma}}$ are orthogonal for all $\sigma, \hat{\sigma} \in \Sigma$. From Lemma 12 it follows that $H_{t,\sigma}$, $< \mathbf{e}(t)\mathbf{u}_\sigma(t) >$ are orthogonal. Finally, we will show that $H_{t+1,\sigma}$, $< e_t\mathbf{u}_{\hat{\sigma}}(t) >$ are orthogonal for $\sigma \neq \hat{\sigma}$. To this end, notice that for all $\sigma, \hat{\sigma} \in \Sigma$, $\mathbf{e}(t)\mathbf{u}_{\hat{\sigma}}(t) \in H^{\mathbf{y},*}_{t+1,\hat{\sigma}}$ and $H^{\mathbf{y}}_{t+1,\sigma} \subseteq H^{\mathbf{y},*}_{t+1,\sigma}$. From Lemma 11 it then follows that $H^{\mathbf{y},*}_{t+1,\sigma}$ and $H^{\mathbf{y},*}_{t+1,\hat{\sigma}}$ are orthogonal for each $\sigma, \hat{\sigma}$. Hence, $\mathbf{e}(t)\mathbf{u}_{\hat{\sigma}}(t)$ is orthogonal to $H^{\mathbf{y}}_{t+1,\sigma} \subseteq H^{\mathbf{y},*}_{t+1,\sigma}$. Since all the involved spaces $H^{\mathbf{y}}_{t+1,\sigma}$ and $< \mathbf{e}(t)\mathbf{u}_\sigma(t) >$ are mutually orthogonal, the closure of their sum equals their direct sum. ■

From Lemma 16 it follows that

$$\mathbf{x}(t+1) = \sum_{\sigma\in\Sigma} E_l[O^{-1}_R(Y_n(t+1)) \mid H^\sigma_t] + E_l[O^{-1}_R(Y_n(t+1)) \mid< \mathbf{e}(t)\mathbf{u}_\sigma(t) >] \tag{38}$$

Define now K_σ as a $n \times p$ matrix such that

$$E_l[O^{-1}_R(Y_n(t+1)) \mid< \mathbf{e}(t)\mathbf{u}_\sigma(t) >] = K_\sigma \mathbf{e}(t)\mathbf{u}_\sigma(t).$$

If $\mathbf{y}$ is full rank, then K_σ is unique and $K_\sigma = E[O^{-1}_R(Y_n(t+1))\mathbf{e}^T(t)\mathbf{u}_\sigma](E[\mathbf{e}(t)\mathbf{e}^T(t)\mathbf{u}^2_\sigma(t)])^{-1}$.

Proof of (19) The second equation of (19) follows directly from the definition of $\mathbf{e}(t) = \mathbf{y}(t) - C\mathbf{x}(t)$. We will concentrate on the first equation. To this end, we will show that

$$E_l[O^{-1}_R(Y_n(t+1)) \mid H^{\mathbf{y}}_{t+1,\sigma}] = \frac{1}{\sqrt{p_\sigma}} A_\sigma \mathbf{x}(t)\mathbf{u}_\sigma(t). \tag{39}$$

From this and from (38) the first equation of (19) follows. From Lemma 11 and the fact that the entries of $\mathbf{x}(t)$ belong to H_t it follows that the entries of $\mathbf{x}(t)\mathbf{u}_\sigma(t)$ belongs to $H^{\mathbf{y}}_{t+1,\sigma}$. Hence, by Lemma 16 in order to show that (39), it is enough to show that $w \in \Sigma^+$, $\sigma \in \Sigma$,

$$E[O_R^{-1}(Y_n(t+1))\mathbf{z}^T_{w\sigma}(t+1)] = \frac{1}{\sqrt{p_\sigma}}A_\sigma E[\mathbf{x}(t)\mathbf{u}_\sigma(t)\mathbf{z}^T_{w\sigma}(t+1)]. \tag{40}$$

If $w\sigma \notin L$, then $\mathbf{z}^T_{w\sigma}(t+1) = 0$ and hence (40) trivially holds. Hence, in the sequel we assume that $w\sigma \in L$. Then from Lemma 15 and Lemma 13 it follows that

$$\frac{1}{\sqrt{p}_\sigma}A_\sigma E[\mathbf{x}(t)\mathbf{u}_\sigma(t)\mathbf{z}^T_{w\sigma}(t+1)] = \frac{1}{\sqrt{p_\sigma}}A_\sigma(\sqrt{p_\sigma}E[\mathbf{x}(t)\mathbf{z}^T_w(t)]) = A_\sigma A_w B_{\hat{\sigma}}. \tag{41}$$

where $\hat{\sigma}$ is the first letter of w. By applying Lemma 13 to $t+1$ instead of t, we obtain $E[O_R^{-1}(Y_n(t+1))\mathbf{z}^T_{w\sigma}(t+1)] = A_\sigma A_v B_{\hat{\sigma}}$. Hence, by combining this with (41), (40) follows.

Proof of (21), (23) Notice that on the one hand, by Lemma 13 $E[\mathbf{x}(t+1)\mathbf{z}^T_\sigma(t+1)] = B_\sigma$ and on the other hand, if we use (19),

$$E[\mathbf{x}(t+1)\mathbf{z}^T_\sigma(t+1)] = \frac{1}{\sqrt{p_\sigma}}A_\sigma E[\mathbf{x}(t)\mathbf{u}_\sigma\mathbf{z}^T_\sigma(t+1)] + K_\sigma E[\mathbf{e}(t)\mathbf{u}_\sigma(t)\mathbf{z}^T_\sigma(t+1)]. \tag{42}$$

Here we used the corollary of Lemma 16 that $\mathbf{x}(t)\mathbf{u}_\sigma(t)$, $\mathbf{e}(t)\mathbf{u}_\sigma(t)$ are orthogonal to $\mathbf{z}_{\hat{\sigma}}(t+1)$ for $\hat{\sigma} \neq \sigma$, $\sigma, \hat{\sigma} \in \Sigma$ and that $\mathbf{x}(t)\mathbf{u}_\sigma(t)$ and $\mathbf{e}(t)\mathbf{u}_\sigma(t)$ are orthogonal. Indeed, from Lemma 16 it follows that $\mathbf{x}(t)\mathbf{u}_\sigma(t)$ and $\mathbf{e}(t)\mathbf{u}_\sigma(t)$ are orthogonal. From Lemma 11 it follows that the entries of $\mathbf{x}(t)\mathbf{u}_\sigma(t)$ belong to $\mathcal{H}^{\mathbf{y}}_{t+1,\sigma} \subseteq \mathcal{H}^{\mathbf{y},*}_{t+1,\sigma}$ and the entries of $\mathbf{e}(t)\mathbf{u}_\sigma(t) = \mathbf{y}(t)\mathbf{u}_\sigma(t) - C\mathbf{x}(t)\mathbf{u}_\sigma(t)$ belong to $\mathcal{H}^{\mathbf{y},*}_{t+1,\sigma}$. The entries of $\mathbf{z}^{\mathbf{y}}_{\hat{\sigma}}(t+1)$ belong to $\mathcal{H}^{\mathbf{y},*}_{t+1,\hat{\sigma}}$. From Lemma 11 it then follows that the spaces $\mathcal{H}^{\mathbf{y},*}_{t+1,\sigma}$, $\mathcal{H}^{\mathbf{y},*}_{t+1,\hat{\sigma}}$ are orthogonal. Using this remark and the equality $\mathbf{z}_\sigma(t+1) = \frac{1}{\sqrt{p_\sigma}}\mathbf{y}(t)\mathbf{u}_\sigma(t) = \frac{1}{\sqrt{p_\sigma}}(C\mathbf{x}(t)\mathbf{u}_\sigma(t) + \mathbf{e}(t)\mathbf{u}_\sigma(t))$, we obtain

$$E[\mathbf{x}(t)\mathbf{u}_\sigma(t)\mathbf{z}^T_\sigma(t+1)] = \frac{1}{\sqrt{p_\sigma}}P_\sigma C^T. \tag{43}$$

In addition, from the discussion above it follows that $T_{\sigma,\sigma} = E[\mathbf{z}_\sigma(t)\mathbf{z}^T_\sigma(t)] = \frac{1}{p_\sigma}CP_\sigma C^T + \frac{1}{p_\sigma}E[\mathbf{e}(t)\mathbf{e}^T(t)\mathbf{u}^2_\sigma(t)])$ and hence

$$E[\mathbf{e}(t)\mathbf{e}^T(t)\mathbf{u}^2_\sigma(t)] = p_\sigma T_{\sigma,\sigma} - CP_\sigma C^T. \tag{44}$$

Combining (43),(44) and (42) yield

$$B_\sigma = \frac{1}{p_\sigma}A_\sigma P_\sigma C^T + \frac{1}{\sqrt{p_\sigma}}K_\sigma(p_\sigma T_{\sigma,\sigma} - CP_\sigma C^T),$$

from which (21) follows easily. If $\mathbf{y}$ is full rank, then the existence of the inverse of $(p_\sigma T_{\sigma,\sigma} - CP_\sigma C^T)$ follows from (44) and the invertibility of $E[\mathbf{e}(t)\mathbf{e}^T(t)\mathbf{u}_\sigma^2(t)]$. If the inverse of $(p_\sigma T_{\sigma,\sigma} - CP_\sigma C^T)$ exists, then (21) implies (23).

Proof that the system satisfies Assumption 2 We have already shown that Parts 1–2 of Assumption 2 are satisfied. Since R is a minimal representation of $\Psi_{\mathbf{y}}$ and $\Psi_{\mathbf{y}}$ is square-summable, from Theorem 3 it follows that $\sum_{\sigma\in\Sigma} A_\sigma^T \otimes A_\sigma^T$ is stable, i.e. Part 4 of Assumption 2 holds. In order to show that Part 5 of Assumption 2 holds, let $\sigma_1, \sigma_2 \in \Sigma$ such that $\sigma_1\sigma_2 \notin L$. Notice that for all $\sigma \in \Sigma$, $v, w \in \Sigma^*$, $CA_wA_{\sigma_2}A_{\sigma_1}A_vB_\sigma = \Lambda_{\sigma v\sigma_1\sigma_2 w}$. Notice that if $\sigma_1\sigma_2 \notin L$, then $\sigma v\sigma_1\sigma_2 w \notin L$, and hence $CA_wA_{\sigma_2}A_{\sigma_1}A_vB_\sigma = \Lambda_{\sigma v\sigma_1\sigma_2 w} = 0$. Since w is arbitrary, it then follows that $A_{\sigma_2}A_{\sigma_1}A_vB_\sigma \in O_R$. As R is observable, i.e. $O_R = \{0\}$, it follows that $(A_{\sigma_2}A_{\sigma_1})A_vB_\sigma = 0$. Since v and σ are arbitrary, the latter implies that $A_{\sigma_2}A_{\sigma_1}W_R = 0$. As R is reachable, i.e. $W_R = \mathbb{R}^n$, it then follows that $A_{\sigma_2}A_{\sigma_1} = 0$. With a similar reasoning, we can show that if $\sigma_1\sigma_2 \notin L$, then for any $w \in \Sigma^*$, $CA_wA_{\sigma_2}B_{\sigma_1} = \Lambda_{\sigma_1\sigma_2 w} = 0$. Hence, observability of R implies $A_{\sigma_2}B_{\sigma_1} = 0$. Finally, from (21), $A_{\sigma_2}A_{\sigma_1} = 0$, $A_{\sigma_2}B_{\sigma_1} = 0$ it follows that $A_{\sigma_2}K_{\sigma_1}Q_{\sigma_1} = 0$ and thus by recalling that $T^{\mathbf{e}}_{\sigma_1,\sigma_1} = \frac{1}{p_{\sigma_1}}Q_{\sigma_1}$, Part 5 of Definition 2 follows.

It is left to show that Part 3 of Assumption 2 holds. To this end, we have to show that for all $v, w \in \Sigma^+$, $|w| \geq |v|$, $E[\mathbf{z}^{\mathbf{x}}_w(t)(\mathbf{z}^{\mathbf{e}}_v(t))^T] = 0$. In order to show that for all $v, w \in \Sigma^+$, $|w| \geq |v|$, $E[\mathbf{z}^{\mathbf{x}}_w(t)(\mathbf{z}^{\mathbf{e}}_v(t))^T] = 0$, we proceed as follows. Since $|w| \geq |v|$, $w = s\hat{v}$ for some $s \in \Sigma^*$, $\hat{v} \in \Sigma^+$, $|\hat{v}| = |v|$. From Lemma 14 it follows that $\mathbf{r}(t) = \begin{bmatrix}\mathbf{x}^T(t), & \mathbf{e}^T(t)\end{bmatrix}^T$ is **RC**. Moreover, $E[\mathbf{z}^{\mathbf{x}}_w(t)(\mathbf{z}^{\mathbf{e}}_v(t))^T]$ is a suitable sub-matrix of $T^{\mathbf{r}}_{w,v}$. If $\hat{v} \neq v$, then from Part 3 of Definition 5 it follows that $T^{\mathbf{r}}_{w,v} = 0$ and hence $E[\mathbf{z}^{\mathbf{x}}_w(t)(\mathbf{z}^{\mathbf{e}}_v(t))^T] = 0$. Assume now that $v = \hat{v}$, i.e. $w = sv$. Recall that he entries of $\mathbf{e}(t)$ are orthogonal to $H^{\mathbf{y}}_t$ (since $\mathbf{e}(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t) \mid H^{\mathbf{y}}_t]$). Moreover, from Lemma 14 it follows that $(\mathbf{y}^T(t), \mathbf{e}^T(t))^T$ is **RC**. Hence, the conditions of Lemma 12 are satisfied for $\mathbf{z}(t) = \mathbf{e}(t)$, $\mathbf{h}(t) = \mathbf{y}(t)$. Therefore, the entries of $\mathbf{z}^{\mathbf{e}}_v(t)$ are orthogonal to $H^{\mathbf{y}}_{t,v}$. Since the entries of $\mathbf{x}(t)$ belong to $H^{\mathbf{y}}_t$, from Lemma 11 it follows that the entries of $\mathbf{z}^{\mathbf{x}}_w(t)$ belong to $H^{\mathbf{y}}_{t,w}$. Notice that if $w = sv$, then $H^{\mathbf{y}}_{t,w} \subseteq H^{\mathbf{y}}_{t,v}$. Hence in this case the entries of $\mathbf{z}^{\mathbf{e}}_v(t)$ are orthogonal to $H^{\mathbf{y}}_{t,w}$ and thus to the entries of $\mathbf{z}^{\mathbf{x}}_w(t)$. Using that $\mathbf{e}(t), \mathbf{x}(t)$ are **RC**, (19) holds and $\mathbf{v}(t) = \mathbf{e}(t)$ satisfies Part 2 of Assumption 2, and $\sum_{\sigma\in\Sigma} A_\sigma^T \otimes A_\sigma^T$ is stable and that the system satisfies Part 4 of Assumption 2, we can show that

$$\mathbf{x}(t) = \lim_{k\to\infty} \sum_{w\in\Sigma^*, |w|\leq k-1} \sum_{\sigma\in\Sigma} A_wB_\sigma\mathbf{z}^{\mathbf{e}}_{\sigma w}(t), \tag{45}$$

where the limit is taken in the mean square sense. From (45) Part 3 of Assumption 2 follows. In order to show (45), we can use a reasoning similar to the proof of Lemma 2. To this end, notice that (19) implies that

$$\mathbf{x}(t) = \sum_{w \in \Sigma^{+}, |w|=k} A_w \mathbf{z}^{\mathbf{x}}_w(t) + \sum_{w \in \Sigma^{*}, |w| \le k-1} \sum_{\sigma \in \Sigma} A_w K_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t),$$

Hence, it is enough to show that $r_k(t) = \sum_{w \in \Sigma^{+}, |w|=k} A_w \mathbf{z}^{\mathbf{x}}_w(k)$ converges to zero in the mean square sense. Like in the proof of Lemma 2, it can be shown that $E[r_k(t) r_k^T(t)] = \sum_{s \in \Sigma^{*}, |s|=k-1} \sum_{\sigma \in \Sigma} A_s S A_s^T = \mathcal{Z}^{k-1}(S)$, where $S = \sum_{\sigma \in \Sigma} A_\sigma T^{\mathbf{x}}_{\sigma,\sigma} A_\sigma^T$, where $\mathcal{Z}$ is the linear map on $\mathbb{R}^{n \times n}$ defined by $\mathcal{Z}(V) = \sum_{\sigma \in \Sigma} A_\sigma V A_\sigma^T$. Since $\sum_{\sigma \in \Sigma} A_\sigma^T \otimes A_\sigma^T$ is the matrix representation of $\mathcal{Z}$ with respect to the basis defined in [41, Section 2.1], [41, Proposition 2.5] implies that $\lim_{k \to \infty} E[||r_k(t)||^2] = \lim_{k \to \infty} \text{trace} E[r_k(t) r_k^T(t)] = \lim_{k \to \infty} \text{trace} \mathcal{R}_{k-1}(S) = 0$.

D. Proof of Theorem 7

Assume that B is a minimal minimal realization of $\mathbf{y}$ and B satisfies Assumption 2. Then by Theorem 5, R_{B} is a representation of $\Psi_{\mathbf{y}}$ and $\mathbf{y}$ satisfies Assumptions 3. Assume that R_{B} is not a minimal representation of $\Psi_{\mathbf{y}}$. Consider a minimal representation R of $\Psi_{\mathbf{y}}$. From Theorem 6 it then follows that there exists a **GBS** realization B_R of $\mathbf{y}$ such that the dimension of B_R equals $\dim R$ and B_R satisfies Assumption 2. From the construction of R_{B} it follows that $\dim \mathrm{B} = \dim R_{\mathrm{B}}$, hence $\dim \mathrm{B} < \dim R = \dim \mathrm{B}_R$. This contradicts to the minimality of B, hence R_{B} has to be minimal. Conversely, assume that R_{B} is minimal, and consider a **GBS** realization B_1 of $\mathbf{y}$ such that B_1 satisfies Assumption 2. Then $\dim \mathrm{B}_1 = \dim R_{\mathrm{B}_1} \le \dim R_{\mathrm{B}} = \dim \mathrm{B}$. Since B_1 was an arbitrary realization of $\mathbf{y}$, it then follows that B is a minimal realization of $\mathbf{y}$. The second statement of the theorem is a direct consequence of the first one and of Theorem 1.

E. Proof of the results related to the realization algorithm

Below we present the proofs of Lemmas 5 – 6 and Theorem 8. To this end, we will need the following auxiliary result.

Lemma 17: Let M_N be a sequence of closed subspaces such that $M_N \subseteq M_{N+1}$ and let M be the closure of the space $\bigcup_{k=1}^{\infty} M_k$. Let h be a mean-square integrable scalar random variable, and let $\mathbf{z}_N = E_l[h \mid M_N]$ and $\mathbf{z} = E_l[h \mid M]$. Then $\lim_{N \to \infty} \mathbf{z}_N = \mathbf{z}$ in the mean-square sense.

Proof of Lemma 17: It is clear that if $\lim_{N\rightarrow\infty} \mathbf{z}_N$ exists and equals z, then $z = E_l[h \mid M]$. Indeed, $z = E_l[h \mid M]$ if and only if $h - z$ is orthogonal to M. If $z = \lim_{N\rightarrow} \mathbf{z}_N$, then, since $h - z_N$ is orthogonal to M_N, it follows that $h - z$ is orthogonal to M_N for all N. Hence, $h - z$ is orthogonal to M, as the latter is the closure of $\bigcup_{N=0}^{\infty} M_N$.

In order to show that $\mathbf{z}_N$ is convergent in the mean-square sense, it is enough to show that z_N is a Cauchy sequence. To this end, define $d_N = ||h - \mathbf{z}_N||$ and notice that for any $s \in M_N$, $d_N \le ||h - s||$, due the the well-known properties of orthogonal projections onto M_N. Since $M_N \subseteq M_{N+1}$, it then follows that $0 \le d_{N+1} \le d_N$, and hence the limit $\lim_{N\rightarrow\infty} d_N = \alpha$ exists. Notice that $< (h - \mathbf{z}_{N+k}), \mathbf{z}_{N+l} >= 0$ for all $0 \le l \le k$. Hence,

$$||h - \mathbf{z}_{N+k}||^2 =< h - \mathbf{z}_{N+k}, h - \mathbf{z}_{N+k} >=< h - \mathbf{z}_{N+k}, h >=< h - \mathbf{z}_{N+k}, h - \mathbf{z}_N >,$$

and thus

$$||\mathbf{z}_{N+k} - \mathbf{z}_N||^2 = ||(\mathbf{z}_{N+k} - h) + (h - \mathbf{z}_N)||^2 = ||h - \mathbf{z}_N||^2 - ||h - \mathbf{z}_{N+k}||^2$$
$$d_N^2 - d_{N+k}^2$$

Since d_N^2 is convergent, it is a Cauchy sequence, and hence for any $\epsilon > 0$ there exists $N_\epsilon > 0$ such that for any $N > N_\epsilon$ and for any $k \ge 0$, $0 < d_N^2 - d_{N+k}^2 < \epsilon$, and hence $||\mathbf{z}_{N+k} - \mathbf{z}_N||^2 < \epsilon$, i.e. $\mathbf{z}_N$ is indeed a Cauchy sequence.

■

Proof of Lemma 5: From Lemma 17 it follows that $\lim_{N\rightarrow\infty} \mathbf{x}_N(t) = \mathbf{x}(t)$. From this and $\mathbf{e}_N(t) = \mathbf{y}(t) - C\mathbf{x}_N(t)$ and $\mathbf{e}(t) = \mathbf{y}(t) - C\mathbf{x}(t)$ it follows that $\lim_{N\rightarrow\infty} \mathbf{e}_N(t) = \mathbf{e}(t)$. Finally, $\lim_{N\rightarrow\infty} \mathbf{x}_N(t)\mathbf{u}_\sigma(t) = \mathbf{x}(t)\mathbf{u}_\sigma(t)$, $\lim_{N\rightarrow\infty} \mathbf{e}_N(t)\mathbf{u}_\sigma(t) = \mathbf{e}(t)\mathbf{u}_\sigma(t)$ follows from Lemma 10. ■

Proof of Lemma 6: The proof is analogous to the proof of (19). More precisely, we define H_t^N as the linear space generated by $\mathbf{z}_w(t)$, $w \in \Sigma^N$, and define $H_t^{\sigma,N}$ as the linear space generated by $\mathbf{z}_w(t)\mathbf{u}_\sigma(t)$, $w \in \Sigma^N$. Similarly to Lemma 16, it then follows that

$$H_{t+1}^{N+1} = \bigoplus_{\sigma\in\Sigma} H_t^{\sigma,N} \oplus \bigoplus_{\sigma\Sigma} < \mathbf{e}_N(t)\mathbf{u}_\sigma(t) > \tag{46}$$

and hence

$$\mathbf{x}_{N+1}(t+1) = \sum_{\sigma\in\Sigma} (E[O_R^{-1}(Y_n(t+1)) \mid H_t^{\sigma,N}] + E[O_R^{-1}(Y_n(t+1)) \mid < \mathbf{e}_N(t)\mathbf{u}_\sigma(t) >]). \tag{47}$$

Define K_σ^N such that

$$E[O_R^{-1}(Y_n(t+1)) \mid < \mathbf{e}_N(t)\mathbf{u}_\sigma(t) >] = K_\sigma^N \mathbf{e}_N(t)\mathbf{u}_\sigma(t). \tag{48}$$

If we show that

$$E[O_R^{-1}(Y_n(t+1)) \mid H_t^{\sigma,N}] = \frac{1}{\sqrt{p_\sigma}} A_\sigma \mathbf{x}_N(t)\mathbf{u}_\sigma(t), \tag{49}$$

then combining this with (48) and (47) we obtain (24). It is left to show that (49) holds. To this end, notice that $E[\mathbf{x}_{N+1}(t+1)\mathbf{z}_w^T(t+1)] = A_v B_{\hat{\sigma}}$ where $\hat{\sigma}$ is the first letter of w, $w \in \Sigma^{N+1}$. The proof of this equality is analogous to that of Lemma 13. The proof of (47) is then analogous to that of (39). Finally (25) follows from (47) in a way similar to the proof of (23). ■

Lemma 18: If $\mathbf{y}$ is a full-rank process and satisfies Assumption 3 and Assumption 4, then for large enough N, T_N is invertable.

Proof of Lemma 18: Since the underlying assumption of this section is that Assumption 3- and Assumption 4 hold, it follows that $\mathbf{y}$ has a **GBS** realization of the form (19). From $\mathbf{y}(t) = C\mathbf{x}(t) + \mathbf{e}(t)$, $\forall t \in Z$ it follows that $\mathbf{z}_w(t) = C\mathbf{z}_w^x(t) + \mathbf{z}^{\mathbf{e}}(t)$, $\forall t \in Z_m$. $T_{v,w} = CE[\mathbf{z}_v^{\mathbf{x}}(t)(\mathbf{z}_w^{\mathbf{x}}(t))^T]C^T + E[\mathbf{z}_v^{\mathbf{e}}(t)(\mathbf{z}_w^{\mathbf{e}}(t))^T]$. Hence, $T_N = R_N + Q_N$, where R_N is the block matrix $R_N = (CE[\mathbf{z}_{v_i}^{\mathbf{x}}(t)(\mathbf{z}_{v_j}^{\mathbf{x}}(t))^T]C^T)_{i,j=1,\ldots,M(N)}$ and Q_N is the block-diagonal matrix, whose ith diagonal $p \times p$ block, $i = 1, \ldots, M(N)$ equals $p_{\hat{v}_i} E[\mathbf{e}(t)\mathbf{e}^T(t)\mathbf{u}_{\sigma_i}^2(t)]$, $v_i = \sigma_i \hat{v}_i$. Since $\mathbf{y}$ is full rank, it then follows that Q_N is strictly positive definite. Notice that R_N is positive semi-definite by definition (as a covariance matrix of $((\mathbf{z}_{v_1}^{\mathbf{x}}(t))^T, \ldots, (\mathbf{z}_{v_{M(N)}}^{\mathbf{x}}(t))^T)^T$). Hence, T_N is strictly positive definite. ■

Proof of Theorem 8: From Part 1 of assumptions of the theorem it follows that $\lim_{M\to\infty} T_{N,M} = T_N$ and $\lim_{M\to\infty} Q_\sigma^{N,M} = Q_\sigma^N$. Since by Lemma 18 T_N is invertable, it follows that $T_{N,M}$ is invertable for large enough M. Moreover, since $\lim_{N\to\infty} Q_\sigma^N = Q_\sigma > 0$, for large enough N and M, $Q_\sigma^{N,M}$ is invertable for all $\sigma \in \Sigma$. Hence, Algorithm 2 is indeed well posed.

Moreover, Part 1 implies that $\lim_{M\to\infty} \Lambda_w^M = \Lambda_w^{\mathbf{y}}$ for all $w \in \Sigma^+$ and hence $\lim_{M\to\infty} H_{\Psi_{\mathbf{y}},n,n+1}^M = H_{\Psi_{\mathbf{y}},n,n+1}$. Hence, for large enough M, rank $H_{\Psi_{\mathbf{y}},\alpha,\beta}^M =$ rank $H_{\Psi_{\mathbf{y}},\alpha,\beta} =$ rank $H_{\Psi_{\mathbf{y}}} = n$. From Algorithm 1 it is clear that its outcome is continuous in the input matrix $H_{\Psi,N,N+1}$, i.e. $\lim_{M\to\infty} {}^M F_\sigma = A_\sigma$, $\lim_{M\to\infty} {}^M G = B$, $\lim_{M\to\infty} {}^M H = C$. Hence, $\lim_{M\to\infty} \widetilde{\Lambda}_{M,N} = \widetilde{\Lambda}_N$ and hence, $\lim_{M\to\infty} \alpha_{N,M} = \widetilde{\Lambda}_N T_N^{-1} = \alpha_N$. This and (27) yields that $P_\sigma^N = \lim_{M\to\infty} P_\sigma^{M,N}$. Using (25) yields $\lim_{M\to\infty} K_\sigma^{M,N} = K_\sigma^N$. Moreover, notice that $\lim_{M\to\infty} Q_{N,M} = (T_{\sigma,\sigma} - CP_\sigma^N C^T)$ and the latter equals $E[\mathbf{e}_N(t)\mathbf{e}_N^T(t)\mathbf{u}_\sigma^2(t)] = Q_\sigma^N$. Finally, from Lemma 5 it follows that $\lim_{N\to\infty} P_\sigma^N = P_\sigma$ $\lim_{N\to\infty} Q_\sigma^N = Q_\sigma$, and by virtue of (25), $\lim_{N\to\infty} K_\sigma^N = K_\sigma$. ■

V. Jump Markov Linear Systems

The goal of this paper is to present a realization theory for a class of discrete-time stochastic hybrid systems known as jump Markov linear systems (JMLS) [41]. In reality, however, we will look at stochastic hybrid systems of a slightly more general form than the ones defined in [41]. The reason is that the more general class generates the same class of output processes as classical JMLS, but it is easier to establish necessary and sufficient conditions for the existence of a realization for the more general class.

Definition 13 (Generalized jump Markov linear system): A *generalized jump Markov linear system* (GJMLS), H, is a system of the form

$$H : \begin{cases} \mathbf{x}(t+1) & = & M_{\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1)}\mathbf{x}(t) + B_{\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1)}\mathbf{v}(t) \\ \mathbf{y}(t) & = & C_{\boldsymbol{\theta}(t)}\mathbf{x}(t) + D_{\boldsymbol{\theta}}\mathbf{v}(t) \end{cases}. \tag{50}$$

Here $\boldsymbol{\theta}$, $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{v}$ are stochastic processes defined on the whole set of integers, i.e. $t \in \mathbb{Z}$. The process $\boldsymbol{\theta}$ is called the *discrete state process* and takes values in the *set of discrete states* $Q = \{1, 2, \dots, d\}$. The process $\boldsymbol{\theta}$ is a stationary ergodic finite-state Markov process, with state-transition probabilities $p_{i,j} = \mathrm{Prob}(\boldsymbol{\theta}(t+1) = j \mid \boldsymbol{\theta}(t) = i) > 0$ for all $i, j \in Q$. Moreover, the probability distribution of the discrete state $\boldsymbol{\theta}(t)$ is given by the vector $\pi = (\pi_1, \dots, \pi_d)^T$, where $\pi_i = \mathrm{Prob}(\boldsymbol{\theta}(k) = i)$. The process $\mathbf{x}$ is called the *continuous state process* and takes values in one of the *continuous-state spaces* $\mathcal{X}_q = \mathbb{R}^{n_q}$, $q \in Q$. More precisely, for any time $t \in \mathbb{Z}$, the continuous state $\mathbf{x}(t)$ lives in the state-space component $\mathcal{X}_{\boldsymbol{\theta}(t)}$. The process $\mathbf{y}$ is the *continuous output process* and takes values in the *set of continuous outputs* $\mathbb{R}^p$. The process $\mathbf{v}$ is the *continuous noise* and takes values in $\mathbb{R}^m$. The matrices M_{q_1,q_2} and B_{q_1,q_2} are of the form $M_{q_1,q_2} \in \mathbb{R}^{n_{q_2} \times n_{q_1}}$ and $B_{q_1,q_2} \in \mathbb{R}^{n_{q_2} \times m}$ for any pair of discrete states $q_1, q_2 \in Q$. Finally, the matrices C_q and D_q are of the form $C_q \in \mathbb{R}^{p \times n_q}$ and $D_q \in \mathbb{R}^{p \times m}$ for each discrete state $q \in Q$. We will make a number of assumptions on the stochastic processes involved.

Assumption 7: Let $\mathcal{D}_t$ be the σ-algebra generated by $\{\boldsymbol{\theta}(t-k)\}_{k \geq 0}$, and let $\mathcal{D}_{t_1,t_2}$, $t_1 \geq t_2$ be the σ-algebra generated by $\{\boldsymbol{\theta}(t)\}_{t=t_2}^{t_1}$. We assume that for all $t \in \mathbb{Z}$,

1) $\mathbf{v}(t)$ is mean square integrable, it is conditionally zero mean given $\mathcal{D}_{t,t+k}$ for all $k \geq 0$, i.e. $E[\mathbf{v}(t) \mid \mathcal{D}_{t+k}] = 0$, and for all $l > 0$, $\mathbf{v}(t)$ and $\mathbf{v}(t-l)$ are conditionally uncorrelated given $\mathcal{D}_{t,t-l}$, i.e. for all $l > 0$ $E[\mathbf{v}(t)\mathbf{v}^T(t-l)^T \mid \mathcal{D}_{t,t-l}] = 0$. Moreover, $Q_q = E[\mathbf{v}(t)\mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t) = q)]$ does not depend on t.

2) The σ-algebras generated by the random variables $\{\mathbf{v}(t-l), l \geq 0\}$ and $\{\boldsymbol{\theta}(t+l), l > 0\}$ are conditionally independent given $\mathcal{D}_t$.
3) For any $t \in Z$, $\mathbf{x}(t)$ belongs to the Hilbert-space generated by the variables $\mathbf{v}(t-k)\chi(\theta(t-k) = q_0, \ldots, \theta(t) = q_k)$ for all $q_0, \ldots, q_k \in Q$, $k > 0$.
4) The Markov process $\boldsymbol{\theta}$ is stationary and ergodic. Therefore, for all $q \in Q$

$$\sum_{s \in Q} \pi_s p_{s,q} = \pi_q. \tag{51}$$

5) Let $n = n_1 + n_2 + \cdots + n_d$. The matrix

$$\widetilde{M} = \begin{bmatrix} p_{1,1} M_{1,1}^T \otimes M_{1,1}^T & p_{1,2} M_{1,2}^T \otimes M_{1,2}^T & \cdots & p_{1,d} M_{1,d}^T \otimes M_{1,d}^T \\ p_{2,1} M_{2,1}^T \otimes M_{2,1}^T & p_{2,2} M_{2,2}^T \otimes M_{2,2} & \cdots & p_{2,d} M_{2,d}^T \otimes M_{2,d}^T \\ \vdots & \vdots & \cdots & \vdots \\ p_{d,1} M_{d,1}^T \otimes M_{d,1}^T & p_{d,2} M_{d,2}^T \otimes M_{d,2}^T & \cdots & p_{d,d} M_{d,d}^T \otimes M_{d,d}^T \end{bmatrix} \in \mathbb{R}^{n^2 \times n^2} \tag{52}$$

is *stable*. That is, for any eigenvalue λ of $\widetilde{M}$, we have $|\lambda| < 1$.
6) For each $q \in Q$, the matrix $D_q Q_q D_q^T$, where $Q_q = E[\mathbf{v}(t)\mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t) = q)]$, is strictly positive definite.

Assumption 7 implies that future discrete states interact with past noises and continuous states only through the past discrete states. It also implies that for any fixed sequence of discrete states, the noise process is a colored noise and the future noise and the current continuous state are uncorrelated. In addition, Assumption 7 imply that the state process $\mathbf{x}(t)$ is wide-sense stationary and the following holds.

Lemma 19: If Assumption 7 holds, then there exists a unique collection of $n_q \times n_q$ matrices P_q with $q \in Q$, such that P_q satisfy

$$P_q = \sum_{s \in Q} p_{q,s}(M_{s,q} P_s M_{s,q}^T + B_{s,q} Q_{s,q} B_{s,q}^T), \tag{53}$$

where $Q_{s,q} = E[\mathbf{v}(t)\mathbf{v}(t)^T \chi(\boldsymbol{\theta}(t) = s, \boldsymbol{\theta}(t+1) = q)]$. In addition $P_q = E[\mathbf{x}(t)\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = q)$ for all $q \in Q$ and $t \in Z$.

We present the proof of Lemma 19 later on in the text. In fact, Lemma 19 follows from Lemma 4 and the relationship between GJMLSs and **GBS**s which will be presented in the sequel. realization by a GJLS system as follows. Next, we define the notion of dimension for GJMLS.

Definition 14 (Dimension of a GJMLS): The *dimension* of a GJMLS H with discrete state process $\boldsymbol{\theta}$ taking values on $Q = \{1, 2, \ldots, d\}$ is defined as

$$\dim H = n_1 + n_2 + \cdots + n_d, \tag{54}$$

where n_i is the dimension of the continuous state space associated with discrete state q, i.e. $n_q = \dim \mathcal{X}_q$, for $q \in Q$.

Remark 8: Notice that two GJLSs can have the same dimension even if the dimensions of the individual continuous components are completely different.

The main motivation for the definitions above is that it allows us to formulate a neat characterization of minimality. In addition, it is intuitively appealing, as the definition of dimension reflects the amount of date which is required to store the state information. Next, we define when a GJMLS is a realization of a given process. For ease of notation, *in the sequel we will keep the discrete state process $\boldsymbol{\theta}$ fixed and whenever we speak of a GJMLS realization of the process $\mathbf{y}$, we will always mean a GJMLS of $\mathbf{y}$ with discrete state process $\boldsymbol{\theta}$*. More precisely, let $\mathbf{y}$ be a stochastic process taking values in $\mathbb{R}^p$.

Definition 15 (Realization by GJMLS): The GJMLS H with discrete state process $\boldsymbol{\theta}$ is said a *realization* of $\mathbf{y}$, if the output process of H equals $\mathbf{y}$. We call a realization H of $\mathbf{y}$ *minimal*, whenever $\dim H \leq \dim H'$ for all GJMLSs H' that are realizations of $\mathbf{y}$.

This section will be devoted to the solution of the following realization problem for GJMLSs with fully observed discrete states.

Problem 1 (Realization problem for jump-markov systems): Given a process $\mathbf{y}$ and find conditions for existence of a GJMLS which is a realization of $\mathbf{y}$ and characterize minimality of GJMLS realizations of $\mathbf{y}$.

A. Relationship between JMLS and GJMLS

Note that the classical definition of discrete-time JMLS [41] differs from (50). The main difference is that in our framework the continuous state transition rule depends not only on the current, but also on the next discrete state. More specifically, a JMLS according to [41] is a GJMLS system of the form

$$\mathbf{S} : \begin{cases} \mathbf{x}(t+1) & = A_{\boldsymbol{\theta}(t)}\mathbf{x}(t) + B_{\boldsymbol{\theta}(t)}\mathbf{v}(t) \\ \mathbf{y}(t) & = C_{\boldsymbol{\theta}(t)}\mathbf{x}(t) + D_{\boldsymbol{\theta}(t)}\mathbf{v}(t) \end{cases}. \tag{55}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the state process, $\mathbf{v}(t) \in \mathbb{R}^m$ is the noise process, $\mathbf{y}(t) \in \mathbb{R}^p$ is the output process and $A_q \in \mathbb{R}^{n \times n}$, $C_q \in \mathbb{R}^{p \times n}$, $B_q \in \mathbb{R}^{n \times m}$ and $D_q \in \mathbb{R}^{p \times m}$ for all $q \in Q = \{1, \ldots, d\}$. In other words a JMLS is just a GMLJS of the form (50) such that $n_q = n$ for all $q \in Q$ and $M_{q_1,q_2} = A_{q_2}$, i.e. M_{q_1,q_2} depends only on q_2 for all $q_1, q_2 \in Q$. In case of JMLS, one does not speak of state-spaces belonging to different discrete states and the most natural candidate for the state-space of a JMLS $\mathbf{S}$ is the space $\mathbb{R}^n$. Therefore, the most natural definition of dimension for a JMLS is the dimension n of its state-space.

The classes of GJMLS and JMLS are equivalent in the following sense. First, it is clear that a classical JMLS also satisfies our definition. Conversely, a GJMLS of the form (50) can be rewritten as a classical JMLS with the same noise and output processes, but with the continuous state process and the system matrices are replaced by a continuous state process and system matrices living in the continuous space $\mathbb{R}^{n_1+n_2+\cdots+n_d}$. More precisely, if H is a GJMLS of the form (50), then define the JMLS

$$\mathbf{S}(H) : \begin{cases} \hat{\mathbf{x}}(t+1) & = \hat{A}_{\boldsymbol{\theta}(t)}\hat{\mathbf{x}}(t) + \hat{B}_{\boldsymbol{\theta}(t)}\mathbf{v}(t) \\ \hat{\mathbf{y}}(t) & = \hat{C}_{\boldsymbol{\theta}(t)}\hat{\mathbf{x}}(t) + D_{\boldsymbol{\theta}(t)}\mathbf{v}(t) \end{cases}, \tag{56}$$

where $\hat{\mathbf{x}}(t) = \begin{bmatrix}\hat{\mathbf{x}}_1^T(t), & \ldots, & \hat{\mathbf{x}}_d^T(t)\end{bmatrix}^T$, $\hat{\mathbf{x}}_q^T(t) = M_{q,\boldsymbol{\theta}(t-1)}\mathbf{x}(t-1) + B_{q,\boldsymbol{\theta}(t-1)}\mathbf{v}(t-1)$, $q \in Q$, and

$$\hat{A}_q = \begin{bmatrix} \delta_{1,q}M_{1,1}, & \delta_{2,q}M_{1,2}, & \ldots & \delta_{d,q}M_{1,d} \\ \delta_{1,q}M_{2,1}, & \delta_{2,q}M_{2,2}, & \ldots & \delta_{d,q}M_{2,d} \\ \vdots & \vdots & \ldots & \vdots \\ \delta_{1,q}M_{d,1}, & \delta_{2,q}M_{d,2}, & \ldots & \delta_{d,q}M_{d,d} \end{bmatrix} \quad \hat{B}_q = \begin{bmatrix} B_{1,q} \\ B_{2,q} \\ \vdots \\ B_{d,q} \end{bmatrix}$$

$$\hat{C}_q = \begin{bmatrix}\delta_{1,q}C_1, & \delta_{2,q}C_2, & \ldots & \delta_{d,q}C_d\end{bmatrix},$$

where $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$ for all $i, j \in Q$. It is easy to see that the output of $\mathbf{S}_H$ and H coincide, i.e. $\hat{\mathbf{y}}(t) = \mathbf{y}(t)$. Hence, a process can be realized by a GJMLS if and only if it can be realized by a JMLS. In addition, notice that is we define the dimension of a JMLS as the dimension n of its state-space, then $\dim \mathbf{S}_H = \dim H$. In other words, the definition of the dimension for a GJMLS becomes the natural definition, once the GJMLS is converted to a JMLS. This is a further argument in favor of the definition of dimension of GJMLS adopted in this paper.

§V-B presents conditions for the existence of a realization of GJMLS and a characterization of minimal GJMLSs. The proofs of §the results of V-B are presented in §V-C.

B. Solution to the realization problem for GJMLS

Below we will present the solution to the realization problem for GJMLS. We will only state the results, their proofs will be presented in §V-C. We start with formulating conditions for existence of a realization by a GJMLS. To this end, we fix a process $\mathbf{y}(t) \in \mathbb{R}^p$ GJMLS and a Markov-process $\boldsymbol{\theta}(t) \in Q = \{1, \dots, d\}$. We will formulate sufficient and necessary conditions for $\mathbf{y}(t)$ to admit a GJMLS realization. In order to formulate the assumptions on $\mathbf{y}$ which characterize realizability, we will recall the terminology of Section III and we will try to interpret $\mathbf{y}(t)$ as a potential output process of a **GBS**. More precisely, we define the alphabet Σ to be the set of pairs of discrete states, i.e. $\Sigma = Q \times Q$. For each letter $(q_1, q_2) \in \Sigma$ let the input processes of B be defined as

$$\mathbf{u}_{(q_1,q_2)}(t) = \chi(\boldsymbol{\theta}(t+1) = q_2, \boldsymbol{\theta}(t) = q_1). \tag{57}$$

Define $p_{(q_1,q_2)} = p_{q_1,q_2}$. Notice that Assumption 1 holds with $\alpha_\sigma = 1$ for all $\sigma \in \Sigma$. We define the set of admissible sequences L (see Definition 4) as

$$L = \{w = (q_1, q_2)(q_2, q_3) \cdots (q_{k-1}, q_k) \mid k \geq 0, q_1, q_2, \dots, q_k \in Q\}. \tag{58}$$

Notice that if $w = \sigma_1\sigma_2 \cdots \sigma_k \notin L$, then $\mathbf{u}_{\sigma_1}(t-k) \cdots \mathbf{u}_{\sigma_k}(t) = 0$. Using the correspondence described above, we can interpret the process $\mathbf{z}^{\mathbf{y}}_w(t)$ defined in (10), i.e. if $w = \sigma_1 \cdots \sigma_k \in \Sigma^+$, $\sigma_1, \dots, \sigma_k \in \Sigma$, with $\sigma_i = (q_{2i-1}, q_{2i})$, for $q_{2i-1}, q_{2i} \in Q$, $i = 1, \dots, k$, then if $w \notin L$, i.e. $q_{2i} \neq q_{2i+1}$ for some $i = 1, \dots, k$, then $\mathbf{z}_w(t) = 0$, and if $q_{2i} = q_{2i+1}$ for all $i = 1, \dots, k$, i.e. if $w \in L$, then

$$\mathbf{z}^{\mathbf{y}}_w(t) = \mathbf{y}(t-k)\chi(\boldsymbol{\theta}(t-k) = s_1, \dots, \boldsymbol{\theta}(t) = s_k)$$

where $s_i = q_{2i-1}$, $i = 1, \dots, k$. In accordance with Notation 3 we drop the superscript $\mathbf{y}$ and we denote $\mathbf{z}^{\mathbf{y}}_w(t)$ by $\mathbf{z}_w(t)$. The terminology above allows us to apply Definition 9 to $\mathbf{y}$ and speak of $\mathbf{y}$ being full rank.

Now we can formulate the assumptions which are necessary and sufficient for existence of a GJMLS realization of $\mathbf{y}$.

Assumption 8: 1) $\{\mathbf{y}(t), \mathbf{z}_w(t) \mid w \in \Sigma^+\}$ is jointly zero-mean, wide-sense stationary, i.e. $E[\mathbf{y}(t)] = 0$, $E[\mathbf{z}_w(t)] = 0$ for all $t \in \mathbb{Z}$ and the covariances $E[\mathbf{z}_w(t)\mathbf{z}_v^T(t)]$, $E[\mathbf{y}(t)\mathbf{z}_w^T(t)]$, $w, v \in \Sigma^+$ are independent of $t \in \mathbb{Z}$,

2) the σ-algebras generated by $\{\mathbf{y}(t-l)\}_{l=0}^{\infty}$ and $\{\boldsymbol{\theta}(t+l)\}_{l=0}^{\infty}$ are conditionally independent w.r.t to the σ-algebra $\mathcal{D}_t$ generated by $\{\boldsymbol{\theta}(t-l)\}_{l=0}^{\infty}$

3) $y(t)$ if a full rank process.

In fact, Assumption 8 not only guarantees existence of a GJMLS realization, but it also guarantees existence of a GJMLS realization which is its own Kalman-filter, i.e. the best possible estimate of its state based on observable is the state itself. In order to state the existence of such a GJMLS, we need additional terminology.

Definition 16 (GJMLS in forward innovation form): We will call a GJMLS H of the form (50) a GJMLS in *forward innovation form*, if the noise process $\mathbf{v}(t)$ equals the innovation process $\mathbf{e}(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w(t) \mid w \in \Sigma^+\}]$ and D_q is the $p \times p$ identity matrix for all $q \in Q$.

With the definitions above, we can state the main result of existence of a GMJLS realization.

Theorem 9 (Existence of a GJMLS Realization): The process $\mathbf{y}$ satisfies Assumption 8 if and only if there exists a GJMLS H of the form (50) which is a realization of $\mathbf{y}$ and which satisfies Assumptions 7. Moreover, H can be chosen to be in forward innovation form.

From the discussion in §V-A and Theorem 9 we can also deduce the following condition for an existence of a realization by JMLS.

Corollary 7: Theorem 9 remains valid if we replace the word GJMLS by JMLS.

The second claim of Theorem 9 is important for filtering. Notice that if H is a GJMLS (respectively JMLS) is in forward innovation form, then it is easy to see that $\mathbf{x}(t) = E_l[\mathbf{x}(t) \mid \{z_w(t) \mid w \in \Sigma^+\}]$, i.e. the Kalman-filter of the H is H itself. Recall that Kalman-filtering of JMLS is a well-established topic [41].

Theorem 9 follows from Theorem 4 by establishing a correspondence between GJMLSs and **GBS**s. This correspondence is interesting on its own right. Moreover, it will help us to formulate the characterization of minimality for GJMLSs. The definition of this correspondence will also explain our choice of working with GJMLSs rather than JMLSs: the correspondence is much simpler for GJMLSs than for JMLSs. For this reason, we will present this correspondence below.

We will use the following notation.

Notation 8 (Identity and zero matrices): In the sequel, we denote by $\mathbf{O}_{k,l}$ the $k \times l$ matrix with all zero entries and we denote by I_k the $k \times k$ identity matrix.
In addition, we will introduce an *auxiliary output process* $\widetilde{\mathbf{y}}(t) \in \mathbb{R}^{pd}$ which is defined as follows

$$\widetilde{\mathbf{y}}(t) = \begin{bmatrix}\mathbf{y}^T(t)\chi(\boldsymbol{\theta}(t)=1), & \mathbf{y}^T(t)\chi(\boldsymbol{\theta}(t)=2), & \ldots, & \mathbf{y}^T(t)\chi(\boldsymbol{\theta}(t)=d)\end{bmatrix}^T \tag{59}$$

Below we will show that GJMLSs realization of $\mathbf{y}$ yield **GBS**s realizations of $\widetilde{\mathbf{y}}$ and vice versa. Moreover, these transformations preserve minimality and isomorphisms. This will enable us to use the existing results on existence of a **GBS** realization and its minimality to prove the corresponding results for GJMLSs. Notice that for

$$\mathcal{E} = \begin{bmatrix}I_p, & \ldots, & I_p\end{bmatrix} \in \mathbb{R}^{p \times pd},$$

$\mathbf{y}(t) = \mathcal{E}\widetilde{\mathbf{y}}(t)$. Hence, if B of the form (8) is a realization of $\widetilde{\mathbf{y}}(t)$, then by replacing the matrices C and D of B with $\mathcal{E}C$ and $\mathcal{E}D$, we obtain a **GBS** realization of $\mathbf{y}$.

In fact, from the definition of $\widetilde{\mathbf{y}}$ we can conclude the following.

Lemma 20: If the process $\mathbf{y}$ satisfies Assumption 8, then $\widetilde{\mathbf{y}}$ also satisfies Assumption 3. In addition, if we define $\widetilde{\mathbf{e}}(t) = \widetilde{\mathbf{y}}(t) - E_l[\widetilde{\mathbf{y}}(t) \mid \{\mathbf{z}^{\widetilde{\mathbf{y}}}_w(t) \mid w \in \Sigma^+\}]$, then

$$\widetilde{\mathbf{e}}(t) = \begin{bmatrix}\mathbf{e}^T(t)\chi(\boldsymbol{\theta}(t)=1), & \ldots, & \mathbf{e}^T(t)\chi(\boldsymbol{\theta}(t)=d)\end{bmatrix}^T.$$

Moreover, the Hilbert-space spanned by the entries of $\{\mathbf{z}_w(t) \mid w \in \Sigma+\}$ coincides with that of spanned by the elements of $\{\mathbf{z}^{\mathbf{y}}_w(t) \mid w \in \Sigma^+\}$.
Next, we associate a generalized bilinear system B_H with a GBJMLS H.

*Definition 17 (**GBS** associated with a GJMLS):* Assume that H is a GJLS of the form (50) and H satisfies Assumptions 7. We will define the **GBS** B_H, referred to as the ***GBS** associated with* H as follows.

$$\mathrm{B}_H \begin{cases} \widetilde{\mathbf{x}}(t+1) = \sum_{\sigma \in \Sigma}(A_\sigma \widetilde{\mathbf{x}}(t) + \widetilde{K}_\sigma \widetilde{\mathbf{v}(t)})\mathbf{u}_\sigma(t) \\ \widetilde{\mathbf{y}}(t) = C\widetilde{\mathbf{x}}(t) + D\widetilde{\mathbf{v}}(t), \end{cases} \tag{60}$$

In order to define the parameters of B, we define $n = n_1 + \cdots + n_d$ and for each $q \in Q$ define the matrices $\mathrm{I}_q \in \mathbb{R}^{n \times n_q}$, $\mathbf{S}_q \in \mathbb{R}^{m \times dm}$

$$\mathbf{S}_q = \begin{bmatrix}\mathbf{O}_{m,(q-1)m}, I_{m \times m}, & \mathbf{O}_{m,(d-q)m}\end{bmatrix}$$

$$\mathbf{I}_q = \begin{bmatrix}\mathbf{O}_{n_q,n_1}, & \ldots & \mathbf{O}_{n_q,n_{q-1}} & , I_{n_q}, & \mathbf{O}_{n_q,n_{q+1}}, & \ldots, \mathbf{O}_{n_q,n_d}\end{bmatrix}^T.$$

Using the matrices above, we define the parameters of B_H as follows.

State $\widetilde{\mathbf{x}}(t)$**.** $\widetilde{\mathbf{x}}(t) = \begin{bmatrix}\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t)=1), & \ldots, & \mathbf{x}^T(t)\boldsymbol{\theta}(t)=d)\end{bmatrix}^T \in \mathbb{R}^n$, $n = n_1 + \cdots + n_d$.

Noise $\widetilde{\mathbf{v}}(t)$**.** $\widetilde{\mathbf{v}}(t) = \begin{bmatrix}\mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t)=1), & \ldots, & \mathbf{v}^T(t)\boldsymbol{\theta}(t)=d)\end{bmatrix}^T \in \mathbb{R}^{dm}$.

Matrices $A_{(q_1,q_2)}$**.** Define for each q_1, q_2 let A_{q_2,q_1} be the $n \times n$ matrix

$$A_{(q_1,q_2)} = \mathbf{I}_{q_2} M_{q_1,q_2} \mathbf{I}_{q_1}^T$$

The matrix $\widetilde{K}_{(q_1,q_2)}$**.** The $n \times md$ matrix $\widetilde{K}_{(q_1,q_2)}$ is defined as

$$K_{(q_1,q_2)} = \mathbf{I}_{q_2} B_{q_1,q_2} \mathbf{S}_{q_1}.$$

Matrix C**.** The $p \times n$ matrix C is defined by

$$C = \begin{bmatrix}\mathbf{I}_1 C_1^T, & \mathbf{I}_2 C_2^T, & \ldots, & \mathbf{I}_d C_d^T\end{bmatrix}^T.$$

That is, C is a diagonal matrix, such that for all $q \in Q$ its diagonal block indexed by row indices $i = (q-1)p, \ldots, qp$ and column indices $j = [n_1 + \cdots + n_{q-1} + 1, \ldots, n_1 + \cdots + n_q]$ equals C_q.

Matrix D The $p \times md$ matrix D is defined by

$$D = \begin{bmatrix}\mathbf{S}_1^T D_1^T, & \mathbf{S}_2^T D_2^T, & \ldots, & \mathbf{S}_d^T D_d^T\end{bmatrix}^T.$$

Lemma 21: The output process of B_H equals $\widetilde{\mathbf{y}}$. If H satisfies Assumptions 7, then B_H satisfies Assumption 2. Moreover, if we define $\hat{D} = \mathcal{E}D = \begin{bmatrix}D_1, & \ldots, & D_d\end{bmatrix}$, then for any $\sigma = (q_1, q_2) \in \Sigma$, $\hat{D}E[\hat{\mathbf{v}}^T(t)\hat{\mathbf{v}}(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2]\hat{D}^T$ is strictly positive definite.

Remark 9: If the GJMLS H is a jump-Markov linear system of the type studied in [41], i.e.

$$\mathbf{H} : \begin{cases} \mathbf{x}(t+1) & = F_{\boldsymbol{\theta}(t)}\mathbf{x}(t) + G_{\boldsymbol{\theta}(t)}\mathbf{v}(t) \\ \mathbf{y}(t) & = \ H_{\boldsymbol{\theta}(t)}\mathbf{x}(t) + L_{\boldsymbol{\theta}(t)}\mathbf{v}(t) \end{cases}. \tag{61}$$

where $F_q \in \mathbb{R}^{n \times n}$, $G_q \in \mathbb{R}^{n \times m}$, $H_q \in \mathbb{R}^{p \times n}$, $G_q \in \mathbb{R}^{p \times m}$, $q \in \Theta$, then we can directly construct a **GBS**

$$\mathrm{B} \begin{cases} \widetilde{\mathbf{x}}(t+1) = \sum_{\sigma \in \Sigma} (\widetilde{A}_\sigma \widetilde{\mathbf{x}}(t) + \widetilde{K}_\sigma \widetilde{\mathbf{v}}(t))\mathbf{u}_\sigma(t) \\ \mathbf{y}(t) = \widetilde{C}\widetilde{\mathbf{x}}(t) + \widetilde{D}\widetilde{\mathbf{v}}(t), \end{cases} \tag{62}$$

whose output is $\mathbf{y}$. In this case, $\widetilde{A}_{(q_1,q_2)}$ is a $nd \times nd$ matrix, all elements of which are zero, except the $n \times n$ block at location (q_1, q_2) which equal F_{q_2}. Similarly, $\widetilde{K}_{(q_1,q_2)}$ is an $nd \times md$

matrix, all elements of which are zero, except the $n \times m$ block at location (q_1, q_2) which equals G_{q_2}. That is,

$$\widetilde{A}_{(q_1,q_2)} = \begin{bmatrix} \delta_{(1,1),(q_1,q_2)} F_1 & \cdots & \delta_{(1,d),(q_1,q_2)} F_d \\ \vdots & \vdots & \vdots \\ \delta_{(d,1),(q_1,q_2)} F_1 & \cdots & \delta_{(d,d),(q_1,q_2)} F_d \end{bmatrix}, \quad \widetilde{K}_{(q_1,q_2)} = \begin{bmatrix} \delta_{(1,1),(q_1,q_2)} G_1 & \cdots & \delta_{(1,d),(q_1,q_2)} G_d \\ \vdots & \vdots & \vdots \\ \delta_{(d,1),(q_1,q_2)} G_1 & \cdots & \delta_{(d,d),(q_1,q_2)} G_d \end{bmatrix}$$

where $\delta_{(i,j),(k,l)} = 1$ if $k = i$ and $j = l$ and $\delta_{(i,j),(k,l)} = 0$ otherwise. The matrices $\widetilde{C}$ and $\widetilde{D}$ are $\widetilde{C} = \begin{bmatrix} C_1, & \dots, & C_d \end{bmatrix}^T$, $\widetilde{D} = \begin{bmatrix} L_1, & \dots, & L_d \end{bmatrix}^T$. The processes $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{v}}$ are defined as $\widetilde{\mathbf{x}}(t) = \begin{bmatrix} \mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = 1), & \dots, & \mathbf{x}^T(t)\boldsymbol{\theta}(t) = d) \end{bmatrix}^T$, $\widetilde{\mathbf{v}}(t) = \begin{bmatrix} \mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t) = 1), & \dots, & \mathbf{v}^T(t)\boldsymbol{\theta}(t) = d) \end{bmatrix}^T$. If H satisfies Assumptions 7, then B defined above satisfies Assumption 2.

We can reverse the construction above, by associating with every **GBS** B a GJMLS H.

*Definition 18 (GJMLS associated with **GBS**):* Let B be a **GBS** of the form

$$\mathbf{x}(t+1) = \sum_{\sigma \in \Sigma} (A_\sigma \mathbf{x}(t) + K_\sigma \widetilde{\mathbf{e}}(t))\mathbf{u}_\sigma(t)$$

$$\widetilde{\mathbf{y}}(t) = C\mathbf{x}(t) + \widetilde{\mathbf{e}}(t)$$

where $\widetilde{\mathbf{e}}(t)$ is the innovation process of $\widetilde{\mathbf{y}}(t)$ defined in Lemma 20. Define the *GJMLS* H_{B} *associated with* B as follows.

$$H_{\mathrm{B}} : \begin{cases} \hat{\mathbf{x}}(t+1) = M_{\boldsymbol{\theta}(t+1),\boldsymbol{\theta}(t)} \hat{\mathbf{x}}(t) + \widetilde{K}_{\boldsymbol{\theta}(t+1),\boldsymbol{\theta}(t)} \mathbf{e}(t) \\ \mathbf{y}(t) = C_{\boldsymbol{\theta}(t)} \hat{\mathbf{x}}(t) + \mathbf{e}(t), \end{cases} \tag{63}$$

where In order to define the parameters of H_{B}, we use the following notation.

For each $q \in Q$, define the matrix $\mathbf{M}_q \in \mathbb{R}^{p \times pd}$ as

$$\mathbf{M}_q = \begin{bmatrix} \mathbf{O}_{p,p(q-1)}, & I_p, & \mathbf{O}_{p,p(d-q-1)} \end{bmatrix}.$$

For each $q \in Q$ define $\mathcal{X}_q \subseteq \mathbb{R}^n$ as the subspace spanned by all the elements belonging to $\mathrm{Im} A_{(q_1,q)} A_w K_{(q_2,q_3)} \mathbf{M}_{q_2}^T$ and $\mathrm{Im} K_{(q_1,q)} \mathbf{M}_{q_1}^T$ for all $q_1, q_2, q_3 \in Q$, $w \in \Sigma^*$, $\sigma \in \Sigma$, $i = 1, \dots, p$. Let $n_q = \dim \mathcal{X}_q$. Let $\Pi_q \in \mathbb{R}^{n \times n_q}$ be such that the columns of Π_q are orthogonal and they span $\mathcal{X}_q$, i.e. $\Pi_q^T \Pi_q = I_{n_q}$ and $\mathrm{Im}\Pi_q = \mathcal{X}_q$. Then Π_q is the matrix representation of the inclusion $\mathcal{X}_q \subseteq \mathbb{R}^n$ and Π_q^T is the matrix representation of the projection of elements of $\mathbb{R}^n$ to $\mathcal{X}_q$.

1) *Continuous state-space for* $q \in Q$*:* $\mathbb{R}^{n_q}$, $n_q = \dim \mathcal{X}_q$.
2) *State process.* The continuous state process $\hat{\mathbf{x}}(t)$ of the GJMLS is obtained from the continuous state $\mathbf{x}(t)$ of the generalized bilinear system 19 as follows. Then let $\hat{\mathbf{x}}(t) =$

$\Pi^T_{\boldsymbol{\theta}(\theta)}(\mathbf{x}(t))$, i.e. $\hat{\mathbf{x}}(t)$ is obtained from $\mathbf{x}(t)$ by viewing it as an element of $\mathcal{X}_{\boldsymbol{\theta}(t)}$ and identifying it with the corresponding vector in $\mathbb{R}^{n_q}$ for $q = \boldsymbol{\theta}(t)$.

3) *System matrices.* For each $q_1, q_2 \in Q$ the matrix $M_{q_1,q_2} \in \mathbb{R}^{n_{q_2} \times n_{q_1}}$ is defines as

$$M_{q_1,q_2} = \Pi^T_{q_2} A_{(q_1,q_2)} \Pi_{q_1}$$

i.e. M_{q_1,q_2} is the matrix representation of the ma[$\mathcal{X}_{q_1} \ni x \mapsto A_{(q_1,q_2)}x \in \mathcal{X}_{q_2}$. For each $q \in Q$ the matrix $C_q \in \mathbb{R}^{p \times n_q}$ as

$$C_q = \mathbf{M}_q C \Pi_q.$$

4) *Noise gain* $\widetilde{K}_{q_1,q_2}$ Let $\widetilde{K}_{q_1,q_2} = \Pi^T_{q_2} K_{(q_1,q_2)} \mathbf{M}^T_{q_1}$.

Lemma 22: Assume that B is in forward innovation form, it satisfies Assumptions 2, and it is a realization of $\widetilde{\mathbf{y}}$. Assume moreover that $\mathbf{y}$ satisfies Part 2 of Assumption 8. Then H_{B} is also a realization of $\mathbf{y}$, it is in forward innovation form, and it satisfies Assumptions 7. Moreover, if the representation R_{B} associated with B is reachable and observable, then $\mathbb{R}^n = \bigoplus_{q \in Q} \mathcal{X}_q$ and hence $\dim \mathrm{B} = \dim H_{\mathrm{B}}$.

Remark 10: In fact, we can convert any **GBS** B of the form

$$\widetilde{x}(t+1) = \sum_{\sigma \in \Sigma} (A_\sigma \widetilde{\mathbf{x}}(t) + K_\sigma \widetilde{\mathbf{v}}(t)) \mathbf{u}_\sigma(t)$$

$$\mathbf{y}(t) = C\widetilde{\mathbf{x}}(t) + D\widetilde{\mathbf{v}}(t)$$

to a jump-Markov linear system of the type defined in [41]:

$$\mathbf{H}: \begin{cases} \mathbf{x}(t+1) & = F_{\boldsymbol{\theta}(t)}\mathbf{x}(t) + G_{\boldsymbol{\theta}(t)}\widetilde{\mathbf{v}}(t) \\ \mathbf{y}(t) & = \ H_{\boldsymbol{\theta}(t)}\mathbf{x}(t) + L_{\boldsymbol{\theta}(t)}\widetilde{\mathbf{v}}(t) \end{cases}. \tag{64}$$

where $\mathbf{x}(t) = \begin{bmatrix} \mathbf{z}^T_1(t), & \dots, \mathbf{z}^T_d(t) \end{bmatrix}^T$, $\mathbf{z}_q(t) = A_{(q,\boldsymbol{\theta}(t-1))}\widetilde{\mathbf{x}}(t-1) + K_{(q,\boldsymbol{\theta}(t-1))}\widetilde{\mathbf{v}}(t-1)$, $q \in Q$, and

$$L_q = D$$

$$H_q = \begin{bmatrix} \delta_{q,1}C, & \delta_{q,2}C, & \cdots & \delta_{q,d}C \end{bmatrix}$$

$$F_q = \begin{bmatrix} \delta_{1,q}A_{(1,1)}, & \delta_{2,q}A_{(1,2)}, & \dots & \delta_{d,q}A_{(1,d)} \\ \delta_{1,q}A_{(2,1)}, & \delta_{2,q}A_{(2,2)}, & \dots & \delta_{d,q}A_{(2,d)} \\ \vdots & \vdots & \dots & \vdots \\ \delta_{1,q}A_{(d,1)}, & \delta_{2,q}A_{(d,2)}, & \dots & \delta_{d,q}A_{(d,d)} \end{bmatrix} \quad G_q = \begin{bmatrix} K_{(1,q)} \\ K_{(2,q)} \\ \vdots \\ K_{(d,q)} \end{bmatrix}$$

where $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$ for all $i, j \in Q$. If B satisfies Assumptions 2, and it is a realization of $\mathbf{y}$, then H is also a ealization of $\mathbf{y}$ and it satisfies Assumptions 7.

Recall the notion of *minimality* of a linear system realization. In particular, recall that a realization by a linear system is minimal if and only if it is reachable and observable. In this subsection, we will formulate similar concepts for GJMLS with fully observed discrete. We first define the notions of reachability and observability for a GJMLS. We then show that a realization by a GJMLS is minimal if and only if it is reachable and observable.

In order to formulate the conditions more precisely, we will need to introduce some notation. In particular, we need to define reachability and observability matrices for GJMLS. To that end, let H be a given GJMLS of the form (50) that satisfies Assumptions 7. Let N be the dimension of H, i.e. $N = \dim H$, and for all $(q_1, q_2) \in Q \times Q = \Sigma$ let

$$\begin{aligned} G_{q_1,q_2} =& E[\mathbf{x}(t)\mathbf{y}^T(t-1)\chi(\boldsymbol{\theta}(t) = q_2, \boldsymbol{\theta}(t-1) = q_1)] = \\ =& p_{q_1,q_2}(M_{q_1,q_2}P_{q_1}C_{q_1}^T + B_{q_1,q_2}Q_{q_1}D_{q_1}^T) \in \mathbb{R}^{q_2 \times p}. \end{aligned} \tag{65}$$

Recall the definition of $L \subset Q \times Q = \Sigma$ from (58).

Notation 9 (Matrix products): We define the following notation for the products of matrices $M_{q_1,q_2} \in \mathbb{R}^{n_{q_1} \times n_{q_2}}$. For any admissible word $w = (q_1, q_2) \cdots (q_{k-1}, q_k) \in L$, where $k > 2$ and $q_1, \ldots, q_k \in Q$, let

$$M_w = M_{q_{k-1},q_k} M_{q_{k-2},q_{k-1}} \cdots M_{q_1,q_2} \in \mathbb{R}^{n_{q_k} \times n_{q_1}} \tag{66}$$

If $w = \epsilon$, then M_ϵ is an identity matrix, dimension of which depends on the context it is used in. If $w \notin L$, then M_w denotes the zero matrix.

Notation 10: For each $q \in Q$, $L^q(N)$ be the set of all words in $w \in L$ such that $|w| \leq N$ and $w = v(q_1, q)$ for some $q_1 \in Q$ and $v \in L$.

Definition 19 (Reachability of a GJMLS): For each discrete state $q \in Q$, define the matrix

$$\mathcal{R}_{H,q} = [M_v G_{q_1,q_2} \mid q_1 \in Q, q_2 \in Q, (q_1, q_2)v \in L^q(N)] \in \mathbb{R}^{n_q \times |L^q(N)|p}. \tag{67}$$

We will say that the GJMLS H is *reachable*, if for each discrete state $q \in Q$, $\operatorname{rank}(\mathcal{R}_{H,q}) = n_q$. Notice that the matrix $R_{H,q}$ is analogous to the controllability matrix for linear systems.

Notation 11: For each $q \in Q$, let $L_q(N)$ be the set of all words in L of length at most N that begin in some pair whose first component is q, i.e. $L_q(N)$ is the set of all words in $w \in L$ such that $|w| \leq N$ and $w = (q, q_2)v$ for some $q_2 \in Q$ and $v \in L$.

Definition 20 (Observability of a GJMLS): For each discrete state $q \in Q$, define the matrix

$$\mathcal{O}_{H,q} = [(C_{q_k} M_v)^T \mid q_{k-1} \in Q, q_k \in Q, v(q_{k-1}, q_k) \in L_q(N)]^T \in \mathbb{R}^{|L_q(N)|p \times n_q}. \tag{68}$$

We will say that a GJMLS H is *observable*, if for each discrete state $q \in Q$, $\mathrm{rank}\ (O_{H,q}) = n_q$. Notice that the matrix $O_{H,q}$ plays a role similar to the observability matrix for linear systems.

Recall from (60) the definition of the **GBS** B_H associated with a GJLS H. Recall from Definition 10 the definition of the representation R_{B_H} associated with the **GBS** B_H. We will denote R_{B_H} by R_H and we will call it the representation associated with the GJMLS H. Recall the definition of reachability of a representation along with the definition of the space O_{R_H} defined in (3). Observability and reachability of a GJMLS H can be characterized in terms of the observability and reachability of the corresponding representation R_H as follows.

Lemma 23: The GJMLS H is reachable if and only if R_H is reachable, and H is observable if and only if R_H is observable.

The lemma above implies that observability and reachability of a GJMLS can be checked by a numerical algorithm.

Definition 21 (Morphism of GJMLSs): Let H be a GJMLS of the form (50) and let $\hat{H}$ is another GJMLS realization of $\mathbf{y}$ given by

$$\begin{aligned} \hat{\mathbf{x}}(t+1) &= \hat{M}_{\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1)} \hat{\mathbf{x}}(t) + \hat{B}_{\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1)} \hat{\mathbf{v}}(t) \\ \hat{\mathbf{y}}(t) &= \hat{C}_{\boldsymbol{\theta}(t)} \hat{\mathbf{x}}(t) + \hat{D}_{\boldsymbol{\theta}(t)} \hat{\mathbf{v}}(t), \end{aligned} \tag{69}$$

where the dimension of the continuous state-space of $\hat{H}$ corresponding to the discrete state q is $\hat{n}_q$. A morphism from H to $\hat{H}$ is a collection of matrices $T = \{T_q \in \mathbb{R}^{\hat{n}_q \times n_q}\}_{q \in Q}$ such that for all $q_1, q_2 \in Q$.

$$T_{q_2} M_{q_1,q_2} = \hat{M}_{q_1,q_2} T_{q_1}, \quad C_{q_1} = \hat{C}_{q_1} T_{q_1}, \quad T_{q_2} G_{q_1,q_2} = \hat{G}_{q_1,q_2}, \tag{70}$$

where G_{q_1,q_2} is defined in (65), and

$$\hat{G}_{q_1,q_2} = \sqrt{p_{q_1,q_2}} (\hat{M}_{q_1,q_2} \hat{P}_{q_1} C_{q_1}^T + \hat{B}_{q_1,q_2} \hat{Q}_{q_1} \hat{D}_{q_1}^T, \tag{71}$$

where $\hat{P}_{q_1} = E[\hat{\mathbf{x}}(t)\hat{\mathbf{x}}^T(t)\chi(\boldsymbol{\theta}(t) = q_1)]$ and $\hat{Q}_{q_1} = E[\hat{\mathbf{v}}(t)\hat{\mathbf{v}}^T(t)\chi(\boldsymbol{\theta}(t) = q_1)]$.

T will be called an isomorphism, if for all $q \in Q$, $n_q = \hat{n}_q$ and T_q is invertible.
Note that $T = (T_q)$ is an GJMLS isomorphism, if and only if the map $\mathbf{S}_T : R_H \to R_{\hat{H}}$ is a representation isomorphism, where $\mathbf{S}_T = \sum_{q \in Q} \mathbf{I}_q T_q \mathbf{I}_q^T$.

We are now ready to state the theorem on minimality of a GJMLS realization.

Theorem 10 (Minimality of a realization by a GJMLS): Let the GJMLS H be a realization of $\mathbf{y}$ of the form (50) and assume that H satisfies Assumption 7. Then, the GJMLS H is a minimal realization of $\mathbf{y}$ if and only if it is reachable and observable. If $\hat{H}$ is another minimal GJMLS realization of $\mathbf{y}$ such that $\hat{H}$ satisfies Assumption 7, then $\hat{H}$ and H are isomorphic.

Remark 11: Notice that in (70) we do not require any relationship between B_{q_1,q_2} and $\hat{K}_{q_1,q_2}$. This is consistent with the situation for linear stochastic systems.

Remark 12 (Realization Algorithms): It is clear that reachability and observability, and hence minimality, of a GJLS can be checked numerically. It is also easy to see that the Algorithm 2 can be adapted to obtain a weak realization H of $\mathbf{y}$.

C. Proofs of the results on realization theory of GJMLSs

Below we present the proofs of the statements presented in §V-B. In addition, we present the proof of Lemma 19.

Proof of Lemma 20: We show that $\widetilde{\mathbf{y}}$ satisfies the parts of Assumptions 3 one by one and then we show that the statement of the lemma for the innovation process of $\widetilde{\mathbf{y}}$ is true.

$\widetilde{\mathbf{y}}$ is an RC process Define the matrix $\mathbf{M}_q \in \mathbb{R}^{p \times dp}$ as

$$\mathbf{M}_q = \begin{bmatrix} \mathbf{O}_{p,(q-1)p}, & I_p, & \mathbf{O}_{p,(d-q-1)p} \end{bmatrix}.$$

It then follows that $\mathbf{z}^{\widetilde{\mathbf{y}}}_w(t) = \mathbf{M}_q^T \mathbf{z}_w^T(t)$ if $w = (q,q_1)v$ for some $q, q_1 \in Q$, $v \in \Sigma^+$. Moreover, notice that $\widetilde{\mathbf{y}}(t)(\mathbf{z}^{\widetilde{\mathbf{y}}}_w(t))^T = \mathbf{M}_{q_2}^T \mathbf{y}(t)\mathbf{z}_w^T(t)\mathbf{M}_q$, where $q \in Q$ is the first component of the first letter of w and $q_2 \in Q$ is the second component of the last letter of w. It is then easy to check that if $\mathbf{y}$ is a **RC** process, then so is $\widetilde{\mathbf{y}}$. In order to see that $\mathbf{y}$ is an **RC** process, notice that the first requirement of Assumption 8 implies that $\mathbf{y}$ satisfies Part 1 of Definition 5. That $\mathbf{y}$ satisfies Part 2 of Definition 5 can be shown as follows. If $w \notin L$, then $\mathbf{u}_w(t) = 0$ by definition of L. Let $w, v \in \Sigma^*$ be such that $w\sigma, v\sigma' \in L$ and $|w| > 0$. It is clear that $\mathbf{z}_{w\sigma}(t)\mathbf{z}_{v\sigma'}(t)$ contains a term $\mathbf{u}_\sigma(t)\mathbf{u}_{\sigma'}(t)$ and the latter term is zero, if $\sigma \neq \sigma'$. Assume that $\sigma = \sigma' = (q_1, q_2)$. Then, using the definition of $\mathbf{z}_w(t), \mathbf{z}_v(t)$, $E[\mathbf{z}_{w\sigma}(t)\mathbf{z}_{v\sigma}^T(t)] = \frac{1}{p_{q_1,q_2}} E[\mathbf{z}_w(t-1)\mathbf{z}_v^T(t-1)\chi(\boldsymbol{\theta}(t-1) = q_1, \boldsymbol{\theta}(t) = q_2)]$. Here, for $v = \epsilon$, $\mathbf{z}_v(t-1) = \mathbf{y}(t-1)$. Using the assumption on conditional independence, $E[\mathbf{z}_w(t-1)\mathbf{z}_v^T(t-1)\chi(\boldsymbol{\theta}(t-1) = q_1, \boldsymbol{\theta}(t) = q_2) \mid \mathcal{D}_{t-1}] = E[\chi(\boldsymbol{\theta}(t-1) = q_1, \boldsymbol{\theta}(t) = q_2) \mid \mathcal{D}_{t-1}]E[\mathbf{z}_w(t-1)\mathbf{z}_v^T(t-1) \mid \mathcal{D}_{t-1}] = p_{q_1,q_2}\chi(\boldsymbol{\theta}(t-1) = q_1)E[\mathbf{z}_w(t-1)\mathbf{z}_v^T(t-1) \mid \mathcal{D}_{t-1}]$. Note that

$E[\mathbf{z}_w(t-1)\mathbf{z}_v^T(t-1)] = E[E[\mathbf{z}_w(t-1)\mathbf{z}_v^T(t-1) \mid \mathcal{D}_{t-1}]]$. Moreover, $w\sigma \in L$, $|w| > 0$ implies that q_1 is the last component of the last letter of w. Hence, $\mathbf{z}_w(t-1)\chi(\boldsymbol{\theta}(t-1) = q_1) = \mathbf{z}_w(t-1)$. From the properties of conditional expectation it follows then that $\chi(\boldsymbol{\theta}(t-1) = q_1)E[\mathbf{z}_w(t-1)\mathbf{z}_v^T(t-1) \mid \mathcal{D}_{t-1}] = E[\mathbf{z}_w(t-1)\chi(\boldsymbol{\theta}(t-1) = q_1)\mathbf{z}_v^T(t-1) \mid \mathcal{D}_{t-1}] = E[\mathbf{z}_w(t-1)\mathbf{z}_v(t-1) \mid \mathcal{D}_{t-1}]$. Combining all these remarks, it follows that $E[\mathbf{z}_w(t-1)\mathbf{z}_v^T(t-1)\chi(\boldsymbol{\theta}(t-1) = q_1, \boldsymbol{\theta}(t) = q_2)] = p_{q_1,q_2}E[\mathbf{z}_w(t-1)\mathbf{z}_v^T(t-1)]$ and hence $E[\mathbf{z}_{w\sigma}(t)\mathbf{z}_{v\sigma}^T(t)] = E[\mathbf{z}_w(t-1)\mathbf{z}_v^T(t-1)]$. That is, $\mathbf{y}$ satisfies Part 2 of Definition 5. By Remark 1, $\mathbf{y}$ then satisfies Part 3 of Definition 5 too.

$\Psi_{\widetilde{\mathbf{y}}}$ is rational and square summable

From the discussion above it follows that if $\Psi_{\widetilde{\mathbf{y}}} = \{T_{(\sigma,i)} \mid \sigma \in \Sigma, i = 1, \ldots, dp\}$, then for all $q \in Q$, $l = 1, \ldots, p$, $T_{\sigma,p(q-1)+l}(v)$ can be written as follows. If q is the first components of σ, then $T_{\sigma,p(q-1)+l}(v) = \mathbf{M}_{q_2}^T S_{\sigma,l}(v)\mathbf{M}_q$ for all $v \in \Sigma^*$ where $\sigma v = s(q_1, q_2)$ for some $s \in \Sigma^*$, $q_1 \in Q$. If q is not the first components of σ, then $T_{\sigma,p(q-1)+l}(v) = 0$. It is not difficult to construct a rational representation of $\Psi_{\widetilde{\mathbf{y}}}$ based on such a representation of $\Psi_{\mathbf{y}}$. Indeed, assume that $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma\in\Sigma}, B, C)$ is a representation of $\Psi_{\mathbf{y}}$. Define $\hat{\mathcal{X}} = \mathbb{R}^{dn}$ and define $\mathbf{H}_q \in \mathbb{R}^{n\times nd}$ by

$$\mathbf{H}_q = \begin{bmatrix} \mathbf{O}_{n,(q-1)n}, & I_n, & \mathbf{O}_{n,(d-q-1)n} \end{bmatrix}.$$

Let $\hat{A}_{(q_1,q_2)} = \mathbf{H}_{q_2}^T A_{(q_1,q_2)}\mathbf{H}_{q_1}$, $\hat{B}_{(q_1,q_2),p(q_1-1)+i} = \mathbf{H}_{q_2}B_{(q_1,q_2),i}$, $i = 1, \ldots, p$ and let $\hat{B}_{(q_1,q_2),l} = 0$ for all $l \neq p(q_1-1)+i$ for some $i = 1, \ldots, p$. Finally, define $\hat{C} = \begin{bmatrix} \mathbf{H}_1^T C^T, & \ldots, & \mathbf{H}_d^T C^T \end{bmatrix}^T$, i.e. $\hat{C}$ is a block diagonal matrix, whose (q,q)th $p\times n$ block equals C. It is then easy to see that $\hat{R} = (\mathbb{R}^{nd}, \{\hat{A}_\sigma\}_{\sigma\in\Sigma}, \hat{B}, \hat{C})$ is a representation of $\Psi_{\widetilde{\mathbf{y}}}$. Square summability of $\Psi_{\widetilde{\mathbf{y}}}$ follows easily from that of $\Psi_{\mathbf{y}}$, by taking into account the relationship $T_{\sigma,p(q-1)+l}(v) = \mathbf{M}_{q_2}^T S_{\sigma,l}(v)\mathbf{M}_q$, $v \in \Sigma^*$, $l = 1, \ldots, p$, $q \in Q$, $\sigma \in \Sigma$, q is the first letter of σ.

Proof of the formula for $\widetilde{\mathrm{e}}(t)$

Finally, from the discussion above it follows that the Hilbert-space spanned by the entries of $\{\mathbf{z}_w(t) \mid w \in \Sigma^+\}$ coincides with that of spanned by the elements of $\{\mathbf{z}_w^{\widetilde{\mathbf{y}}}(t) \mid w \in \Sigma^+\}$. If $\mathbf{z}(t) = E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w(t) \mid w \in \Sigma^+\}]$, then define $s(t) = \begin{bmatrix} \mathbf{z}^T(t)\chi(\boldsymbol{\theta}(t) = 1), & \ldots, & \mathbf{z}^T(t)\chi(\boldsymbol{\theta}(t) = d) \end{bmatrix}^T$. We claim that $s(t) = E_l[\widetilde{\mathbf{y}}(t) \mid \{\mathbf{z}_w^{\widetilde{\mathbf{y}}}(t) \mid w \in \Sigma^+\}]$. Indeed, $s(t)$ belongs to the Hilbert-space spanned by the entries of $\{\mathbf{z}_w^{\widetilde{\mathbf{y}}}(t) \mid w \in \Sigma^+\}$. Moreover, if q is the first component of the first letter of w and q_1 is the second component of the last letter of w, then $E[\widetilde{\mathbf{y}}(t)(\mathbf{z}_w^{\widetilde{\mathbf{y}}}(t))^T] = \mathbf{M}_{q_1}^T E[\mathbf{y}(t)\mathbf{z}_w^T(t)]\mathbf{M}_q = \mathbf{M}_{q_1}^T E[\mathbf{z}(t)\mathbf{z}_w^T(t)]\mathbf{M}_q = E[s(t)(\mathbf{z}_w^{\widetilde{\mathbf{y}}}(t))^T]$. From this, the claim of the lemma regarding $\widetilde{\mathbf{e}}(t)$ follows easily. ■

Proof of Lemma 21: First, we show that B_H is well-defined and the output of B_H equals $\widetilde{\mathbf{y}}$. For this, we have to show that $\hat{\mathbf{x}}(t)$ indeed satisfy (60). From this and the definition of $\widetilde{\mathbf{y}}(t)$ it follows easily that the outputs of H and B_H are equal. We show that the various parts of Definition 2 hold one by one. First of all, Assumption 4 on ergodicity of $\boldsymbol{\theta}$ means that the framework of Section III can be used as it was explained before.

$\widetilde{\mathbf{v}}(t)$ satisfies Part 1 of Assumption 2

First, we will show that $\mathbf{v}$ is an **RC** process. From Part 1, Assumption 7 it follows that $\mathbf{z}^{\mathbf{v}}_w(t)$ is zero-mean. Moreover, for any $w,v \in \Sigma^+$, $|w|=k<|v|=l$, $E[\mathbf{z}^{\mathbf{v}}_w(t)(\mathbf{z}^{\mathbf{v}}_v(t))^T] = E[E[\mathbf{z}^{\mathbf{v}}_w(t)(\mathbf{z}^{\mathbf{v}}_v(t))^T \mid D_{t-k,t}]]$. If $v=ss^{'}$ for some $s,s^{'} \in \Sigma^+$, $|s^{'}|=|w|$ and $w \neq s^{'}$ then clearly $\mathbf{u}_v(t)\mathbf{u}_w(t)=0$ and hence $E[\mathbf{z}^{\mathbf{v}}_w(t)(\mathbf{z}^{\mathbf{v}}_v(t))^T]=0$. Otherwise, if $w=s^{'}$, then notice that $\mathbf{u}_v(t)$ is a product of variables $\chi(\boldsymbol{\theta}(t-r)=q)$ for some $q \in Q$ and $r=0,\ldots,l-1$ multiplied by a constant. Hence, by Part 1 of Assumption 7 $E[\mathbf{z}^{\mathbf{v}}_w(t)\mathbf{z}^{\mathbf{v}}_v(t) \mid D_{t-l,t}] = \frac{1}{p_w}E[\mathbf{u}_v(t) \mid D_{t-k}]E[\mathbf{v}(t-k)\mathbf{v}(t-l))^T \mid D_{t-l,t}]=0$. Hence, $E[\mathbf{z}^{\mathbf{v}}_w(t)(\mathbf{z}^{\mathbf{v}}_v(t))^T]=0$ for any $w \neq v$, $|w| \neq |v|$. If $w \neq v$ but $|w|=|v|$, the $\mathbf{u}_w(t)\mathbf{u}_v(t)=0$ and hence $E[\mathbf{z}^{\mathbf{v}}_w(t)(\mathbf{z}^{\mathbf{v}}_v(t))^T]=0$. Finally, if $w=v$ and $|w|=|v|=k$, then using Assumption 7, Part 2 yields $E[\mathbf{z}^{\mathbf{v}}_w(t)(\mathbf{z}^{\mathbf{v}}_w(t))^T \mid D_{t-k,t}] = \frac{1}{p_w}E[\mathbf{u}_w(t) \mid D_{t-k,t}]E[\mathbf{v}(t-k)\mathbf{v}^T(t-k) \mid D_{t-k}] = \chi(\boldsymbol{\theta}(t-k)=q)E[\mathbf{v}(t-k)\mathbf{v}^T(t-k) \mid D_{t-k,t}] = E[\mathbf{v}(t-k)\mathbf{v}^T(t-k)\chi(\boldsymbol{\theta}(t-k)=q) \mid D_{t-k,t}]$ where is assumed to be of the form $w=(q,q_1)s$ for some $q,q_1 \in Q$, $s \in \Sigma^*$. Hence, $E[\mathbf{z}^{\mathbf{v}}_w(t)\mathbf{z}^{\mathbf{v}}_w(t))^T] = E[E[\mathbf{z}^{\mathbf{v}}_w(t)(\mathbf{z}^{\mathbf{v}}_w(t))^T \mid D_{t-k,t}]] = E[\mathbf{v}(t-k)\mathbf{v}^T(t-k)\chi(\boldsymbol{\theta}(t-k))]$ and the latter does not depend t by Assumption 7, Part 2. Hence, we have shown that $E[\mathbf{z}^{\mathbf{v}}_w(t)\mathbf{z}^{\mathbf{v}}_w(t))^T]$ does not depend on t. Finally, notice that $E[\mathbf{v}(t)(\mathbf{z}^{\mathbf{v}}_w(t))^T|D_{t-k,t}] = \frac{1}{\sqrt{p_w}}\mathbf{u}_w(t)E[\mathbf{v}(t)\mathbf{v}^T(t-k) \mid D_{t-k,t}]=0$ does not depend on t and hence $E[\mathbf{v}(t)(\mathbf{z}^{\mathbf{v}}_w(t))^T]=0$ also does not depend on t. Hence, $\mathbf{v}(t)$ satisfies Part 1 of Definition 5. Finally, from the discussion above it follows that $T_{w,v}=0$ for $w \neq v$, and $T_{w,w}=0$ for $w \notin L$ and $T_{w,w}=E[\mathbf{v}(t)\mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t)=q)]$ where $q \in Q$ is such that $w=(q,q_1)s$ for some $q_1 \in Q$, $s \in \Sigma^*$. This implies that Part 2 of Definition 5 is satisfied. By Remark 1 this already implies Part 3 of Definition 5. Hence, $\mathbf{v}$ is indeed an **RC** process. Next, we show that $\widetilde{\mathbf{v}}$ is an **RC** process too. It follows that all the entries of $\mathbf{z}^{\widetilde{\mathbf{v}}}_w(t)$ are zero except the one which corresponds to the qth block of p rows, where q is the first components of the first letter of w. The latter entry equals $\mathbf{z}^{\mathbf{v}}_w(t)$, It is then easy to see that Part 1 of Definition 5 hold. Consider any two $w,v \in \Sigma^+$, and let the first component of the first letter of w and v be respectively $q_1,q_2 \in Q$. Then non-zero $p \times p$ block of $E[\widetilde{\mathbf{v}}(t)(\mathbf{z}^{\widetilde{\mathbf{v}}}_w(t))^T]$ is the one indexed by (q_1, q_1). Similarly, the only non-zero $p \times p$ block of $E[z^{\widetilde{\mathbf{v}}}_v(t)(\mathbf{z}^{\widetilde{\mathbf{v}}}_w(t))^T]$ is the is the one indexed by $q_1 \times q_2$. Here, we viewed both matrices as $d \times d$ matrices of $p \times p$ block. The respective non-zero entries are $E[\mathbf{z}^{\mathbf{v}}(t)(\mathbf{z}^{\mathbf{v}}_w(t)^T]$ and $E[\mathbf{z}^{\mathbf{v}}_v(t)(\mathbf{z}^{\mathbf{v}}_w(t))^T]$. Since $\mathbf{v}$ is **RC**, it follows that $\widetilde{\mathbf{v}}$ satisfies Part 2 and 3 of Definition 5.

$\widetilde{\mathbf{v}}(t)$ satisfies Part 2 of Assumption 2

The orthogonality of $\mathbf{z}^{\widetilde{\mathbf{v}}}_w(t)$ and $\mathbf{z}^{\widetilde{\mathbf{v}}}_v(t)$ for $w \neq v$ follows from the proof that $\widetilde{\mathbf{v}}(t)$ satisfies Part 1 of Assumption 2.

$\mathbf{x}(t)$ and $\mathbf{v}(t)$ satisfy Part 3 of Assumption 2 The first statement of Part 3 of Assumption 2 is a direct consequence of Part 3 of Assumption 7 and the fact that the sum of entries of $\widetilde{\mathbf{v}}(t)$ equals $\mathbf{v}(t)$.

Part 5 of Assumption 2 holds From the construction of $A_{(q_1,q_2)}$ it follows that the only non-zero column of $A_{(q_1,q_2)}$ is the one indexed by $j = (\sum_{q=1}^{q_1-1} n_q) + 1, \ldots, \sum_{q=1}^{q_1} n_q$, and the only non-zero rows are the ones indexed by $i = (\sum_{q=1}^{q_2-1} n_q) + 1, \ldots, \sum_{q=1}^{q_2} n_q$. Hence, $A_{(q_3,q_4)}A_{(q_1,q_2)}$ is necessarily zero if $q_2 \neq q_3$. The latter condition is equivalent to $(q_1, q_2)(q_3, q_4) \notin L$. Similarly, the only non-zero rows of $\hat{K}_{(q_1,q_2)}$ are the ones indexed by $i = (\sum_{q=1}^{q_2-1} n_q) + 1, \ldots, \sum_{q=1}^{q_2} n_q$, so again $A_{(q_3,q_4)}\hat{K}_{(q_1,q_2)} = 0$ for $q_3 \neq q_2$.

Part 4 of Assumption 2 holds It is easy to see that $\sum_{(q_1,q_2)\in\Sigma} p_{q_1,q_2} A^T_{(q_1,q_2)} \otimes A^T_{(q_1,q_2)} = \widetilde{M}$ and hence Part 4 of Assumption 2 follows directly from Part 5 of Assumption 7.

Proof that $\hat{D}E[\hat{\mathbf{v}}^T(t)\hat{\mathbf{v}}(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t{+}1) = q_2)]\hat{D}^T > 0$. Notice that $\hat{D}E[\hat{\mathbf{v}}^T(t)\hat{\mathbf{v}}(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)]\hat{D}^T = D_{q_1}E[\mathbf{v}(t)\mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)]D^T_{q_1}$. From Part 1 it follows that $E[\mathbf{v}(t)\mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2) \mid D_t] = E[\mathbf{v}(t)\mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t) = q_1)]E[\chi(\boldsymbol{\theta}(t+1) = q_2) \mid D_t] \mid D_t] = p_{q_1,q_2}E[\mathbf{v}(t)\mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t) = q_1) \mid D_t]$ and hence $E[\mathbf{v}(t)\mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)] = p_{q_1,q_2}E[\mathbf{v}(t)\mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t) = q)] = p_{q_1,q_2}Q_{q_1}$. Hence, $DQ_{(q_1,q_2)}D^T = p_{q_1,q_2}D_{q_1}Q_{q_1}D^T_{q_1}$. Since $p_{q_1,q_2} > 0$, by Part 6 of Assumption 7, the above matrix is strictly positive definite. ■

Proof of Lemma 22: The first, we argue that H_{B} is well-defined and its output equals $\mathbf{y}$. The only non-trivial thing is to prove that $\hat{\mathbf{x}}(t)$ is well defined and that the output of H_{B} is $\mathbf{y}$. First, notice that Lemma 20 implies that

$$K_{(\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1)}\widetilde{\mathbf{e}}(t) = K_{(\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1))}\mathbf{M}_{\boldsymbol{\theta}(t)}\mathbf{e}(t)$$

It then follows that

$$\hat{\mathbf{x}}(t+1) = \Pi^T_{\boldsymbol{\theta}(t+1)}\mathbf{x}(t+1) = \Pi^T_{\boldsymbol{\theta}(t+1)}A_{(\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1))}\mathbf{x}(t) + \Pi^T_{\boldsymbol{\theta}(t+1)}K_{(\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1)}\widetilde{\mathbf{e}}(t) \tag{72}$$

Note that from Lemma 2 it follows that $\mathbf{x}$ is an **RC** process and that $\mathbf{x}(t)\chi(\boldsymbol{\theta}(t) = q)$ belongs to $\mathcal{X}_q = \mathrm{Im}\Pi_q$ almost surely. Hence, $\Pi_{\boldsymbol{\theta}(t)}\Pi_{\boldsymbol{\theta}(t)}^T\mathbf{x}(t) = \mathbf{x}(t)$ and thus $A_{\boldsymbol{\theta}(t+1),\boldsymbol{\theta}(t)}\mathbf{x}(t) = A_{\boldsymbol{\theta}(t+1),\boldsymbol{\theta}(t)}\Pi_{\boldsymbol{\theta}(t)}\hat{\mathbf{x}}(t)$. Substituting this into (72) yields that

$$\hat{\mathbf{x}}(t+1) = M_{\boldsymbol{\theta}(t+1),\boldsymbol{\theta}(t)}\hat{x}(t) + \widetilde{K}_{\boldsymbol{\theta}(t+1),\boldsymbol{\theta}(t)}\mathbf{e}(t).$$

Hence, the first equation of H_{B} holds. Notice that $\mathbf{M}_{\boldsymbol{\theta}(t)}\widetilde{\mathbf{y}}(t) = \mathbf{y}(t)$ and $\mathbf{M}_{\boldsymbol{\theta}(t)}\widetilde{\mathbf{e}}(t) = \mathbf{e}(t)$. Moreover, by the discussion above it follows that $C\mathbf{x}(t) = C\Pi_{\boldsymbol{\theta}(t)}\hat{\mathbf{x}}(t)$. By multiplying $\widetilde{\mathbf{y}}(t) = C\hat{x}(t) + \widetilde{\mathbf{e}}(t)$ with $\mathbf{M}_{\boldsymbol{\theta}(t)}$ we obtain

$$\mathbf{y}(t) = C_{\boldsymbol{\theta}(t)}\hat{\mathbf{x}}(t) + \mathbf{e}(t).$$

That is, $\mathbf{y}$ is indeed the output of H_{B}.

Next, we show that H_{B} satisfies each of the assumptions of Assumption 7.

Part 1 of Assumption 7 Since $\mathbf{y}(t)$ is the output of B, by Theorem 4 it is **RC**. Moreover, because B satisfies Assumption 2, the innovation process is **RC** too. Hence, $E[\mathbf{z}^{\mathbf{e}}_w(t+|w|)] = 0$ for any $w \in \Sigma^+$, which implies that $E[\mathbf{e}(t) \mid \mathcal{D}_{t+k}] = 0$ for any $k \geq 0$. Notice that for any $w \in \Sigma^+$, $|w| = l$ the variables $\mathbf{u}_w(t+1)$ generate the σ-algebra $D_{t-l,t}$. Notice that for any $w \in \Sigma^+$, $|w| = l-1$, $\sigma \in \Sigma$ $E[\mathbf{e}(t)\mathbf{e}^T(t-l)\mathbf{u}_{w\sigma}(t+1)] = \sqrt{p_{w\sigma}}E[\mathbf{z}^{\mathbf{e}}_{\sigma}(t+1)(\mathbf{z}^{\mathbf{e}}_{w\sigma}(t+1))^T] = 0$. Hence, $E[\mathbf{e}(t)\mathbf{e}^T(t-l) \mid \mathcal{D}_{t,t-l}] = 0$. Finally $E[\mathbf{e}(t)\mathbf{e}^T(t)\chi(\boldsymbol{\theta}(t) = q)] = \sum_{q_2 \in Q} E[\mathbf{z}^{\mathbf{e}}_{(q,q_2)}(t+1)(\mathbf{z}^{\mathbf{e}}_{(q,q_2)}(t+1))^T]$ and the latter does not depen on t due to the fact that $\mathbf{e}(t)$ is **RC**.

Part 2 of Assumption 7 Let $\mathcal{F}_1$ be the σ-algebra generated by the variables $\{y(t-l)\}_{l\geq 0}$ and denote by $\mathcal{F}_1 \vee \mathcal{D}_t$ the smallest σ-algebra which contains $\mathcal{F}_1$ and $\mathcal{D}_t$. Let $\mathcal{F}_2$ be the σ-algebra generated by $\{\boldsymbol{\theta}(t+l)\}_{l>0}$ and notice that by assumption $\mathcal{F}_2$ and $\mathcal{F}_1$ are conditionaly independent w.r.t. $\mathcal{D}_t$. From the elementary properties of conditional independence and the fact that $\mathcal{F}_1$ and $\mathcal{F}_2$ are conditionally independent w.r.t. $\mathcal{D}_t$ it follows that $\mathcal{F}_1 \vee \mathcal{D}_t$ and $\mathcal{F}_2$ are also conditionally independent w.r.t. $\mathcal{D}_t$.

Hence, it is enough to show that for $l \geq 0$, $\mathbf{e}(t-l)$ is $\mathcal{F}_1 \vee \mathcal{D}_t$ measurable. From this and the discussion above it then follows that the σ-algebra generated by $\{e(t-l)\}_{l=0}^{\infty}$ and $\mathcal{F}_2$ are conditionaly independent w.r.t $\mathcal{D}_t$. Notice that $\mathbf{e}(t-l)$ belongs to the Hilbert-space generated by $\{\mathbf{y}(t-l), \mathbf{z}_w(t-l) \mid w \in \Sigma^+\}$, and hence by Lemma 1, $\mathbf{e}(t-l)$ is measurable w.r.t the σ-algebra generated by $\{y(t-l), \mathbf{z}_w(t-l) \mid w \in \Sigma^+\}$. The latter σ-algebra is contained in $\mathcal{F}_1 \vee \mathcal{D}_t$ and hence $\mathbf{y}$ is $\mathcal{F}_1 \vee \mathcal{D}_t$ measurable, as required.

Part 3 of Assumption 7 This is a direct consequence of Part 3 of Assumption 2.

Part 4 of Assumption 7 This a direct consequence of Assumption 2.

Part 5 of Assumption 7 It then follows that $M_{q_1,q_2} \otimes M_{q_1,q_2} = (\Pi_{q_2}^T \otimes \Pi_{q_2}^T)(A_{(q_1,q_2)} \otimes A_{(q_1,q_2)})(\Pi_{q_1} \otimes \Pi_{q_1})$. Let $P = (P_1, \dots, P_d)$ a d tuple of matrices $P_q \in \mathbb{R}^{n_q \times n_q}$ such that if P is interpreted as a $\sum_{q \in Q} n_q^2$ vector $\phi(P)$, then $\widetilde{M}^T \phi(P) = \lambda \phi(P)$ for some $\lambda \in \mathbb{C}$. It then follows that $\lambda P_q = \sum_{r \in Q} p_{r,q} M_{r,q} P_r M_{r,q}^T$. Notice that with $\hat{P}_q = \Pi_q P_q \Pi_q^T = (\Pi_q \otimes \Pi_q)\phi(P)$, $\lambda P_q = \sum_{r \in Q} p_{r,q} \Pi_q^T A_{(r,q)} \hat{P}_r A_{(r,q)}^T \Pi_q$. By applying from the left Π_q and from the right Π_q^T to both sides of the equation, we get $\lambda \hat{P}_q = \sum_{r \in Q} p_{r,q} \Pi_q \Pi_q^T A_{(r,q)} \hat{P}_r A_{(r,q)}^T \Pi_q \Pi_q^T$. Notice that $\mathcal{X}_q = \mathrm{Im}\Pi_q$ and that $A_{(r,q)}\mathcal{X}_r \subseteq \mathcal{X}_q$. Hence, $A_{(r,q)}\Pi_r = \Pi_q S$ for some $S \in \mathbb{R}^{n_r \times n}$. By exploiting $\Pi_q^T \Pi_q = I_{n_q}$, it follows that $\Pi_q \Pi_q^T A_{(r,q)} \Pi_r = A_{(r,q)}^T \Pi_r$. Thus, by taking into account that $\hat{P}_r = \Pi_r P_r \Pi_r^T$, $r \in Q$,

$$\lambda \hat{P}_q = \sum_{r \in Q} p_{r,q} A_{(r,q)} \hat{P}_r A_{(r,q)}^T.$$

Note that $A_{(r,q)}\Pi_{r_1} = 0$ for $r_1 \neq r$, since $A_{(r,q)}|_{\mathcal{X}_{r_1}} = 0$, since $\mathcal{X}_{r_1}$ belongs to the linear span of elements of $\mathrm{Im}A_{(r_1,q_1)}$ and $\mathrm{Im}K_{(r_1,q_1)}$, $q_1 \in Q$, and Part 5 of Assumption 2. Hence, if $\hat{P} = \sum_{q \in Q} \hat{P}_q$, then $A_{(r,q)} \hat{P}_r A_{(r,q)}^T = A_{(r,q)} \hat{P} A_{(r,q)}^T$. Denote by $\mathcal{Z}$ the linear map $\mathbb{R}^{n^2 \times n^2} \mapsto \sum_{(r,q) \in Q \times Q} p_{r,q} A_{(r,q)} V A_{(r,q)}^T$. From the discussion above it follows that $\hat{P}$ is an eigenvector of $\mathcal{Z}$ corresponding to the eigenvalue λ. From [41, Chapter 2] it follows $\sum_{(r,q) \in Q \times Q} p_{r,q} A_{(r,q)} \otimes A_{(r,q)}$ is just a matrix representation of $\mathcal{Z}$. Then Part 4 of Assumption 2 implies that the eigenvalues $(\sum_{(r,q) \in Q \times Q} p_{r,q} A_{(r,q)} \otimes A_{(r,q)})^T = \sum_{(r,q) \in Q \times Q} p_{r,q} A_{(r,q)}^T \otimes A_{(r,q)}^T$ all inside the unit disk. Since takings transposes does not change the eigenvalues, it then follows that all the eigenvalues of $\sum_{(r,q) \in Q \times Q} p_{r,q} A_{(r,q)} \otimes A_{(r,q)}$, and hence of $\mathcal{Z}$, are inside the unit disk as well. Since λ was an arbitrary eigenvalue of $\widetilde{M}^T$, and $\widetilde{M}$ and $\widetilde{M}^T$ have the same eigenvalues, it follows that Part 5 of Assumption 7 holds.

Part 6 of Assumption 7 A direct consequence of Part 5 of Definition 2.

Proof that $\mathbb{R}^n = \bigoplus_{q \in Q} \mathcal{X}_q$

Consider the matrix B_σ of B defined in (18). It then follows that $K_{(q_1,q_2)} Q_{(q_1,q_2)} = B_{(q_1,q_2)} - A_{(q_1,q_2)} P_{(q_1,q_2)} C^T$. where $Q_{q_1,q_2} = E[\tilde{\mathbf{e}}(t)\tilde{\mathbf{e}}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)]$. From Lemma 20 it follows that $Q_{q_1,q_2} = \mathbf{M}_{q_2}^T E[\mathbf{e}(t)\mathbf{e}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)]\mathbf{M}_{q_1}$. Since $\mathbf{e}(t)$ and $\boldsymbol{\theta}(t+1)$ are conditionally independent given $\mathcal{D}_t$, it follows that $E[\mathbf{e}(t)\mathbf{e}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)] = E[\mathbf{e}(t)\mathbf{e}^T(t)\chi(\boldsymbol{\theta}(t) = q_1)]p_{q_1,q_2} > 0$. Notice, moreover, that $\mathbf{M}_{q_1}\mathbf{M}_{q_1}^T = I_{q_1}$. Hence, by multiplying $K_{(q_1,q_2)} Q_{(q_1,q_2)} = B_{(q_1,q_2)} - A_{(q_1,q_2)} P_{(q_1,q_2)} C^T$ by $\mathbf{M}_{q_1}^T (E[\mathbf{e}(t)\mathbf{e}^T(t)\chi(\boldsymbol{\theta}(t) = q_1)])^{-1} p_{q_1,q_2}^{-1}$

from the right, we obtain that $K_{(q_1,q_2)}\mathbf{M}_{q_1}^T$ belongs to the linear span of elements of the form $\mathrm{Im}B_{(q_1,q_2)}$ and $A_{(q_1,q_2)}z$, $z \in \mathbb{R}^n$.

Also notice that $\mathbf{M}_q C A_w B_\sigma = E[\widetilde{\mathbf{y}}(t)(\mathbf{z}^{\widetilde{\mathbf{y}}}_{\sigma w}(t))^T] = 0$, if the last component of the last letter of σw is not $q \in Q$. Since B is reachable, any $z \in \mathbb{R}^n$ is a linear combination of vectors from $\mathrm{Im}A_w B_\sigma$ for some $w \in \Sigma^*$, $\sigma \in \Sigma$. Hence, $\mathbf{M}_q C A_{(q_1,q_2)} = 0$ and $\mathbf{M}_q C B_{(q_1,q_2)} = 0$ for all $q_1, q_2, q \in Q$ such that $q_2 \neq q$. Combining this with the definition of $\mathcal{X}_q$, $q \in Q$ and the fact derived above that $K_{(q_1,q_2)}$ is spanned by elements $\mathrm{Im}A_{(q_1,q_2)}$, $\mathrm{Im}B_{(q_1,q_2)}$, it follows that $\mathbf{M}_q C x = 0$ for all $x \in \mathcal{X}_{q_1}$, $q_1 \neq q$.

We are now ready to prove that $\mathbb{R}^n = \bigoplus_{q \in Q} \mathcal{X}_q$. From the discussion above, it follows that $\mathcal{X}_{q_1} \cap \mathcal{X}_{q_2} = \{0\}$. Indeed, if $x \in \mathcal{X}_{q_1} \cap \mathcal{X}_{q_2}$, then for $q \neq q_1$, $\mathbf{M}_q C x = 0$, and since $q_1 \neq q_2$ and $x \in \mathcal{X}_{q_2}$, $\mathbf{M}_{q_1} C x = 0$. Hence, $Cx = 0$. Moreover, notice that $\mathcal{X}_q \subseteq \ker A_{(q_3,q_4)}$ for $q \neq q_3$, since $A_{(q_3,q_4)} K_{q_5,q} = 0$ and $A_{(q_3,q_4)} A_{(q_5,q)} = 0$ for all $q_5 \in Q$. By applying this result to $q = q_1$ and $q = q_2$, it follows that $A_{(q_3,q_4)} x = 0$ for any $q_1, q_4 \in Q$ and hence $A_w x = 0$ for any $w \in \Sigma^+$. That is, $C A_w x = 0$ for all $w \in \Sigma^*$, i.e. $x \in O_{R_{\mathrm{B}}}$. Since B is observable, it then follows that $x = 0$.

It is left to show that $\mathbb{R}^n = \sum_{q \in Q} \mathcal{X}_q$. To this end, consider the definition of R_{B}. As it was already mentioned, $\mathbf{x}(t)\chi(\boldsymbol{\theta}(t) = q)$ belongs to $\mathcal{X}_q$ for $q \in Q$ almost everywhere. Hence, the columns of $P_{(q_1,q_2)} = E[\mathbf{x}(t)\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)]$ belong to $\mathcal{X}_{q_1}$: take any $M \in \mathbb{R}^{n - n_{q_1} \times n}$ such that $\mathcal{X}_{q_1} = \ker M$; then $M\mathbf{x}(t)\chi(\boldsymbol{\theta}(t) = q) = 0$ almost everywhere, and hence $M P_{(q_1,q_2)} = 0$. It then follows that $\mathrm{Im}A_{(q_1,q_2)} P_{(q_1,q_2)} C^T \subseteq \mathcal{X}_{q_2}$. From the previous discussion it follows that $K_{(q_1,q_2)} Q_{(q_1,q_2)} = p_{q_1,q_2} K_{(q_1,q_2)} \mathbf{M}_{q_1}^T E[\mathbf{e}(t)\mathbf{e}^T(t)\chi(\boldsymbol{\theta}(t) = q_1)]\mathbf{M}_{q_1}$ and hence $\mathrm{Im}K_{(q_1,q_2)} Q_{(q_1,q_2)} \subseteq \mathcal{X}_{q_2}$. Combining all this with the definition of R_{B} it follows that $\mathrm{Im}B_{(q_1,q_2)} \subseteq \mathcal{X}_{q_2}$. Since $A_{(q,q_3)}(\mathcal{X}_q) \subseteq \mathcal{X}_{q_3}$, we obtain that $\mathrm{Im}A_w B_{(q_1,q_2)}$ always belongs to $\mathcal{X}_q$, where q is the last component of the last letter of $(q_1, q_2)w$. From reachability of R_{B} we then obtain that $\mathbb{R}^n = \sum_{q \in Q} \mathcal{X}_q$, as claimed. ■

Now we can also easily prove Lemma 19. In fact, we will prove first a technical result, relating state covariances of H and B_H. From this Lemma 19 follows easily.

Lemma 24: Assume that H satisfies Assumption 7. Let $\hat{P}_{q_1,q_2} = E[\hat{\mathbf{x}}(t)\hat{\mathbf{x}}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)]$. Then for $P_q = E[\mathbf{x}(t)\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = q)]$,

$$\hat{P}_{q_1,q_2} = p_{q_1,q_2} \mathbf{I}_{q_1} P_{q_1} \mathbf{I}_{q_1}^T.$$

Similarly, if $\hat{Q}_{q_1,q_2} = E[\hat{\mathbf{v}}(t)\hat{\mathbf{v}}(t)\chi(\boldsymbol{\theta}(t) = 1, \boldsymbol{\theta}(t+1) = q_2)]$, and $Q_{q_1} = E[\mathbf{v}(t)\mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t) = q_1)]$, then

$$\hat{Q}_{q_1,q_2} = p_{q_1,q_2}\mathbf{S}_{q_1}^T Q_{q_1}\mathbf{S}_{q_1}.$$

Proof of Lemma 24: The second statement of the lemma was already shown in the proof of Lemma 21, while showing that B_H satisfies Part 5 of Assumption 2 holds.

We proceed with the proof of the first statement. From the construction of $\hat{\mathbf{x}}(t)$ it follows that $\hat{P}_{(q_1,q_2)} = \mathbf{I}_{q_1}E[\mathbf{x}(t)\mathbf{x}(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)]\mathbf{I}_{q_1}^T$. Hence, it is enough to show that $E[\mathbf{x}(t)\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)] = p_{q_1,q_2}P_{q_1}$. To this end, notice that $E[\mathbf{x}(t)\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)] = E[E\mathbf{x}(t)\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2 \mid \mathcal{D}_t]]$. Also notice that from Part 3 of Assumption 7 it follows that $\mathbf{x}(t)$ is measurable w.r.t. the σ-algebra generated by $\{\mathbf{v}(t-l)\}_{l\geq 0}$. Indeed, Part 3 of Assumption 7 and Lemma 1 implies that $\mathbf{x}(t)$ measurable w.r.t. to the σ-algebra generated by $\{\mathbf{v}(t-l)\}_{l\geq 0}$.

From Part 2 of Assumption 7 it then follows that $\mathbf{x}(t)$ and $\boldsymbol{\theta}(t), \boldsymbol{\theta}(t+1)$ are conditionally independent given $\mathcal{D}_t$. Hence $E[\mathbf{x}(t)\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2) \mid \mathcal{D}_t] = p_{q_1,q_2}E[\mathbf{x}(t)\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = q_1) \mid \mathcal{D}_t]$. Combining this with the discussion above yields that $E[\mathbf{x}(t)\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)] = p_{q_1,q_2}E[\mathbf{x}(t)\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = q_1)]$. ■

Proof of Lemma 19: Consider the **GBS** B_H associated with H. From the construction of the matrices of B_H it follows that the solutions to (53) and those of (14) interpreted for $\mathrm{B} = \mathrm{B}_H$ can be related as follows. Suppose that $\{P_q\}_{q\in Q}$ is a solution to (53). From Lemma 24 it follows that $\hat{Q}_{(q_1,q_2)} = p_{q_1,q_2}\mathbf{S}_{q_1}^T Q_{q_1}\mathbf{I}_{S_{q_1}}$. Define $\hat{P}_{(q_1,q_2)} = p_{q_1,q_2}\mathbf{I}_{q_1}P_{q_1}\mathbf{I}_{q_1}^T$. Notice that $\mathbf{I}_q^T\mathbf{I}_q = I_{n_q}$ and $\mathbf{S}_q\mathbf{S}_q^T = I_m$. If we multiply (53) by $\mathbf{I}_q$ from the right and by $\mathbf{I}_q^T$ from the right, then using the discussion above and the definition of $A_{(q_1,q_2)}$, K_{q_1,q_2} we readily obtain that $\{\hat{P}_{(q_1,q_2)}\}_{(q_1,q_2)\in Q\times Q}$ satisfies (14). In addition, notice that the correspondence between $p_{q_1,q_2}P_{q_1}$ and $\hat{P}_{(q_1,q_2)}$ is injective, since $\mathbf{I}_{q_1}$ is full column rank for all $q_1 \in Q$. Since by Lemma 4 (14) has precisely one solution, this implies that (53) has at most one solution.

Next, we show that (53) has a solution. To this end, notice that the unique solution of (14) is of the form $\hat{P}_{(q_1,q_2)} = E[\hat{\mathbf{x}}(t)\hat{\mathbf{x}}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)]$. Notice that the only non-zero block of $\hat{P}_{(q_1,q_2)}$ is the one which corresponds to $p_{q_1,p_2}E[\mathbf{x}(t)\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = q_1)]$. Define now $P_q = E[\mathbf{x}(t)\mathbf{x}^T(t)\chi(\boldsymbol{\theta}(t) = q)]$, $q \in Q$. From the discussion above and Lemma 24 and the definition of the matrices $A_{(q_1,q_2)}$ and $K_{(q_1,q_2)}$ it is easy to see that $\{P_q\}_{q\in Q}$ satisfies (53). ■

Now we are ready to present the proof of Theorem 9.

Proof of Theorem 9: **Necessity**

If $\mathbf{y}$ has a realization by a GJMLS which satisfies Assumption 7, then by Lemma 21, $\mathbf{y}$ can be realized by a **GBS** which satisfies Assumption 2. By Theorem 5, the latter implies that $\mathbf{y}$ satisfies Assumption 3. Moreover, the second statement of Lemma 21 together with Theorem 5 implies that $\mathbf{y}$ is full rank. Hence, $\mathbf{y}$ satisfies the first part of Assumption 8.

Finally, the validity of Part 2 of Assumption 8 can be obtained as follows. Let $\mathcal{F}_t$ be the σ-algebra generated by $\{\mathbf{v}(t-l)\}_{l\le 0}$. Let $\mathcal{D}_t^+$ be the σ-algebra generated by $\{\boldsymbol{\theta}(t+l)\}_{l\ge 0}$. From Part 3 of Assumption 7 and $\mathbf{y}(t) = C_{\boldsymbol{\theta}(t)}\mathbf{x}(t) + D_{\boldsymbol{\theta}(t)}\mathbf{v}(t)$ it follows that $\mathbf{y}(t)$ is measurable with respect to the joint σ-algebra $\mathcal{F}_t \vee \mathcal{D}_t$. Hence, the σ-algebra $\mathcal{H}_t$ generated by $\{\mathbf{y}(t-l)\}_{l\ge 0}$ is a sub-algebra of $\mathcal{F}_t \vee \mathcal{D}_t$. Since by Part 2 of Assumption 7 $\mathcal{F}_t$ and $\mathcal{D}_t^+$ are conditionaly independent given $\mathcal{D}_t$, from the well-known properties of conditional independence it follows that $\mathcal{F}_t \vee \mathcal{D}_t$ and $\mathcal{D}_t^+$ are conditionally independent too. Hence, $\mathcal{H}_t$ and $\mathcal{D}_t^+$ are conditionally independent given $\mathcal{D}_t$,

Sufficiency Assume that $\mathbf{y}$ satisfies Assumption 8. From Theorem 4 it follows that $\mathbf{y}$ admits a **GBS** Σ realization in forward innovation form which satisfies Assumption 2. From Lemma 22 it then follows that the GJMLS H_Σ associated with Σ is a realization of $\mathbf{y}$ and it satisfies Assumption 7.

■

Proof of Lemma 23: Consider the **GBS** BS_H associated with H from (60). Then it is easy to see that $R_H = (\mathbb{R}^n, \{\sqrt{p_\sigma}A_\sigma\}_{\sigma\in\Sigma}, B, C)$, where $B = \{B_{(\sigma,j)} \mid \sigma \in \Sigma, j = 1, \ldots, p\}$ and with $B_\sigma = \begin{bmatrix} B_{\sigma,1} & \ldots B_{\sigma,p} \end{bmatrix}$,

$$B_\sigma = \sqrt{p_\sigma}(A_\sigma \hat{P}_\sigma C^T + K_\sigma Q_\sigma D_\sigma^T)$$

where $\hat{P}_\sigma = E[\hat{\mathbf{x}}(t)\hat{\mathbf{x}}^T(t)\mathbf{u}_\sigma^2(t)]$. From Lemma 24 it then follows that

$$B_{(q_1,q_2)} = \sqrt{p_{q_1,q_2}}\mathbf{I}_{q_2}^T G_{q_1,q_2}\mathbf{M}_{q_1},$$

where $\mathbf{M}_q = \begin{bmatrix} \mathbf{O}_{p,p(q-1)}, & I_p, & \mathbf{O}_{p,p(d-q-1)} \end{bmatrix} \in \mathbb{R}^{p\times pd}$.

Note that R_H is reachable if and only the elements of $\mathrm{Im}\sqrt{p_w}A_w B_{(q_1,q_2)}$, $w \in (Q\times Q)^*$, $|w| \le n-1$, $(q_1,q_2) \in Q\times Q$, span the whole space. Notice that $A_w B_{(q_1,q_2)} = 0$ if $(q_1,q_2)w \notin L$ and that $B_{(q_1,q_2)} \in \mathrm{Im}\mathbf{I}_q$, and $A_w B_{(q_1,q_2)}$ belongs to $\mathrm{Im}\mathbf{I}_q$, if w ends in a letter (q_3,q). Hence, reachability of R_H is equivalent to requiring that the span of columns of $A_{(q,q_3)w}B_{(q_1,q_2)}, B_{(q,q_4)}$ for all q_1, q_2, q_3, q_4, $|w| \leq n-2$, $w \in L$ equals $\mathrm{Im}\mathbf{I}_q$ for all $q \in Q$. Notice that $\mathbf{M}_q$ is full row rank, hence $\mathrm{Im}A_{(q,q_3)w}B_{(q_1,q_2)} = \mathrm{Im}\mathbf{I}_q M_{(q,q_3)w}G_{(q_1,q_2)}\mathbf{M}_{q_1} = \mathrm{Im}\mathbf{I}_q M_{(q,q_3)w}G_{(q_1,q_2)}$ and $\mathrm{Im}B_{(q_4,q)} = \mathrm{Im}\mathbf{I}_q G_{q_4,q}\mathbf{M}_{q_4} = \mathrm{Im}\mathbf{I}_q G_{q_4,q}$ for all q_1, q_2, q_3, q_4, $|w| \leq n-2$, $w \in L$. It then follows that the span of those vectors equals $\mathrm{Im}\mathbf{I}_q \mathcal{R}_{H,q}$. Since $\mathrm{rank}\ \mathbf{I}_q = n_q$, reachability of R_H is indeed equivalent to $\mathrm{rank}\ \mathcal{R}_{H,q} = n_q$ for all $q \in Q$.

From the definition of R_H and B_H it follows that $\mathbf{M}_r C A_w A_{(q_1,q)} \mathbf{I}_q = 0$ if $(q_1, q)w$ does not end in a letter (q_2, r), $q_2 \in Q$, and $\mathbf{M}_r C A_w A_{(q_1,q)} \mathbf{I}_q = C_r M_{w(q_1,q)}$ otherwise, for any $r, q_1, q \in Q$, $w \in \Sigma^*$. Hence, $\ker C A_w \mathbf{I}_q = \ker C_r M_w$ for all $w \in L$ such that w ends in (q_2, r). Notice that $C\mathbf{I}_q = C_q$. Finally, we remark that $w \notin L$, then $CA_w = 0$ and if w does not start with a letter of the form (q, q_1), then $CA_w \mathbf{I}_q = 0$. From the discussion above it then follows that $O_{R_H} \cap \mathrm{Im}\mathbf{I}_q = \mathbf{I}_q(\mathcal{O}_{H,q})$.

Assume now that R_H is observable, i.e. $O_{R_H} = \{0\}$. Since $\mathbf{I}_q$ is full column rank, we then get that $\mathcal{O}_{H,q} = \{0\}$, $q \in Q$. Conversely, assume that $\mathcal{O}_{H,q} = \{0\}$ for all $q \in Q$. It then follows that $O_{R_H} \cap \mathrm{Im}\mathbf{I}_q = \{0\}$. Let $x = (x_1^T, \ldots, x_d^T)^T \in \mathbb{R}^n$, $x_q \in \mathbb{R}^{n_q}$, $q \in Q$, and assume that $x \in O_{R_H}$. Notice that $Cx = \begin{bmatrix}(C_1x_1)^T, & \ldots, & (C_dx_d)^T\end{bmatrix}^T$ and $Cx_q = \mathbf{M}_q C_q x_q = C\mathbf{I}_q x$, $q \in Q$. Hence, $Cx = 0$ is equivalent to $C_q x_q = 0$. Moreover, for any $q_1, q_2 \in Q$, $A_{(q_1,q_2)}x = A_{(q_1,q_2)}\mathbf{I}_{q_1}x_{q_1}$ and $A_{(q_1,q_2)}\mathbf{I}_q x_q = 0$ for $q \neq q_1$. Hence, $x \in O_{R_H}$ implies that $CA_w\mathbf{I}_q x_q = 0$ for any $q \in Q$, $w \in \Sigma^*$, $|w| \leq n-1$. Hence, $\mathbf{I}_q x_q \in O_{R_H} \cap \mathrm{Im}\mathbf{I}_q$. Since we have shown above that $O_{R_H} \cap \mathrm{Im}\mathbf{I}_q = \{0\}$, it follows that $\mathbf{I}_q x_q = 0$, $q \in Q$. Since $\mathbf{I}_q$ is full column rank, it follows that $x_q = 0$ for all $q \in Q$. Hence, $x = 0$. ■

Proof of Theorem 10:

Minimality $\implies$ reachability and observability. Assume that H is a minimal realization of $\mathbf{y}$ and assume that it is not reachable or observable. Consider the **GBS** B_H associated with H. From Lemma 21 it follows that B_H is a realization of $\widetilde{\mathbf{y}}$. From Lemma 23 it follows R_{B_H} cannot be reachable and observable. Then by Theorem 7 B_H cannot be minimal. Take a minimal realization B of $\widetilde{\mathbf{y}}$ in forward innovation form. Then $\dim \mathrm{B} < \dim \mathrm{B}_H = \dim H$. Construct the GJMLS H_{B} associated with B. By Lemma 22, H_{B} is a realization of $\mathbf{y}$ and $\dim H_{\mathrm{B}} = \dim \mathrm{B} < \dim H$. This contradicts to minimality of H and hence a contradiction.

Reachability and observability $\implies$ minimality Assume that H is reachable and observable but it is not a minimal realization of $\mathbf{y}$. Consider the associated **GBS** B_H. From Lemma 23 it follows that $R_H = R_{\mathrm{B}_H}$ is reachable and observable. From Theorem 7 and Lemma 21 it then follows that B_H is a minimal realization of $\widetilde{\mathbf{y}}$. Assume that H is not minimal. Then there exists a GJMLS $\hat{H}$ such that $\dim \hat{H} < \dim H$, $\hat{H}$ is a realization of $\mathbf{y}$ and it satisfies Assumption 7. From Lemma 22 it then follows that $\mathrm{B}_{\hat{H}}$ is a realization of $\widetilde{\mathbf{y}}$. Since $\dim \hat{H} = \dim \mathrm{B}_{\hat{H}}$ and $\dim H = \dim \mathrm{B}_H$, it follows that $\dim \mathrm{B}_{\hat{H}} < \dim \mathrm{B}_H$, which contradicts the minimality of B_H.

Minimal realizations are isomorphic If H and $\hat{H}$ are two minimal realizations of $\mathbf{y}$ such that they both satisfy Assumption 7, then by Lemma 21 the **GBS**s B_H and $\mathrm{B}_{\hat{H}}$ are minimal realizations of $\widetilde{\mathbf{y}}$ which satisfy Assumption 2. From Theorem 7 it then follows that the representations $R_H = R_{\mathrm{B}_H}$ and $R_{\hat{H}} = R_{\mathrm{B}_{\hat{H}}i}$ are isomorphic and they are both reachable and observable. Consider this isomorphism $\mathbf{S} : R_H \to R_{\hat{H}}$. It is easy to see that $\mathbf{S}$ is then an isomorphism between H and $\hat{H}$. ■

VI. Discussion and Conclusion

We have presented a realization theory for stochastic jump-linear systems. The theory relies on the solution of a generalized bilinear filtering/realization problem. This solution represents an extension of the known results on linear and bilinear stochastic realization/filtering.

We would like to extend the presented results to more general classes of hybrid systems. In particular, we would like to develope realization theory for jump-linear systems with partially observed discrete states. Necessary conditions for existence of a realization by a system of this class were already presented in [40]. Another line of research we would like to pursue is to use the presented theory for developing subspace identification algorithms for stochastic jump-linear systems. Note that the classical stochastic bilinear realization theory gave rise to a number of subspace identification algorithms, see [16], [18], [17], [15]. It is very likely that the presented results will lead to very similar subspace identification algorithms.

Acknowledgement. This work was supported by grants NSF EHS-05-09101, NSF CAREER IIS-04-47739, and ONR N00014-05-1083.

References

[1] B. L. Ho and R. E. Kalman, "Effective construction of linear state-variable models from input/output data," in *Proc. Third Annual Allerton Conf. on Circuit and System Theory*. University of Illinois, 1965, pp. 449–459. 2, 7

[2] R. E. Kalman, P. L. Falb, and M. A. Arbib, *Topics in mathematical system theory*. New York: McGraw-Hill Book Co., 1969. 2

[3] L. Ljung, *System Identification: theory for the user*. Prentice Hall, 1987. 2

[4] P. V. Overschee and B. D. Moor, *Subspace Identification For Linear Systems: Theory, Implementation, Applications*. Kluwer Academic Publishers, 1996. 2

[5] R. W. Brockett, "On the algebraic structure of bilinear systems," in *Theory and Applications of Variable Structure System*, R. Mohler and A. Ruherti, Eds. New York, Academic, 1972, p. 1972. 2

[6] M. Fliess, "Sur la realization des systemes dynamiques bilineaires," *C. R. Académie des Sciences Paris*, vol. A, no. 277, pp. 243–247, 1973. 2

[7] P. D'Alessandro, A. Isidori, and A. Ruberti, "Realization and structure theory of bilinear dynamical systems," *SIAM Journal on Control and Optimization*, vol. 12, no. 3, pp. 517–535, 1974. 2

[8] A. Isidori, "Direct construction of minimal bilinear realizations from nonlinear input-output maps," *IEEE Transactions on Automatic Control*, pp. 626–631, 1973. 2

[9] A. Isidori and A. Ruberti, "Realization theory of bilinear systems," in *Geometric Methods in System Theory*, D. Q. Mayne and R. W. Brockett, Eds. D. Reidel Publishing Co., 1973. 2

[10] E. D. Sontag, "Bilinear realizability is equivalent to existence of a singular affine differential I/O equation," *Systems & Control Letters*, vol. 11, no. 3, pp. 181–187, 1988. 2

[11] H. Sussman, "Semigroup representations, bilinear approximations of input-output maps, and generalized inputs," in *Mathematical Systems Theory*, 1975, pp. 172–192. 2

[12] ——, "Minimal realizations and canonical forms for bilinear systems," *J. Franklin Institute*, vol. 301, no. 6, pp. 593–604, 1976. 2

[13] U. Desai, "Realization of bilinear stochastic systems," *IEEE Transactions on Automatic Control*, vol. 31, no. 2, 1986. 2, 4, 11, 13, 16, 18, 20, 23, 28

[14] A. E. Frazho, "On stochastic bilinear systems," in *Modelling and application of stochastic processes*. Boston, MA: Kluwer Acad. Publ., 1986, pp. 215–241. 2, 13, 18, 20, 28

[15] H. Chen and J. Maciejowski, "New subspace identification method for bilinear systems," University of Cambridge, Tech. Rep. CUED/FINFENG/TR.337, 2000. 2, 4, 32, 70

[16] M. Favoreel, "Subspace methods for identification and control of linear and bilinear systems," Ph.D. dissertation, K.U. Leuven, 1999. 2, 4, 32, 70

[17] V. Verdult and M. Verhaegen, "Identification of multivariable bilinear state space systems based on subspace techniques and separable least squares optimization," *International Journal of Control*, vol. 74, no. 18, pp. 1824–1836, 2001. 2, 4, 70

[18] ——, "Subspace identification of multivariable linear parameter-varying systems," *Automatica*, vol. 38, pp. 805–814, 2002. 2, 4, 32, 70

[19] M. Fliess, "Realizations of nonlinear systems and abstract transitive lie-algebras," *Bulletin of Americal Mathematical Society*, vol. 2, no. 3, 1980. 2

[20] B. Jakubczyk, "Construction of formal and analytic realizations of nonlinear systems," in *Feedback Control of Linear and Nonlinear Systems*, ser. Lecture Notes in Control and Information Sciences, Isodori and Hinrichsen, Eds., vol. 39, 1982, pp. 147 – 156. 2

[21] ——, "Realization theory for nonlinear systems, three approaches," in *Algebraic and Geometric Methods in Nonlinear Control Theory*, M. Fliess and M. Hazewinkel, Eds. D.Reidel Publishing Company, 1986, pp. 3–32. 2

[22] F. Celle and J. Gauthier, "Realizations of nonlinear analytic input-output maps," *Math. Systems Theory*, vol. 19, pp. 227 – 237, 1987. 2

[23] C. Reutenauer, “The local realization of generating series of finite lie-rank,” in *Algebraic and Geometric Methods in Nonlinear Control Theory*, M. Fliess and M. Hazewinkel, Eds. D.Reidel Publishing Company, 1986, pp. 33–43. 2

[24] Y. Wang and E. D. Sontag, “Realization and input/output relations: the analytic case,” in *Proceedings of the 28th Conference on Decision and Control*, 1989, pp. 1975–1980. 2

[25] B. Jakubczyk, “Existence and uniqueness of realizations of nonlinear systems,” *SIAM Journal on Control and Optimization*, vol. 18, no. 4, pp. 455–471. 2

[26] H. Sussman, “Existence and uniqueness of minimal realizations of nonlinear systems,” in *Mathematical Systems Theory*, vol. 3, 1977, pp. 263–284. 2

[27] Z. Bartosiewicz, “Realizations of polynomial systems,” in *Algebraic and geometric methods in nonlinear control theory*, ser. Math. Appl. Dordrecht: Reidel, 1986, vol. 29, pp. 45–54. 2

[28] E. D. Sontag, *Polynomial Response Maps*, ser. Lecture Notes in Control and Information Sciences. Springer Verlag, 1979, vol. 13. 2, 5, 7

[29] Y. Wang and E. Sontag, “On two definitions of observation spaces,” *Systems and Control Letters*, vol. 13, pp. 279–289, 1989. 2

[30] ——, “Algebraic differential equations and rational control systems,” *SIAM Journal on Control and Optimization*, no. 30, pp. 1126–1149, 1992. 2

[31] ——, “Generating series and nonlinear systems: analytic aspects, local realizability and i/o representations,” *Forum Mathematicum*, no. 4, pp. 299–322, 1992. 2

[32] R. Grossman and R. Larson, “An algebraic approach to hybrid systems,” *Theoretical Computer Science*, vol. 138, pp. 101–112, 1995. 2

[33] M. Petreczky, “Realization theory for linear switched systems,” in *Proceedings Sixteenth International Symposium on Mathematical Theory of Networks and Systems*, 2004, draft available at `http://www.cwi.nl/~mpetrec`. 2

[34] ——, “Realization theory for linear switched systems: Formal power series approach,” *Systems and Control Letters, to appear*, 2007. 2

[35] ——, “Realization theory for bilinear switched systems,” in *Proceedings of 44th IEEE Conference on Decision and Control*, 2005, pp. 690–695. 2

[36] ——, “Realization theory for bilinear hybrid systems,” in *11th IEEE Conference on Methods and Models in Automation and Robotics 2005*, cD-ROM only. 2

[37] ——, “Hybrid formal power series and their application to realization theory of hybrid systems,” in *Proc. 17th International Symposium on Mathematical Networks and Systems*, 2006. 2

[38] M. Petreczky and J.-B. Pomet, “Realization theory of nonlinear hybrid systems,” in *Taming Heterogenity and Complexity of Embedded Control (CTS-HYCON Workshop on Embedded and Hybrid Control)*. International Scientific & Technical Encyclopedia (ISTE), London, 2006. 3

[39] M. Petreczky, “Realization theory for discrete-time piecewise-affine hybrid systems,” in *Proceedings 17th International Symposium on Mathematical Networks and Systems*, 2006. 3

[40] M. Petreczky and R. Vidal, “Metrics and topology for nonlinear and hybrid systems,” in *Hybrid Systems: Computation and Control*. Springer Verlag, 2007. 3, 4, 5, 70

[41] O. Costa, M. Fragoso, and R. Marques, *Discrete-Time Markov Jump Linear Systems*. Springer Verlag, 2005. 3, 9, 10, 37, 45, 48, 50, 53, 55, 57, 65

[42] M. Petreczky, “Realization theory of hybrid systems,” Ph.D. dissertation, Vrije Universiteit, Amsterdam, 2006, available online: www.googlsite.com/mihaly.petreczky. 5, 6, 7, 8

[43] J. Berstel and C. Reutenauer, *Rational series and their languages*, ser. EATCS Monographs on Theoretical Computer Science. Springer-Verlag, 1984. 5

[44] E. D. Sontag, “Realization theory of discrete-time nonlinear systems: Part I – the bounded case,” *IEEE Transactions on Circuits and Systems*, vol. 26, no. 5, pp. 342–356, 1979. 5, 7, 8

[45] D. E. Sontag, “A remark on bilinear systems and moduli spaces of instantons,” *Systems and Control Letters*, vol. 9, no. 5, pp. 361–367, 1987. 8

[46] P. Bilingsley, *Probability and measure*. Wiley, 1986. 12, 15

[47] G. Popescu, “Positive-definite functions on free semigroups,” *Can. J. Math.*, vol. 48, no. 4, pp. 887 – 896, 199. 18

[48] P. Caines, *Linear Stochastic Systems*. John Wiley and Sons, New-York, 1998. 20, 24

[49] G. Kotsalis, A. Megretski, and M. A. Dahleh, “Balanced truncation for a class of stochastic jump linear systems and model reduction of hidden markov models,” *IEEE Transactions on Automatic Control*, vol. 53, no. 11, 2008. 21

[50] V. Verdult and M. Verhaegen, “Subspace identification of piecewise linear systems,” in *Proceedings of the 43rd IEEE Conference on Decision and Control*, 2004, pp. 3838–3843. 32